PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS

MARIA CLARA GIASSETTI MEDEIROS CORRADINI LOPES

DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS E OPACIDADE ALGORÍTMICA

CAMPINAS

2026

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS
ESCOLA DE CIÊNCIAS HUMANAS, JURÍDICAS E SOCIAIS (ECHJS)
PROGRAMA DE PÓS GRADUAÇÃO *STRICTO SENSU* EM DIREITOS HUMANOS E DESENVOLVIMENTO SOCIAL
MARIA CLARA GIASSETTI MEDEIROS CORRADINI LOPES

DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS E OPACIDADE ALGORÍTMICA

Dissertação apresentada ao Programa de Pós-Graduação *stricto sensu* em DIREITOS HUMANOS E DESENVOLVIMENTO SOCIAL do Escola de Ciências Humanas, Jurídicas e Sociais da Pontifícia Universidade Católica de Campinas, como exigência para obtenção do título de Mestre em Direito.

Orientador: Prof. Dr. Claudio José Franzolin

CAMPINAS
2026

Giassetti Medeiros Corradini Lopes, Maria Clara

L864d Direito à explicação em decisões automatizadas e opacidade algorítmica / Maria Clara Giassetti Medeiros Corradini Lopes. - Campinas: PUC-Campinas, 2026.

100 f.

Orientador: Cláudio José Franzolin.

Dissertação (Mestrado em Direitos Humanos e Desenvolvimento Social) - Faculdade de Direito, Escola de Ciências Humanas, Jurídicas e Sociais, Pontifícia Universidade Católica de Campinas, Campinas, 2026.
Inclui bibliografia.

1. Direitos humanos. 2. Inteligência artificial. 3. Direito à explicação.

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS
ESCOLA DE CIÊNCIAS HUMANAS, JURÍDICAS E SOCIAIS
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM DIREITO

MARIA CLARA GIASSETTI MEDEIROS
DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS E OPACIDADE ALGORÍTMICA

Este exemplar corresponde à redação final da Dissertação de Mestrado em Direito da PUC-Campinas, e aprovada pela Banca Examinadora.

APROVADA: **13 de fevereiro de 2026.**

DR. YURI NATHAN DA COSTA LANNES (FDF)

DRA. ALETEIA HUMMES THAINES (PUC-CAMPINAS)

DRA. CLAUDIO JOSE FRANZOLIN Presidente (PUC-CAMPINAS)

AGRADECIMENTOS

Conquistas dessa magnitude merecem agradecimentos. E, de antemão, já antecipo que meus agradecimentos serão longos, porque para além da certeza de que nada nessa vida se conquista sozinho, eu tive a sorte de ter pessoas e instituições incríveis me apoiando e incentivando a chegar até aqui.

A despeito de formalidades protocolares, agradeço meu grande amor, Mateus Corradini Lopes, pela relação de apoio e incentivo mútuos que construímos, por caminharmos juntos, lado a lado, por não nos deixarmos desistir dos nossos sonhos e mais, por sonharmos juntos!

À minha mãe, obrigada por acreditar em mim, mesmo quando eu travava ferrenhas batalhas com o sistema tradicional de ensino. Quem diria que eu chegaria aqui. Essa conquista também é sua! Obrigada pelos privilégios que, com muito suor, você me proporcionou.

Ao Ricardo Giassetti, meu maior referencial cultural. Se esta pesquisadora se dedicou à leitura de tantos referenciais para desenvolver este trabalho, é porque o seu incentivo começou na minha infância.

À minha família Corradini Lopes, que agradeço em nome de Gisely e Marcelo. O amor que vocês me deram foi disruptivo. Obrigada por vibrarem por mim e comigo!

À família que eu escolhi, Gabriel Reis Garcia, Rafael Reis Garcia e Gustavo de Moraes Martins Pires. Obrigada por todo o apoio de sempre! Vocês foram e são essenciais!

Dito isso, meus mais sinceros agradecimentos ao CNPq, pelo financiamento desta pesquisa (Processo 130692/2024-7). Como acadêmica, aproveito o espaço para ressaltar a importância ímpar dos órgãos de fomento e financiamento à pesquisa brasileira que, apesar dos sensíveis cortes que vêm sofrendo nos últimos anos, persistem em constante luta pelos interesses da humanidade. Que fique registrado: valorizemos a ciência!

Ao meu orientador, Cláudio José Franzolin, me faltam palavras para agradecer. Que felicidade é poder caminhar ao seu lado e levar o seu nome marcado na minha trajetória. Muito obrigada por me acolher, incentivar, advertir, cuidar e desempenhar a tarefa de docência e orientação com tamanha leveza. Meu

carinho e admiração são, sem dúvida, inversamente proporcionais ao pouco tempo que tive o prazer de ser sua orientanda.

Ao Programa de Pós-Graduação em Direitos Humanos e Desenvolvimento Social da Pontifícia Universidade Católica de Campinas, em nome de todo o corpo docente, obrigada! Minha passagem pela PUC-Campinas ficou marcada em mim pelos mais valiosos conhecimentos que pude receber, diretamente das mãos de professores espetaculares. Em especial professores Pedro Peruzzo, por ensinar que questões tão complexas podem e devem ser abordadas com igual sensibilidade; Guilherme Cabral, por determinar meu rigor metodológico; Vinícius Casalino, pelas inúmeras reflexões sobre modelo econômico e Silvio Beltramelli, por acentuar a importância dos dados concretos para provocar mudanças práticas.

Ao meu querido supervisor de estágio docência, Felipe Grizotto, que me acolheu com conselhos sempre perspicazes e cuidadosos! Obrigada!

Ao Instituto Nacional de Pesquisa e Promoção de Direitos Humanos (INPPDH), na pessoa do professor e presidente César Augusto Nunes que, para além de uma grande referência, tenho a sorte de chamar de amigo. Obrigada pelas inúmeras oportunidades, em especial pela supervisão no estágio docência, desenvolvido na Universidade Presbiteriana Mackenzie que, ao ensejo, agradeço em nome do professor Leopoldo Soares.

Ao professor Yuri Lannes, a quem devo não somente pela introdução ao assunto que veio a ser meu tema de pesquisa, mas também pela parceria que frequentemente demando.

À Universidad Pontificia Bolivariana, de Medellín - cuja honra de ter realizado estágio de pesquisa é inestimável -, agradeço na pessoa do professor Freddy Orlando Santamaría Velasco, por possibilitar o enriquecimento não só da pesquisa aqui apresentada, mas também desta pesquisadora em sua totalidade.

Dito isso, não poderia deixar de agradecer à Cristiane de Moraes Ferreira Martins, em especial pela conversa que me impediu de desistir de prestar o processo seletivo que me trouxe até aqui. Essa é mais uma prova de que pequenos gestos têm o poder de influenciar os rumos das vidas ao nosso redor.

Por fim, aos dionisíacos amigos com os quais o mestrado me brindou: Rafael R., Beatriz O., Isadora, Joyce, Rafael G., Beatriz S. e Daniela. Vocês tornaram tudo mais leve e divertido. Obrigada!

RESUMO

Um dos maiores problemas da sociedade da informação, atualmente, é a opacidade algorítmica inerente aos sistemas de inteligência artificial utilizados em decisões automatizadas. Esse aspecto engendra não apenas insegurança jurídica, mas também impacta diretamente na vida dos titulares de dados, em geral, ameaçando converter o ser humano em mero objeto destituído de agência e autonomia. A inescrutabilidade das chamadas "caixas-pretas", além de ferir o direito à informação, reflete em afronta a diversos outros direitos humanos e fundamentais, como a dignidade da pessoa humana, a igualdade e o devido processo legal. A Lei Geral de Proteção de Dados (LGPD) e o Regulamento Geral sobre a Proteção de Dados (GDPR) preveem princípios de transparência e direitos de revisão adotados com o fito de sanar a problemática em comento. Dentre as salvaguardas possíveis, o Direito à Explicação foi eleito para exemplificar a discussão sobre o controle democrático da tecnologia. Nesse ínterim, a indagação que se faz é se existe um direito de explicação positivado nesses ordenamentos e como esse direito é interpretado e utilizado para garantir a inteligibilidade das decisões. Com isso demonstra-se que a explicação atua como um princípio instrumental indispensável para a efetivação da autodeterminação informativa e da não discriminação. Partindo desse ponto central, pretende-se discutir a natureza jurídica deste direito - se autônomo ou derivado -, suas limitações frente ao segredo industrial e à complexidade técnica, além dos desafios regulatórios de sua implementação. Para tanto, adotou-se o método do Direito Comparado, observando os parâmetros metodológicos de análise contextual, alinhado a técnicas de pesquisa bibliográfico-documental e exploratória qualitativa, para uma melhor compreensão das assimetrias entre o modelo brasileiro e o europeu. Conclui-se pela insuficiência da mera transparência (falácia da transparência) e pela necessidade de estruturar uma Governança de IA robusta, capaz de traduzir a opacidade técnica em transparência qualificada, mantendo a observância aos direitos humanos.

Palavras-chave: Direito à explicação. Opacidade algorítmica. Direitos humanos. Transparência. Decisão automatizada.

ABSTRACT

One of the biggest problems in today's information society is the algorithmic opacity inherent in the artificial intelligence systems used in automated decisions. This aspect not only creates legal uncertainty, but also directly impacts the lives of data subjects in general, threatening to turn human beings into mere objects devoid of agency and autonomy. The inscrutability of so-called “black boxes,” in addition to violating the right to information, reflects an affront to various other human and fundamental rights, such as human dignity, equality, and due process of law. The General Data Protection Law (LGPD) and the General Data Protection Regulation (GDPR) provide for principles of transparency and rights of review adopted in order to address the problem in question. Among the possible safeguards, the Right to Explanation was chosen to exemplify the discussion on democratic control of technology. In the meantime, the question that arises is whether there is a right to explanation enshrined in these legal systems and how this right is interpreted and used to ensure the intelligibility of decisions. This demonstrates that explanation acts as an indispensable instrumental principle for the realization of informational self-determination and non-discrimination. Starting from this central point, we intend to discuss the legal nature of this right—whether autonomous or derivative—its limitations in relation to trade secrets and technical complexity, as well as the regulatory challenges of its implementation. To this end, the Comparative Law method was adopted, observing the methodological parameters of contextual analysis, aligned with bibliographic-documentary and qualitative exploratory research techniques, for a better understanding of the asymmetries between the Brazilian and European models. It concludes that mere transparency is insufficient (transparency fallacy) and that there is a need to structure robust AI governance capable of translating technical opacity into qualified transparency, while maintaining respect for human rights.

Keywords: Right to explanation. Algorithmic opacity. Transparency. Automated decision-making. Human Rights

LISTA DE IMAGENS

SUMÁRIO

1 INTRODUÇÃO

A crescente adoção de sistemas de inteligência artificial (IA) em decisões que afetam a vida dos indivíduos — como concessão de crédito, seleção para empregos e até sentenças judiciais — levanta questões fundamentais sobre transparência e *accountability*. No centro desse debate está o direito à explicação, isto é, a exigência de que os cidadãos possam compreender as razões por trás das decisões automatizadas. No entanto, a chamada opacidade algorítmica, resultante da complexidade dos modelos de IA e da proteção de segredos comerciais, impede que esses processos sejam plenamente compreendidos e questionados.

A presente dissertação alinha-se organicamente à Área de Concentração em Direitos Humanos e Desenvolvimento Social, uma vez que a expansão das tecnologias disruptivas - especificamente a Inteligência Artificial e a tomada de decisão automatizada - impõe novos e complexos desafios à proteção da dignidade humana. O estudo em questão transcende a mera análise jurídico-formal para adentrar na dimensão ética e social da tecnologia. Ele aprofunda-se na compreensão de como a opacidade algorítmica e a ausência de *accountability* em decisões automatizadas configuram uma ameaça substancial à proteção de direitos humanos como, por exemplo, a dignidade da pessoa humana, da igualdade e do devido processo legal. Ao atuar como "caixas-pretas", os algoritmos podem não apenas reproduzir vieses e preconceitos preexistentes, mas também os amplificar, tornando-os invisíveis e inquestionáveis. Nesse contexto, a falta de transparência e o poder desproporcional conferido a esses sistemas subvertem princípios fundamentais de justiça e equidade, erigindo barreiras ao desenvolvimento social inclusivo e à proteção dos direitos humanos em um cenário cada vez mais mediado por dados e automação.

Assim, o desenvolvimento social, na era da informação, não pode ser dissociado da capacidade dos indivíduos de compreenderem e contestarem as estruturas tecnológicas que governam aspectos essenciais de suas vidas, como acesso ao crédito, emprego e justiça.

A opacidade algorítmica, ao operar como uma barreira invisível, ameaça converter o ser humano em mero objeto de dados, destituído de agência e autonomia. Portanto, investigar o direito à explicação não é uma tarefa meramente técnica ou consumerista, mas uma exigência de curadoria dos direitos humanos.

Trata-se de assegurar que o avanço tecnológico sirva ao desenvolvimento social inclusivo, impedindo a consolidação de uma 'caixa-preta' social que perpetua discriminações e viola o devido processo legal substancial.

Ademais, a pesquisa insere-se na Linha de Pesquisa de Direitos Humanos e Políticas Públicas, porquanto a regulação da IA é, essencialmente, uma questão de política pública que visa proteger direitos fundamentais.Ademais, a pesquisa insere-se na Linha de Pesquisa de Direitos Humanos e Políticas Públicas, porquanto a regulação da Inteligência Artificial é, essencialmente, uma questão crucial de política pública que transcende aspectos meramente técnicos e econômicos. O foco central desta política é estabelecer um arcabouço normativo e ético robusto que vise, primordialmente, proteger os direitos fundamentais e garantir a dignidade humana em um cenário de rápida e profunda transformação tecnológica. A IA, em seu desenvolvimento e aplicação, pode gerar riscos significativos de discriminação, vieses algorítmicos, opacidade decisória e vigilância massiva, impactando diretamente os direitos à igualdade, à privacidade, ao devido processo legal e à não discriminação. Portanto, a intervenção regulatória é indispensável para mitigar esses riscos e assegurar que o avanço tecnológico esteja em consonância com os princípios democráticos e os direitos humanos universalmente reconhecidos.

A comparação entre o Brasil (Sul Global) e a União Europeia (vanguarda regulatória) é estratégica para analisar como as decisões políticas e a formulação de normas impactam a soberania digital e a proteção local dos direitos. O foco está em como o Estado e a sociedade civil podem, por meio de políticas públicas eficazes, enfrentar a assimetria de poder entre as grandes corporações de tecnologia e os cidadãos, exigindo respostas jurídicas e políticas que garantam a efetividade dos direitos humanos na sociedade em rede.

A opacidade de decisões automatizadas pode comprometer direitos como o à informação e ao contraditório, uma vez que indivíduos são afetados sem compreender o porquê de tais resultados. Embora esses sistemas opacos de inteligência artificial levantem questões éticas profundas, como as que envolvem o consentimento e a autonomia sobre as próprias vidas, a presente análise se concentra na viabilidade do direito à explicação de decisões tomadas por sistemas de inteligência artificial.

A relevância da temática em tela é inegável, haja vista que a IA vem sendo empregada em setores sensíveis, a exemplo da justiça, saúde, finanças e

segurança, onde as decisões algorítmicas podem ocasionar consequências profundas e duradouras.

A preocupação global e nacional com a governança da Inteligência Artificial (IA) é crescente, como atestam relatórios e decisões de alta relevância emitidos por diversos organismos e pela comunidade internacional. Nesse contexto, o Conselho de Direitos Humanos da ONU e o Conselho Nacional dos Direitos Humanos no Brasil têm promovido debates que sublinham a necessidade urgente de um arcabouço regulatório que incorpore princípios como transparência, responsabilidade e equidade para lidar com o impacto da IA nos direitos humanos.

Na esfera regulatória, a União Europeia, por meio da Lei da IA do Parlamento Europeu, estabeleceu obrigações para os desenvolvedores, graduadas conforme o nível de risco dos sistemas, e concedeu aos cidadãos o direito de apresentar queixas e de obter explicações sobre decisões que afetem seus direitos.

Adicionalmente, o desenvolvimento de tecnologias de vigilância, como o *Spyware*, intensificou essa preocupação, motivando o Chefe de Direitos Humanos da ONU a pleitear uma moratória no uso de certas tecnologias. Esse panorama reforça a urgência e a pertinência do presente estudo.

Isto é, a temática desta dissertação justifica-se em razão da especial relevância da questão no campo jurídico e social, haja vista que a falta de explicabilidade pode comprometer direitos fundamentais, como o devido processo legal e a igualdade. A ausência de fontes heterônomas para promover esse direito reforça a necessidade de um aprofundamento teórico e legislativo, visando equilibrar inovação tecnológica e proteção de direitos. Assim, esta dissertação busca explorar os desafios e as possíveis soluções para assegurar a transparência e a justiça no uso da IA.

Assim, esta pesquisa centra-se no seguinte problema de pesquisa: existe um direito de explicação positivado na GDPR e na LGPD, e como esse direito é interpretado e utilizado?

Nesse sentido, a relevância da investigação é reforçada pelo próprio fundamento axiológico da legislação brasileira. A Lei Geral de Proteção de Dados (LGPD), longe de ser um estatuto meramente procedimental, declara expressamente em seu artigo 1º que a disciplina do tratamento de dados tem como objetivo a proteção dos direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

Ao estabelecer a dignidade da pessoa humana como alicerce, a LGPD reconhece que a autodeterminação informativa é pressuposto para o exercício da cidadania. Consequentemente, a ausência de explicação sobre decisões automatizadas não configura apenas uma falha de serviço, mas uma afronta direta a esses fundamentos basilares. Se o indivíduo não pode conhecer os critérios que definem seu perfil ou determinam seu acesso a bens da vida, sua dignidade é vilipendiada pela inescrutabilidade da máquina. É sob essa ótica analítica, e não puramente tecnicista, que se deve interpretar a tensão entre o segredo industrial e o direito à explicação.

Sustenta-se como hipótese para as problemáticas assinaladas pela pesquisa a existência de um direito de explicação positivado tanto na GDPR quanto na LGPD, mas sua aplicação e interpretação variam significativamente entre os dois contextos. Essa variação pode ser influenciada por fatores culturais, políticos e sociais, o que pode impactar a efetividade do direito à explicação e a proteção dos indivíduos contra a opacidade algorítmica.

O objetivo geral consiste em analisar fontes normativas de modo a avaliar o conteúdo, sua delimitação conceitual e características do direito à explicação.

Os objetivos específicos, por sua vez, tangem (i) apontar o direito à explicação no âmbito da GDPR e da LGPD (ii) abordar sobre o conceito de opacidade algorítmica e sua interface com o direito à transparência do usuário, sob a perspectiva das referidas fontes normativas (iii) debater o direito à explicação e sua interface com o direito fundamental à autodeterminação informativa, dignidade da pessoa humana, igualdade, devido processo legal, informação, contraditório e não discriminação (v) identificar as dificuldades e desafios na implementação do direito à explicação em ambos os sistemas legais.

A presente pesquisa faz uso de uma compreensão analítica sobre direito à explicação e opacidade algorítmica, a partir do método do direito comparado como instrumento à sua realização, na medida em que parte da dialeticidade entre a GDPR e a LGPD para aferir definições sobre o direito à explicação.

Escolheu-se a comparação da legislação brasileira com a europeia porque a última tem assumido a vanguarda regulatória, com propostas que buscam estabelecer diretrizes para o uso ético da IA (como a GDPR e o AI Act), de sorte que

o estudo dessa jurisdição permite, em tese, compreender as tendências globais que influenciam outras regiões, como o Sul Global.

Nada obstante, para escolha da norma comparada, levou-se em consideração o impacto global e a reprodução de assimetrias - visto que a regulação de IA nesses espaços tem efeitos diretos nos países do Sul Global, seja pela exportação de tecnologias, seja pela imposição indireta de padrões regulatórios (em que pese essa influência raramente considerar as realidades locais, reforçando uma lógica neocolonial, na qual países periféricos se tornam meros consumidores de normas e tecnologias desenvolvidas no Norte Global), de sorte que analisar a UE, com aplicação de filtros, permite questionar como essas regulações afetam a soberania digital e a autonomia regulatória de outros países.

Por fim, considera-se que o estudo comparativo dessas legislações deve incluir uma reflexão sobre a falta de protagonismo do Sul Global na definição dessas normas. Isto é, em que pese o recorte comparativo escolhido, a análise da legislação sobre IA na Europa não significa apenas estudar modelo normativo sabidamente dominante, mas também compreender como essas regulações impactam globalmente e reforçam ou desafiam relações de poder desiguais no campo da tecnologia e do direito.

É certo que, ao adotar a referida técnica (direito comparado), esta pesquisa ponderou questões epistemológicas e metodológicas, especialmente no que se refere à tendência de focar no restrito conjunto normativo dos suspeitos habituais países do norte global - os quais geralmente são tidos como prósperos e estáveis. Tal cuidado se deu a fim de não levantar uma presunção de que as experiências legislativas desses países são mais representativas, amplas e importantes que as das demais localidades.

Tal reflexão sobre os problemas atrelados à aplicação da comparação entre direitos, que foi apresentada por Hirschl (2014, p. 16), aponta que a ênfase no norte global geralmente é uma questão política, na medida que pode refletir uma preferência normativa por valores que esses países defendem - o que pode distorcer a compreensão e a avaliação de sistemas normativos mais diversos. Além disso, o autor destaca que a falta de consideração das particularidades culturais e contextuais dos países do sul global pode resultar em uma análise superficial e inadequada das realidades constitucionais e dos direitos humanos em diferentes contextos (Hirschl, 2014, p. 16).

Nesse sentido, Hirschl (2014, p. 54) destaca a dificuldade de equilibrar a universalidade dos direitos com as tradições locais e as normas específicas de cada jurisdição. A referência a um conjunto limitado de jurisdições pode também criar um "senso de moda" na escolha de quais normas e práticas constitucionais são consideradas válidas ou desejáveis, o que pode não refletir a diversidade e a complexidade das realidades constitucionais globais.

Apesar de todas as dificuldades apresentadas, a importância da comparação de direitos remanesce na presente pesquisa, porquanto possibilita uma análise crítica e aprofundada sobre como diferentes sistemas jurídicos enfrentam a questão da explicabilidade das decisões automatizadas, permitindo identificar avanços, lacunas e desafios na proteção dos direitos fundamentais. Além disso, a técnica do direito comparado permite avaliar como a GDPR e a LGPD incorporam o direito à explicação e a transparência algorítmica, fornecendo subsídios para a construção de modelos normativos mais robustos e adequados a diferentes realidades. Frisa-se que tal abordagem também viabiliza uma reflexão sobre o impacto global dessas regulações e sobre a necessidade de evitar a imposição acrítica de padrões jurídicos do Norte Global sobre outras regiões, garantindo que o debate sobre a opacidade algorítmica leve em consideração as particularidades culturais, políticas e sociais de cada contexto.

Diante desse panorama crítico, cumpre destacar de plano que aplicação da técnica de direito comparado se atentou às considerações tecidas por Hirschl (2014), especialmente no que tange à metodologia, com a análise contextual, à prática de “*cherry-picking*”, diversidade de fontes, interação entre normas e práticas e reflexões axiológicas.

Isto é, seguindo os parâmetros definidos por Hirschl, a comparação de direitos deve ser multifacetada e baseada em uma compreensão crítica e contextualizada do direito comparado, devendo ser realizada de maneira cuidadosa e reflexiva, ponderando tanto os métodos utilizados quanto os objetivos da comparação. Do ponto de vista do método, é necessário, segundo o autor, que seja rigoroso e sistemático, utilizando princípios de seleção de casos, tais quais, "casos mais semelhantes" e "casos mais diferentes", a fim de garantir que a comparação seja significativa e informativa. Isso implica, portanto, uma análise cuidadosa das semelhanças e diferenças entre os sistemas jurídicos em questão.

Demais disso, diante da complexidade dos temas abordados, utiliza-se do mecanismo de interdisciplinaridade, especialmente das disciplinas relacionadas à ciência da computação, física, direito empresarial, direito constitucional, e ciências sociais.

A organização deste trabalho foi pensada em 5 (cinco) partes, sendo que, entre introdução e conclusão, existem 3 (três) capítulos.

O primeiro capítulo, intitulado “Fontes Heterônomas sobre a Proteção de Dados e Informações da Pessoa Humana” é dedicado à fundamentação teórica, delimitação dos conceitos de algoritmos, *machine learning* e IA. Nele também é abordado o problema da opacidade algorítmica e suas diferentes causas, como o segredo comercial e a complexidade técnica, estabelecendo a base para o restante da dissertação.

O segundo capítulo, intitulado “Direito de Explicação e sua Compreensão Teórica” contém o cerne da pesquisa. Neste capítulo analisa-se a positivização, as características e as limitações do direito à explicação na LGPD e na GDPR. A seção explora se o direito é autônomo ou uma extensão de outros direitos fundamentais, além de discutir seu conflito com o segredo industrial e a complexidade da IA.

O terceiro capítulo de apresentação sintetiza as descobertas e apresenta uma análise comparativa dos direcionamentos legislativos no Brasil e na Europa. Este discute a viabilidade da adequação das leis e propõe soluções para mitigar a opacidade, com ênfase na transparência e no devido processo no âmbito privado.

Do ponto de vista metodológico, esta pesquisa adotou caráter bibliográfico-documental exploratório qualitativo. Tendo em vista se tratar de uma pesquisa essencialmente sobre a existência de um direito de explicação relacionado à opacidade algorítmica na GDPR e na LGPD, alguns parâmetros foram estabelecidos para fins de limitação do marco teórico. Para seleção da bibliografia base da presente pesquisa, que além dos textos normativos comparados - GDPR e LGPD -, é composta majoritariamente por artigos científicos, foram aplicados critérios objetivos de busca nos sites oficiais.

O Portal de Periódicos da CAPES - CAFE foi escolhido como repositório para realização das buscas, já que contém em si mesmo os principais repositórios (Scielo, Web of Science e Scopus). Como termos de pesquisa (buscadores) foram inseridos: "opacidade algorítmica"; "algorithmic opacity"; "Automated Decision"; "Decisão automatizada"; "Right to explanation"; "Direito à explicação"; "Right to

explanation" E "Automated Decision"; "Direito à explicação" E "decisão automatizada"; "right to explanation" E "Automated Decision" E "algorithmic opacity"; "Direito à explicação" E "decisão automatizada" E "opacidade algorítmica"; "AI" E "explainability" E "automated decision"; "IA" E "direito à explicação" E "decisão automatizada"; "AI" E "right to explanation" "IA" E "direito à explicação"; "IA" E "explicabilidade "AI" E "explainability"; "Automated Decision" E "algorithmic opacity"; "decisão automatizada" E "opacidade algorítmica"; "transparency" E "AI" E "automated decision"; "transparência" E "IA" E "decisão automatizada".

Tais termos foram extraídos dos diversos modos que a literatura jurídica brasileira e internacional se referem à opacidade algorítmica. A busca foi realizada por meio de expressões variadas, pois o uso de termos diversos gerou resultados também distintos. E, considerando a falta de um padrão nominal para a opacidade algorítmica e a inexistência de um artigo que preveja especificamente esse direito, o objetivo foi cercar, de forma mais abrangente possível, os casos relevantes para essa pesquisa.

Também houve aplicação de filtro temporal e de área de concentração. Para composição do *corpus* deste trabalho, a pesquisa se ateve aos resultados de artigos escritos a partir de 2014, haja vista que, apesar do tema objeto do estudo ser recente, observou-se que valiosos textos que fundamentam esta pesquisa surgiram pelo menos no referido ano. Os trabalhos também foram filtrados por sua concentração em “ciências sociais aplicadas”. O resultado encontrado totalizou 686 artigos. Nestes, foram aplicados três novos filtros: primeiramente, limitou-se os resultados às bases de dados da Scielo, Scopus e Web of Science; depois, exclui-se aqueles não foram revisados por pares e, posteriormente, exclui-se aqueles que não tinham relação direta com a temática tratada.

A análise do material obtido foi desenvolvida pautando-se nos seguintes questionamentos: (1) Como a GDPR e a LGPD definem o direito à explicação e quais são as suas principais características? (2) Quais são as semelhanças e diferenças nas abordagens da GDPR e da LGPD em relação à opacidade algorítmica? (3) De que maneira a aplicação do direito à explicação varia entre os contextos jurídico e cultural da União Europeia e do Brasil? (4) Quais implicações o direito à explicação tem para a proteção dos direitos dos indivíduos em diferentes jurisdições? (5) Quais são os principais desafios e dificuldades enfrentados na implementação do direito à explicação na GDPR e na LGPD? (6) Como a técnica do

direito comparado pode contribuir para uma melhor compreensão das práticas de proteção de dados entre a GDPR e a LGPD?

O material foi complementado com outros artigos, livros e teses/dissertações, à medida em que a pesquisadora teve contato.

2 FONTES HETERÔNOMAS SOBRE A PROTEÇÃO DE DADOS E INFORMAÇÕES DA PESSOA HUMANA

2.1 Delimitação conceitual de algoritmos, machine learning, inteligência artificial e caixa-preta

O ponto de partida do presente trabalho precisa ser a conceituação de termos essenciais à compreensão da discussão proposta. Sendo a presente pesquisa interdisciplinar, interseccionando o ramo do direito e da computação, existem diversos termos que são utilizados sem a devida acurácia, elevando, assim, a erro o receptor da informação.

Algoritmo, como explica Mittelstadt *et al* (2016, p. 2) pode possuir diversos sentidos a depender do que está sendo veiculado, como nos casos de ciência da computação, matemática e discursos públicos. Os autores concluem, assim, pela necessidade de dividir, de pronto, a definição formal e popular do termo em questão. O conceito formal de algoritmo enquanto construção matemática limita-se a uma sequência finita de regras (instruções e passos lógicos) bem definidas e não ambíguas, elaborada com o objetivo de resolver um problema ou realizar uma tarefa específica (Knuth, 1968).

Isto é, como lecionado por Araújo (2024), “os algoritmos funcionam segundo uma dinâmica de *inputs* e *outputs*, situação na qual programa-se uma saída para uma entrada pré-definida”. Nada obstante, este autor destaca que, sob a ótica da transparência algorítmica, “a intenção primária seria basicamente entender quais seriam as “entradas” predispostas e as “saídas” e, pois, a influência desse mecanismo na esfera de direitos dos indivíduos”. Entretanto, conforme destacado por Hoffman-Reim (2022), a partir desses mecanismos podem surgir também dois outros produtos: *impact* (relativo aos impactos diretos em relação ao indivíduo cujos dados estão sendo tratados) e *outcomes* (resultados gerais com repercussão para além desse indivíduo).

Doutro bordo, Mittelstadt *et al* (2016, p. 2) aponta que o uso leigo de 'algoritmo' abrange a sua implementação tecnológica e aplicação para uma tarefa específica. Isto é, um algoritmo totalmente configurado incorpora a estrutura matemática abstrata implementada num sistema para análise de tarefas num domínio específico. A configuração para uma tarefa ou conjunto de dados não altera a representação matemática ou a implementação do sistema, mas sim ajusta o

funcionamento do algoritmo a um caso ou problema específico. Em ambos os casos, o termo refere-se à implementação e interação de algoritmos num programa, software ou sistema de informação, e não à construção matemática em si[1].

A Inteligência Artificial é, segundo o English Oxford Living Dictionary Online, “a teoria e o desenvolvimento de sistemas computacionais capazes de realizar tarefas que normalmente requerem inteligência humana, como percepção visual, reconhecimento de fala, tomada de decisões e tradução entre idiomas”. Segundo a OECD Recommendation of the Council on Artificial Intelligence, 2019, trata-se de um sistema computacional que “executa ações normalmente associadas à inteligência humana, operando com diversos níveis de autonomia, a fim de realizar previsões, recomendações ou decisões que influenciam ambientes reais ou virtuais”. Nas palavras de McCarthy (1955), é a produção de máquinas inteligentes, “*especialmente programas de computador inteligentes*". A IA funciona, em geral, mediante a análise de um grande volume de dados e a identificação de padrões, utilizando métodos como, por exemplo, *machine learning* e *deep learning*.

A inteligência artificial (IA) é um ramo da ciência da computação que se concentra na elaboração de sistemas computacionais que raciocinem, tomem decisões e resolvam problemas, muitas vezes mimetizando competências humanas. De acordo com Nilsson (2009), é um conjunto de técnicas para a construção de máquinas inteligentes, capazes de resolver problemas que requerem inteligência humana. Como propõem Freitas e Freitas (2021), é um “sistema algorítmico adaptável, relativamente autônomo, emulatório da decisão humana”. John McCarthy (1989) definiu a inteligência artificial como A Inteligência Artificial (IA) é a ciência e a engenharia dedicadas à criação de máquinas inteligentes, em especial programas de computador com essa capacidade. Embora relacionada ao objetivo de utilizar computadores para compreender a inteligência humana, a IA não se restringe necessariamente aos métodos observáveis em sistemas biológicos.

---

[1] O uso leigo de 'algoritmo' também inclui a implementação da construção matemática numa tecnologia e uma aplicação da tecnologia configurada para uma tarefa específica. Um algoritmo totalmente configurado incorporará a estrutura matemática abstrata que foi implementada num sistema para análise de tarefas num domínio analítico específico. Tendo em conta esta clarificação, a configuração de um algoritmo para uma tarefa ou conjunto de dados específicos não altera a sua representação matemática subjacente ou a implementação do sistema; trata-se antes de um ajustamento adicional do funcionamento do algoritmo em relação a um caso ou problema específico. Em ambos os casos, o termo é utilizado não em referência ao algoritmo como uma construção matemática, mas sim à implementação e interação de um ou mais algoritmos num determinado programa, software ou sistema de informação. (Tradução livre).

Posteriormente, levando em conta o grande avanço que tal tecnologia vem sofrendo desde os anos 90, o Parlamento Europeu, vendo importância em legislar sobre o assunto, promulgou a Resolução de 20 de outubro de 2020, que, em seu artigo 3º, A, define o sistema de IA como:

> [...] um sistema baseado em software ou integrado em dispositivos de físicos e que apresenta um comportamento que simula inteligência, nomeadamente recolhendo e tratando dados, analisando e interpretando o seu ambiente e tomando medidas – com um determinado nível de autonomia – para atingir objetivos específicos

Já no Brasil, Novais e Freitas (2018, p. 14) definem a IA como “Os Sistemas Inteligentes têm como um dos seus propósitos, o de habilitar o computador para que este execute funções que são desempenhadas pelo ser humano fazendo uso do conhecimento e do raciocínio”. Na mesma linha Nunes (2023, p. 4) assevera que a inteligência artificial costuma ser conceituada como “a capacidade da máquina de interpretar dados de forma racional ou humana, tomando decisões autênticas com base em informações preexistentes”. O referido autor destaca, ainda, que segundo Morais (2021) o termo é empregado em sentido lato, referindo-se a programas computacionais aptos a reproduzir alguma habilidade humana. É certo, todavia, que Estudos comportamentais apontam que vieses cognitivos influenciam decisões humanas, levando pesquisadores a rejeitar a capacidade de "pensar humanamente" como traço definidor da IA. Por isso, segundo Morais (2021), parte da doutrina define IA como a capacidade da máquina de "agir racionalmente".

Ademais, importante notar que a inteligência artificial vai muito além da mera automação. Nesta última, a máquina é meramente pré-configurada por seres humanos, que tendem a operá-la, monitorá-la e reconfigurá-la continuamente. Na inteligência artificial, por sua vez, a máquina mantém um grau de independência maior, interpretando o ambiente, produzindo raciocínios completos e, não raro, remodelando por conta própria suas avaliações e decisões ao longo do tempo

Atualmente, no âmbito legislativo, o Projeto de Lei nº 21/2020, que se encontra no Senado Federal, que visa estabelecer normas para o desenvolvimento e aplicação da IA no país, conceitua o sistema de Inteligência artificial, em seu artigo 2º, I, como:

> o sistema baseado em processo computacional que pode, para um determinado conjunto de objetivos definidos pelo homem, fazer previsões e recomendações ou tomar decisões que influenciam ambientes reais ou virtuais[2].

Assim, é possível observar que não existe um consenso quanto ao conceito de IA[3]. Entretanto, com base nos autores mencionados, pode-se sustentar que a inteligência artificial compreende um sistema computacional que visa emular a inteligência humana, permitindo que, após coletar dados e informações, tire suas próprias conclusões para assim tomar medidas para chegar em determinado objetivo.

Por sua vez, *machine learning* (aprendizado de máquina, em português) é definido por Surden (2014) como uma técnica que permite que algoritmos mudem seu comportamento para melhorar seu desempenho à medida que recebem mais dados. Significa que, em vez de serem programados manualmente para realizar tarefas específicas, os algoritmos de aprendizado de máquina "aprendem" a partir da experiência, ajustando-se e aperfeiçoando-se com base nos dados que processam, de sorte a possibilitar resultados automatizados que se aproximam daqueles que teriam sido encontrados por uma pessoa em situação semelhante.

Também pode ser definido como ‘‘qualquer metodologia e conjunto de técnicas que possam empregar dados para criar novos padrões e conhecimentos e gerar modelos que possam ser usados para previsões eficazes sobre os dados” (Van Otterlo, 2013, tradução livre). Isto é, se refere ao processo de aprendizagem dos computadores, no qual, ao analisar grandes quantidades de dados, reconhecem padrões e apresentam soluções (*outputs*) para tarefas que não foram pré-programados.

Frisa-se que “programar” significa estabelecer um algoritmo. No caso do *machine learning*, é necessário que não exista um algoritmo estabelecido de forma explícita, isto é, que um programador não tenha escrito uma sequência de passos para realização de uma determinada tarefa.

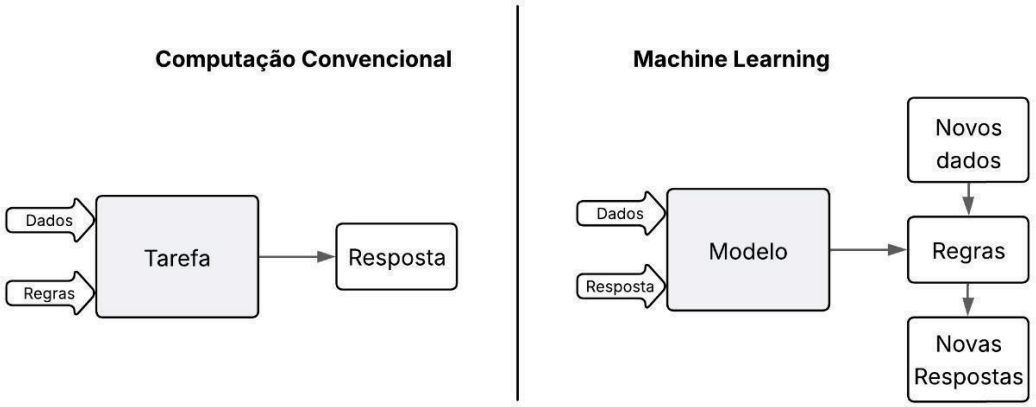
---

[2] BRASIL. Câmara dos Deputados. Projeto de Lei 21/2020. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Disponível em: < https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928> . Acesso em: 09.04.2025.

[3] IRIS - INSTITUTO DE REFERÊNCIA EM INTERNET E SOCIEDADE. Inteligência Artificial no Brasil: a Estratégia Brasileira de Inteligência Artificial (EBIA) e o Projeto de Lei nº 21/2020. Disponível em: https://irisbh.com.br/inteligencia-artificial-no-brasil-a-estrategia-brasileira-de-inteligencia-artificial-ebia-e-o-projeto-de-lei-no-21-2020/. Acesso em: 1 abr. 2025.

O aprendizado da máquina é, noutras palavras, a habilidade que a IA possui de acumular experiências “pessoais”, permitindo que ela atue de maneiras diferentes em situações idênticas (Medon, 2022). Ou seja, permite que a máquina aprenda com suas experiências externas, desenvolvendo-se e aprimorando-se para que não repita um erro já cometido (Frazão, 2019). Dessa forma, os sistemas de *machine learning* utilizam padrões e inferências a partir dos dados para melhorar seu desempenho em tarefas específicas ao longo do tempo, visando encontrar um resultado ótimo.

Figura 1. Fluxograma diferenciando computação convencional e *machine learning*



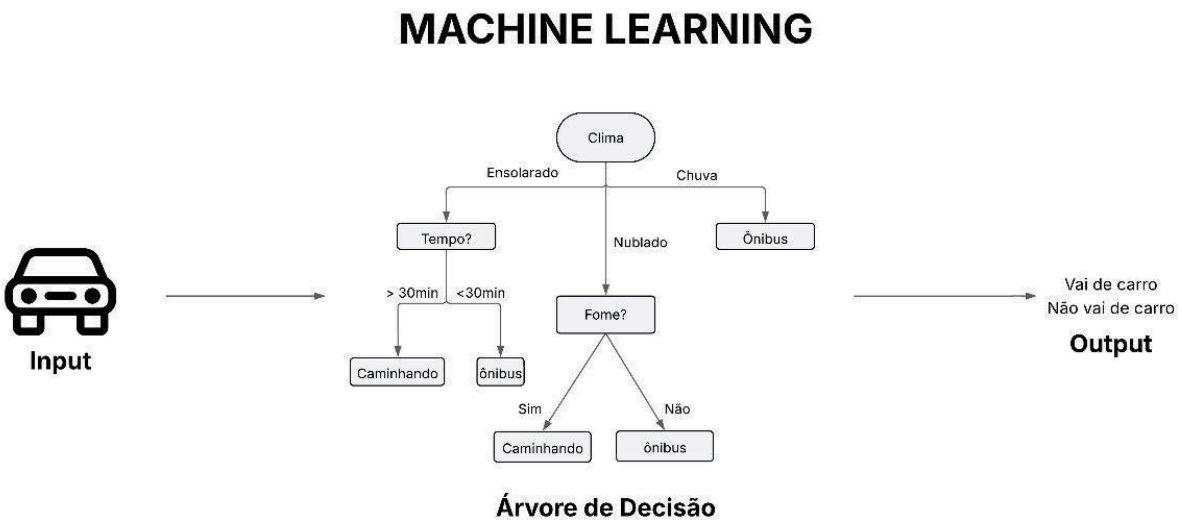
Fonte: autoria própria (2025)

O *deep learning*, por sua vez, é um tipo específico mais recente e avançado de aprendizado de máquina, utilizado para padrões complexos, com ainda maiores quantidades de dados, possuindo maiores camadas de redes neurais que o *machine learning*, ou seja, passando por um processo de aprendizagem mais profundo e, portanto, podendo realizar tarefas mais complexas. Nunes (2023, p. 5) complementa:

> valendo-se de múltiplas camadas para extrair progressivamente recursos de nível superior da entrada bruta. Neste sentido, enquanto o aprendizado de máquina usa algoritmos para analisar dados, aprender com esses dados e tomar decisões informadas com base no que aprendeu, o aprendizado profundo estrutura algoritmos em camadas para criar uma “rede neural artificial” que pode aprender e tomar decisões inteligentes por conta própria.

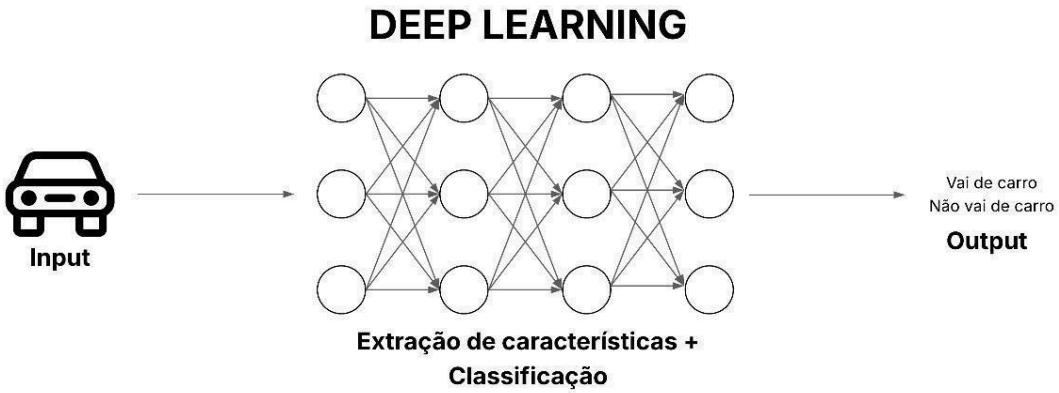
Vale observar a imagem abaixo, que ilustra a complexidade e multiplicidade dos processos de aprendizagem mencionados:

Figura 2. Fluxograma *machine learning*.



Fonte: autoria própria (2025)

Figura 3. Fluxograma *deep learning*.



Fonte: autoria própria (2025)

Essa subcategoria de algoritmos, chamada de *machine learning*, utiliza grandes volumes de dados para criar modelos preditivos e é encontrada em duas formas principais: aprendizado supervisionado e não supervisionado.

2.2 Breve perspectiva histórica da inteligência artificial

Com intuito de ambientar a temática de pesquisa, especialmente tendo em vista a recente criação do mecanismo objeto do estudo - a inteligência artificial - e dos novos ramos de implementação, de sorte que a legislação ainda não teve tempo hábil para acompanhar os avanços, entendeu-se relevante contextualizar, ainda que brevemente, a evolução histórica da inteligência artificial até o momento.

Para tanto, considerou-se o recorte geográfico dos Estados Unidos e da Europa. A escolha de tais regiões se deu em razão da centralidade tecnológica atual - os EUA lideram a pesquisa e a inovação no setor, com grandes empresas como Google, Microsoft e OpenAI - mas, especialmente, em razão da facilidade ao acesso de conteúdo histórico sobre a temática, que é vastamente mencionada nas referências bibliográficas consultadas. Não se desconsidera, entretanto, a importância do avanço tecnológico atinente à inteligência artificial nas demais localidades do mundo.

Pois bem. Em geral, a IA funciona mediante a análise de um grande volume de dados e a identificação de padrões, utilizando métodos como, por exemplo, *machine learning* e *deep learning*. Sobre as características da IA, Pires (2017, P. 242) aduz:

> Para definir a Inteligência Artificial, Russel e Norvig identificam duas principais características: uma associada como processo de raciocínio e motivação, e outra ligada ao comportamento. Nesse sentido, a principal diferença entre um algoritmo convencional e a IA está, justamente, na habilidade de acumular experiências próprias e extrair delas aprendizado, como um autodidata. Esse aprendizado, denominado de *machine learning*, permite à IA atuar de forma diferente em uma mesma situação, a depender da sua performance anterior - o que é muito similar à experiência humana.

Do ponto de vista econômico, Teixeira e Cheliga (2021) asseveram que a ascensão do mercantilismo na Europa, entre os séculos XV e XVIII, provocou profundas transformações no modelo econômico e social vigente. Até então, a produção, que era essencialmente local e baseada no trabalho manual e familiar, passou a focar na fabricação de bens destinados à venda e exportação.

Esse novo cenário abriu caminho para a Primeira Revolução Industrial, iniciada por volta de 1760, que trouxe uma mudança radical na forma como os bens eram produzidos. A introdução de máquinas movidas a vapor substituiu grande parte

da mão de obra manual nas fábricas, resultando em um expressivo aumento da produtividade e no fortalecimento do comércio internacional.

Os mesmos autores (Teixeira e Cheliga, 2021) apontam que a segunda Revolução Industrial, ocorrida entre 1850 e o início do século XX, foi marcada pelo aprimoramento das tecnologias existentes e pelo surgimento de novas invenções, como o avião. No século XX, por volta da década de 1940, teve início a Terceira Revolução Industrial, período caracterizado pelo rápido avanço da tecnologia e da ciência: a criação da internet, em 1969, e os progressos impulsionados pela Guerra Fria, especialmente na corrida espacial, modificaram profundamente a dinâmica da sociedade.

Em meados dessa revolução surge a inteligência artificial. Seu nascimento remonta a Alan Turing, cientista britânico, cujas pesquisas geraram a formalização do conceito de algoritmo, apresentado em um relatório publicado em Princeton, em 1936, no qual ele desenvolveu o modelo abstrato conhecido como "Máquina de Turing". Esse modelo representou uma concepção pioneira de computador, focando exclusivamente nos aspectos lógicos do seu funcionamento, incluindo memória, estados e transições.

Além disso, durante a Segunda Guerra Mundial, Turing desempenhou um papel crucial como chefe do departamento "Hut 8", encarregado da análise da criptografia da frota naval alemã. Nesse contexto histórico, ele liderou o projeto da "Bombe", um dispositivo eletromecânico, que visava a criptoanálise das mensagens cifradas, especialmente as codificadas pela máquina de cifra Enigma, amplamente utilizada pela Marinha Alemã para comunicação segura (Newman, 1955).

A Segunda Guerra Mundial (1939-1945), despida de qualquer ética e cuidado, garantiu o desenvolvimento tecnológico acelerado sob a perspectiva de melhorar o poder bélico. Nesse ínterim, destaca-se a própria ARPANET, os estudos anatômicos e psicológicos realizados nos campos de concentração e as invenções da área eletrônica. Chegado seu fim, os cientistas apresentaram suas descobertas no evento que ficou conhecido como Simpósio de Hixon, em 1948 (Barbosa, Bezerra, 2020) .

A criação dos neurônios artificiais, por Warren McCulloch e Walter Pitts, em 1943, é um dos primeiros marcos científicos bibliográficos da IA. O referido trabalho, baseando-se em conhecimentos de fisiologia e função dos neurônios, propôs a programação do computador de tal maneira que, havendo neurônios, estes poderiam

estar ligados ou desligados, fazendo conexões e ativações de acordo com a cadeia que se relacionasse, simulando, assim, o funcionamento do sistema nervoso humano. Marvin Minsky, oito anos após a publicação do referido artigo, construiu a primeira máquina de rede neural artificial, chamada de Stochastic Neural Analog Reinforcement Calculator (SNARC. Como explicado por Barbosa (2020, p. 94), "*Minsky testou os recursos de aprendizado fazendo a máquina navegar em um labirinto virtual, no intuito de verificar se a máquina seria capaz de aprender a sair sozinha do referido labirinto virtual*".

Em 1950 foi desenvolvido o Teste de Turing, o qual buscava verificar se uma máquina seria capaz de emular a comunicação escrita humana sem que o receptor notasse se tratar de um programa de computador (Russel, 2009).

Todavia, foi somente em 1956, durante a Conferência do Dartmouth College, em New Hampshire (EUA), que se utilizou pela primeira vez o termo "inteligência artificial" como uma nova área do conhecimento (Russel, 2009; Stanford). Na ocasião, foi apresentado o primeiro software de IA em execução. Escrito por Allen Newell, Herbert Simon e J. C. Shaw. O programa se chamava *Logic Theorist* e emulava como o cérebro humano processa informações matemáticas, além de testar a veracidade dos teoremas matemáticos.

Pouco tempo depois, em 1958, John McCarthy, que havia sido o organizador da Conferência supra mencionada, inventou uma linguagem de programação chamada *Lisp*, que foi utilizada como padrão por longo período. Em 1959, lançou o artigo científico intitulado "Programs with Common Sense", propondo a utilização da lógica como maneira de raciocínio para solucionar problemas, bem como para armazenar o conhecimento necessário para a referida solução. Os primeiros sistemas especialistas, que codificam o conhecimento de especialistas em determinada área, por sua vez, surgem apenas mais tarde. É o caso do Dendral, em meados de 1965, - sistema que prestava à análise da estrutura de química orgânica de uma molécula - e do MYCYN, em 1975, que recomendava antibióticos para infecções bacterianas (Negnevistky e Ensmenger, 2004).

No ano seguinte, o termo *machine learning* foi mencionado pela primeira vez, referindo-se a um sistema em que o computador tem a habilidade de aprender uma função sem que tenha sido diretamente programado para. Isto é, a partir de uma

introdução de dados em um algoritmo a máquina passa a aprender e executar tarefas automaticamente. Foi em 1978 que Tom Mitchell sugeriu o primeiro algoritmo de aprendizagem de máquina, o qual foi utilizado por longos anos na área. Nessa toada, destaca-se que, somente nos anos 80 é que a inteligência artificial passou a ter relevância, visto que, nas palavras de Negnevistky e Ensmenger (2004):

> [...] até meados da década de 1980, havia carência de dados, capacidade de processamento limitada dos recursos disponíveis e desenvolvimento precário de softwares para que a Inteligência Artificial pudesse atingir todas as suas possibilidades e potencialidades.

O desenvolvimento da inteligência artificial se intensificou nos anos 2000, abrangendo carros autônomos, algoritmos de sugestão e armamento bélico autônomo. Em 2008, o processamento de linguagem natural impulsionou assistentes virtuais como Siri (2011), Alexa, Cortana e Google Assistente. Em 2012, o *deep learning* (um *machine learning* mais complexo) foi usado pela Google para reconhecimento de gatos em vídeos do YouTube.

Nesse contexto, a discussão sobre ética na IA ganhou força, exigindo adaptações sociais e normas regulamentadoras para lidar com os desafios tecnológicos e os direitos em risco. O processo legislativo lento, evidenciado pela demora na regulamentação da internet (como o Marco Civil da Internet de 2014 e o GDPR de 2018), cria lacunas regulatórias que, no caso da IA e sistemas automatizados de decisão, podem afetar direitos fundamentais (como privacidade, informação e não discriminação), demandando a adoção rápida e eficaz de mecanismos de garantia.

2.3 Opacidade algorítmica e sua interface com o direito à informação do cidadão à luz dos direitos fundamentais

Sob argumento de que os estudantes não estavam aprendendo o suficiente porque os professores não faziam um bom trabalho, em 2009, o prefeito de Washington D.C., juntamente com a recém nomeada reitora das escolas desta cidade, pretendendo melhorar o desempenho das escolas municipais,

implementaram um sistema avaliativo da performance dos professores. O desempenho do profissional passou a ser avaliado a partir de uma ferramenta chamada IMPACT, criada com o auxílio da Mathematica Policy Research, de Princeton. Ao final do ano letivo, aqueles classificados entre os 2% piores foram demitidos. O novo sistema de pontuação, conhecido como modelagem de valor agregado, serviria para medir a eficácia dos professores ao ensinar matemática e habilidades linguísticas, sendo que a pontuação gerada pelo algoritmo comporia metade da avaliação final.

O caso é relatado por Cathy O'Neil (2021), sobre o caso de Sarah Wysocki, professora do quinto ano, que, apesar de aparentemente não precisar se preocupar com a implementação do plano mencionado - haja vista frequentemente receber excelentes avaliações de seu diretor e dos pais dos alunos -, foi classificada com uma performance insuficiente e dispensada. Significa dizer que a pontuação auferida pelo algoritmo da ferramenta implementada prevaleceu sobre os comentários positivos dos gestores e da comunidade escolar. Naturalmente, Wysocki procurou entender a baixa pontuação a ela atribuída pelo algoritmo. Todavia, como O'Neil (2021) relata, não havia alguém que pudesse explicar[4].

Nesse mesmo sentido, Yuval Harari (2016, p. 43) reflete

> [...] algoritmos realmente importantes – como algoritmos de busca do Google- são desenvolvidos por equipes enormes. Cada membro somente entende uma parte do quebra-cabeça e ninguém entende o algoritmo como um todo. Além disso, com o surgimento da aprendizagem automática e das redes neurais artificiais, mais e mais algoritmos se desenvolvem independentemente, aprimorando a si mesmos e aprendendo com os próprios erros. Eles analisam quantidades astronômica de dados, que nenhum humano é capaz de abranger, e aprendem a reconhecer padrões e a adotar estratégias que escapam à mente humana. O algoritmo-semente pode de início ser desenvolvido por humanos, mas ele cresce, segue o próprio caminho e vai onde humanos nunca foram antes - até onde nenhum humano pode segui-lo.

[4] As sentenças das ADMs caem como mandamentos dos deuses algorítmicos. O modelo em si é uma caixa preta, cujo conteúdo é segredo corporativo ferozmente protegido. Isso permite que consultorias como a Mathematica cobrem mais, mas serve também a outro propósito: se as pessoas sendo avaliadas são mantidas no escuro, por esse modo de pensar, haverá menos chance de tentarem burlar o sistema. Em vez disso, terão simplesmente de trabalhar mais duro, seguir as regras e rezar para que o modelo registre e aprecie seus esforços. Mas se os detalhes são escondidos, também é mais difícil questionar ou contestar os resultados. (O'Neil, 2021, p. 11).

Os algoritmos operam, basicamente, com parâmetros de classificação. Isto é, os processos de tomada de decisão automatizada são pautados em processos matemáticos de classificação, realizados através de “árvores de decisão”, que resultam em um comando automatizado imprimido em forma de oferta de um produto, serviço ou assunto (Araújo, 2024). Significa dizer que, em se repetindo o processo diversas vezes, são criadas diversas e inúmeras árvores de decisão, formando as “florestas aleatórias”. A partir disso, dada uma condição “se”, tem-se uma decisão da máquina (Shane, 2022 *in* Araújo, 2024).

O conceito de "black box" (caixa-preta) em relação a algoritmos de inteligência artificial refere-se a sistemas cujos processos internos (a formação das árvores de decisão) são incompreensíveis ou não transparentes, mesmo que se possa observar os *inputs* (dados de entrada) e *outputs* (resultados) desses sistemas. Assim, a falta de entendimento sobre como as saídas são produzidas a partir das entradas é o que caracteriza uma "caixa-preta". Significa dizer que seu funcionamento interno não é visível, conhecido ou entendível - por alguma razão é misterioso. Neste sentido, o matemático francês Cédric Villani (2018, p. 70) destaca que “esta falta de conhecimento se deve principalmente hoje à mudança de paradigma introduzida pelo advento da aprendizagem, especialmente a aprendizagem profunda”.

Aliás, nesse quesito, importa mencionar que a diferenciação entre aprendizado de máquina supervisionado e não supervisionado reside no tipo dos dados utilizados no treinamento do modelo. Como explicado por Araújo (2024)[5], enquanto o aprendizado supervisionado utiliza-se de dados rotulados, em que cada entrada possui uma saída correspondente, o aprendizado não supervisionado não trabalha com rotulagem de dados, buscando padrões e estruturas por conta própria.

O desconhecimento acerca dos processos decisórios do código gera, por consequência, opacidade do algoritmo, além de preocupações sobre a

[5] Quando há supervisão, alguns dados são rotulados de forma que a resposta obtida possa ser melhor medida, obtendo-se melhor precisão acerca do aprendizado. Quando não há supervisão, esses rótulos inexistem, e a própria máquina deve por conta própria reconhecer os padrões. Portanto, quando se fala em opacidade, há de se considerar, sobretudo, a qualidade dos dados com os quais trabalha o processo automatizado, se há ou não supervisão sobre o aprendizado da máquina e como este se dá. (Araújo, 2024, p. 277).

responsabilidade, a ética e a confiança nos sistemas que utilizam essa tecnologia. Ocorre que a transparência é altamente associada à conferência de confiabilidade do modelo de Inteligência Artificial, de sorte que sua ausência, seja por qual for o motivo, além de apresentar possíveis riscos à sociedade e à democracia, também engendra desconfianças. Isto é, a opacidade algorítmica contribui, instiga ou, no mínimo, corrobora na manutenção da falta de neutralidade dos algorítmicos. Nesse sentido, Nunes (2023, p. 11) afirma que “a promoção da inteligência artificial explicável pode aumentar a confiabilidade e a legitimidade dos modelos algorítmicos”, de sorte a explicitar a importância de sanar problemas relativos à falta de explicabilidade algorítmica.

2.3.1 Direito de explicação e opacidade algorítmica: características e classificações

Segundo o dicionário brasileiro Michaelis de língua portuguesa, “opacidade” é um substantivo feminino utilizado para expressar (1) a qualidade ou estado de algo que é opaco, sem transparência; (2) a propriedade de bloquear a passagem de luminosidade e, ainda, de forma figurativa para aduzir (3) sombra ou (4) ausência de clareza, de transparência, de precisão.

Em transferência, a opacidade algorítmica é descrita como a dificuldade ou impossibilidade de compreender claramente o funcionamento interno e as decisões tomadas por algoritmos, em especial os de *machine learning* ou *deep learning*. Por isso, Henry Surden (2014) utiliza o termo "opacidade tecnológica" para descrever “qualquer momento que um sistema tecnológico se engaje em comportamentos que, embora apropriados, podem ser difíceis de entender ou prever, do ponto de vista humano”. Em complementação, Mittelstadt *et al* (2016, p. 3) elucidam que os algoritmos classificam novas entradas (como imagens) definindo regras. Isso pode ser feito com entradas rotuladas (aprendizado supervisionado) ou o próprio algoritmo define os modelos (aprendizado não supervisionado). O operador humano não precisa entender a lógica dessas regras de decisão. Essa capacidade de aprendizado concede aos algoritmos um grau de autonomia cujo impacto é incerto,

pois as tarefas são difíceis de prever ou explicar, inibindo a identificação e correção de desafios éticos[6].

Burrel (2016, p. 3), neste mesmo sentido, assevera "*Eu indico a opacidade em termos de legibilidade do código*". Sardinha (2022, p. 19[7]) explica a opacidade como a impermeabilidade à luz: um objeto opaco, como a porta de um carro, é intransponível à visão, ao contrário de uma janela; uma janela pintada é parcialmente opaca/transparente, mostrando a relação inversa entre opacidade e transparência.

A opacidade algorítmica é abordada por Burrel (2016, p. 1) como um problema em mecanismos socialmente relevantes de classificação e ranqueamento, como filtros de spam, detecção de fraude, mecanismos de busca, tendências de notícias, segmentação de mercado, publicidade, qualificação de seguros/empréstimos e pontuação de crédito. Estes são exemplos de mecanismos de classificação aos quais nossos dados pessoais e de rastreamento estão sujeitos diariamente em sociedades capitalistas avançadas e conectadas em rede. Tais mecanismos frequentemente dependem de algoritmos computacionais, incluindo algoritmos de aprendizado de máquina.

---

[6] O trabalho do algoritmo envolve a colocação de novas entradas em um modelo ou estrutura de classificação. As tecnologias de reconhecimento de imagens, por exemplo, podem decidir quais tipos de objetos aparecem em uma imagem. O algoritmo "aprende" definindo regras para determinar como as novas entradas serão classificadas. O modelo pode ser ensinado ao algoritmo por meio de entradas rotuladas manualmente (aprendizado supervisionado); em outros casos, o próprio algoritmo define os modelos mais adequados para dar sentido a um conjunto de entradas (aprendizado não supervisionado) (Schermer, 2011; Van Otterlo, 2013). Em ambos os casos, o algoritmo define regras de tomada de decisão para lidar com novas entradas. De forma crítica, o operador humano não precisa entender a lógica das regras de tomada de decisão produzidas pelo algoritmo (Matthias, 2004: 179). Como essa explicação sugere, as capacidades de aprendizado concedem aos algoritmos algum grau de autonomia. O impacto dessa autonomia deve permanecer incerto até certo ponto. Como resultado, as tarefas executadas pelo aprendizado de máquina são difíceis de prever antecipadamente (como uma nova entrada será tratada) ou de explicar posteriormente (como uma determinada decisão foi tomada). A incerteza pode, portanto, inibir a identificação e a correção de desafios éticos no projeto e na operação de algoritmos. (Tradução livre).

[7] Para entender a opacidade, é importante entender o que "opaco" significa. Um objeto opaco é completamente impermeável à luz, o que significa que você não pode vê-lo. Por exemplo, uma porta de carro é completamente opaca. A janela acima da porta, no entanto, não é opaca, pois você pode ver através dela. Se a janela estiver pintada, é parcialmente opaca e parcialmente transparente. Quanto menos transparente a janela, maior a sua opacidade. Em outras palavras, transparência e opacidade são inversamente relacionadas (OPACIDADE In: DICIONÁRIO TechLib,). Um exemplo de opacidade é dado por Juan Pablo Marín Díaz e Juliana Galvis), em um artigo no Goethe Institut que diz que: "se compararmos frases neutras como 'vamos comer comida italiana', ou 'vamos comer comida mexicana', em princípio as duas frases devem ter o mesmo nível de positividade. Porém, como os algoritmos são treinados com dados que podem ser tendenciosos, acabam dando uma pontuação mais positiva ao italiano que ao mexicano". É a chamada opacidade algorítmica (Díaz; Galvis).

A análise do conceito e seus desdobramentos é especialmente relevante no contexto da Inteligência Artificial, haja vista que a previsibilidade e a explicabilidade dos algoritmos são aspectos cruciais para aferir confiança e aceitação pública. Dessa forma, a falta de clareza sobre os processos decisórios algorítmicos levanta sérias preocupações éticas e legais, que perpassam questões sobre equidade de justiça social, sobretudo em setores mais sensíveis como justiça criminal, saúde e finanças, onde as decisões algorítmicas podem ter impactos profundos e duradouros na vida das pessoas.

Mittelstadt (2016, p. 2) alerta: "*a lacuna resultante entre o projeto e a operação de algoritmos e nossa compreensão de suas implicações éticas pode ter consequências graves que afetam indivíduos, grupos e segmentos inteiros de uma sociedade*". E complementa:

> No entanto, os algoritmos são eticamente desafiadores não apenas por causa da escala de análise e da complexidade da tomada de decisões. A incerteza e a opacidade do trabalho que está sendo feito pelos algoritmos e seu impacto também são cada vez mais problemáticos. Tradicionalmente, os algoritmos exigem que as regras e os pesos para a tomada de decisões sejam definidos individualmente e programados "à mão". Embora isso ainda seja verdade em muitos casos (o algoritmo PageRank do Google é um exemplo de destaque), os algoritmos dependem cada vez mais de capacidades de aprendizado (Tutt, 2016).

Frank Pasquale (2015, p. 6/7) indica que a opacidade algorítmica resulta de três categorias de sigilo, cada qual com motivações distintas: sigilo real (*real secrecy*), que se destina à proteção da privacidade e da intimidade, mantendo certos fatos em reserva; sigilo legal (*legal secrecy*), que corresponde à imposição legal ou contratual de preservar o sigilo de determinadas informações; e ofuscação (*obfuscation*), que consiste em uma estratégia de sigilo que visa dificultar a compreensão e desestimular o interesse em obter acesso a certas informações

Outrossim, na busca por distinguir as formas de opacidade, Burrel (2016) descreveu três formas: (i) opacidade como segredo corporativo ou estatal intencional, (ii) opacidade como analfabetismo técnico e (iii) uma opacidade que surge das características dos algoritmos de aprendizado de máquina e da escala necessária para aplicá-los de forma útil (complexidade das operações internas dos sistemas de IA).

Isto é, a opacidade algorítmica representa uma barreira significativa à transparência e à compreensão dos sistemas de decisão automatizados, sendo resultado tanto de decisões corporativas quanto de limitações técnicas e complexidades estruturais dos próprios algoritmos. Sua análise é essencial para enfrentar os desafios éticos e sociais da inteligência artificial em contextos sensíveis como saúde, justiça e finanças.

2.3.1.1 Óbice à explicabilidade das decisões algorítmicas por sigilo industrial

Observa Sandving (*et al*., 2014, p. 9) que esse “jogo de gato e rato” torna totalmente improvável que a maioria dos algoritmos seja (ou necessariamente deva ser) divulgada ao público geral. Dito isso, uma alternativa óbvia seria que os algoritmos privados e fechados fossem colocados em código aberto, especialmente porque modelos de negócio bem-sucedidos surgiram a partir do movimento de código aberto.

Trata-se de opacidade proposital, atrelada à autoproteção das empresas que pretendem manter seus segredos comerciais e, principalmente, sua vantagem competitiva de mercado. Burrel (2016, p. 3) pontua que

> No entanto, não se trata apenas de um mecanismo de busca competindo com outro para manter seu 'molho secreto' em segredo. Também acontece que as plataformas e os aplicativos dominantes, especialmente aqueles que usam algoritmos para classificação, recomendação, tendências e filtragem, atraem aqueles que querem 'jogá-los' como parte de estratégias para garantir a atenção do público em geral. (Tradução livre).

Por outro lado, Pasquale propõe, em uma análise mais cética, que a atual extensão da opacidade algorítmica é um produto de regulamentações fracas e atrasadas. Em seu livro *The black box society: the secret algorithms that control money and information,* este autor argumenta a existência de uma situação de opacidade enquanto “incompreensibilidade remediável”, em que o adversário é a própria regulamentação: “e se os financistas mantiverem seus atos opacos de propósito, precisamente para evitar ou confundir a regulamentação?” (Pasquale, 2015, p. 2).

Ou seja, a opacidade algorítmica, segundo Pasquale (2015) poderia ser atribuída à autoproteção intencional das corporações em nome de vantagem

competitiva, mas isso também poderia ser um disfarce para uma nova forma de ocular regulamentações contornadas, a manipulação de consumidores e/ou padrões de discriminação.

Posto isto, a opacidade por sigilo intencional evidencia o conflito entre interesses corporativos e a transparência algorítmica, com empresas resguardando seus algoritmos como ativos estratégicos. Tal postura, ainda que legítima do ponto de vista competitivo, compromete a possibilidade de auditoria pública e levanta preocupações quanto à manipulação, discriminação algorítmica e insuficiência regulatória.

2.3.1.2 Óbice à explicabilidade das decisões algorítmicas por barreira técnica (analfabetismo técnico)

A opacidade algorítmica também pode decorrer da falta de conhecimento da maior parcela da população para escrever e ler os códigos, que são escritos em linguagem de programação específicas, como C ou Python. Isto é, como apontado por Burrel (p. 4), "*um bom código tem dupla função*." A dupla função se refere à capacidade de interpretação do código por humanos (o programador) e pelo dispositivo computacional (Mateas e Montfort, 2005). Isto é,

> A arte e o 'ofício' da programação consistem, em parte, em gerenciar essa função mediadora e implicam algumas 'melhores práticas' bem conhecidas, como a escolha de nomes de variáveis sensatos, incluindo 'comentários' (comunicação unilateral com programadores humanos omitida quando o código é compilado para a máquina) e a escolha da formulação de código mais simples, se todas as coisas forem iguais. (BURREL, p. 4). (Tradução livre).

Nesse aspecto, o Burrel (p. 4) aduz que, para combater tal forma de opacidade, seria necessário empregar amplos esforços educacionais para tornar o público ser adequadamente informado sobre esses mecanismos e seus impactos, colocando a população em posição de avaliador e crítico direto.

Assim é que a opacidade derivada do analfabetismo técnico revela um distanciamento entre a população e o entendimento dos algoritmos, reforçando a necessidade de educação digital e letramento computacional. Sem tais esforços, a

sociedade permanece vulnerável a decisões algorítmicas não compreendidas, comprometendo a autonomia crítica dos cidadãos.

2.3.1.3 Óbice à explicabilidade das decisões algorítmicas por complexidade do sistema

A opacidade técnica por complexidade algorítmica, por fim, relaciona-se à capacidade plena de entender um algoritmo em ação, operando com dados. Isto é, supera as dificuldades atreladas à quantidade de linhas e páginas do código, à multiplicidade de interligações entre módulos e sub-rotinas, entre outros. Isso porque, dada a massiva quantidade de dados utilizados para treinamento dos modelos algoritmos utilizados na era do Big Data, a interação entre estes e o código (por mais que seja escrito de forma transparente e clara), gera complexidade - e opacidade.

Nesse aspecto, Nunes (2023, p. 8) explica que esse tipo de opacidade decorre da complexidade das operações internas dos sistemas que extrapola a mera dificuldade de acesso ou leitura do código, culminando na incapacidade de entendimento da análise preditiva realizada, “porque, além de processar um volume incomensurável de dados, o modelo algorítmico frequentemente (re)adapta sua lógica de decisão interna, à medida que “aprende” com os dados de treinamento”.

É certo, dessarte, que a discussão acerca da opacidade algorítmica e do conceito de *black box* revela os múltiplos níveis de complexidade e desafio ético, técnico e regulatório que envolvem os sistemas de inteligência artificial. É possível observar que a opacidade não se limita a uma simples falta de transparência técnica, mas abrange camadas mais profundas, comprometendo a previsibilidade e a explicabilidade dos modelos - aspectos fundamentais para assegurar confiança, legitimidade e justiça nos contextos sociais em que são aplicados (como saúde, justiça criminal e finanças).

3 O DIREITO DE EXPLICAÇÃO E SUA COMPREENSÃO TEÓRICA

Retornando ao caso dos professores de Washington, que foram avaliados e - eventualmente - demitidos com base no sistema de avaliação que utilizava o algoritmo IMPACT, muitos reclamaram das pontuações arbitrárias e exigiram os detalhes que as compunham. A resposta obtida nunca foi suficiente: "*é um algoritmo, disseram-lhes. É muito complexo. Isso desencorajou muitos a fazer mais pressão*" (O'Neil, 2021, p. 11). Naturalmente, o caráter técnico da linguagem matemática e estatística utilizado em sistemas automatizados de decisão frequentemente acaba por intimidar e desencorajar questionamentos mais incisivos, especialmente entre aqueles que não possuem conhecimento específico suficiente. Nesse contexto, o questionamento feito pela professora Sarah Bax ao gestor distrital responsável diz respeito a como seria possível avaliar profissionais com base em critérios cuja lógica nem mesmo os gestores sabem explicar (O'Neil, 2021). É justamente a partir deste questionamento que surge a questão da explicabilidade algorítmica e do direito à explicação.

Ainda segundo O'Neil (2021), o problema relativo à opacidade algorítmica está atrelado à ausência de transparência dos modelos de inteligência artificial, que tornam seu funcionamento invisível a todos - excetuando-se somente alguns modeladores -, remontando ao conceito de opacidade algorítmica e *black box*, já abordados neste trabalho. Considerando este contexto, o presente capítulo tem como pano de fundo questões demasiado caras e sensíveis que, dado o escopo proposto no presente trabalho, não serão abordadas em profundidade, mas que merecem ser brevemente destacadas.

A explicabilidade algorítmica enquanto direito resvala diretamente no enviesamento algorítmico, atrelando-se, por sua vez, à autodeterminação informativa, igualdade e não discriminação, direito do consumidor e mesmo controle social. Assim, visando assegurar direitos fundamentais e interesses sociais, numa tentativa de equilibrar interesses privados e públicos, surge o direito de explicação, ora abordado.

A característica da explicabilidade algorítmica é a característica ou funcionalidade de um modelo algorítmico que permite compreender os detalhes e motivos do seu funcionamento, ou seja, explicar ou ao menos expor de maneira simplificada, seu processo decisório (Molnar, 2021). O direito de explicação,

refere-se, assim, ao direito do cidadão de receber informações que o possibilite compreender o porquê de determinada decisão a seu respeito ter sido tomada. Sua aplicação se dá no âmbito de decisões automatizadas, as quais são realizadas por inteligência artificial.

Beck *et al*. (2022) abordam o desafio de garantir direitos fundamentais diante de tecnologias de difícil acesso e compreensão, especialmente ante à expansão das decisões automatizadas em diversas áreas da sociedade, alertando para a necessidade de situar parâmetros concretos de aplicação da regulação sob o enfoque dos reais impactos e refletindo o direito de explicação como forma de prevenção e mitigação.

> No campo das decisões automatizadas, é preciso compreender o que a lei pretende defender com "direito à explicação", já que a tecnologia possui arquitetura inacessível para a maioria dos indivíduos. Dada a ampliação das decisões automatizadas nas mais diversas instituições e no cotidiano das pessoas, deve-se discutir os impactos reais que uma decisão automatizada pode gerar na vida de um indivíduo e problematizar as formas de exercício do direito à explicação. Não se trata, portanto, de alinhar os princípios da lei, dos quais decorrem dois direitos fundamentais, da transparência e da informação, mas situar parâmetros concretos de aplicação da regulação no Brasil. Se a tecnologia é um sistema opaco por sua arquitetura, como prevenir e mitigar os riscos dessa tecnologia? (Beck *et al*, 2022, p. 516)

É certo, todavia, que os parâmetros acerca da execução e aplicação do referido direito ainda são tênues. Segundo Wachter (2018), as informações devem ser prestadas de forma suficiente e inteligíveis, permitindo a plena compreensão acerca da lógica, da forma e dos critérios adotados para o tratamento dos dados pessoais, de forma que, consequentemente, seja possível prever os impactos da decisão automatizada.

A discussão sobre os desafios relacionados à falta de transparência em sistemas algorítmicos notadamente está em voga, também buscando apontar caminhos para lidar com esse problema. Segundo Lipton (2018), tornar os modelos de aprendizado de máquina mais compreensíveis é fundamental para fomentar tanto a confiança quanto a responsabilidade no uso dessas tecnologias. O autor sugere diferentes estratégias para alcançar essa clareza, como tornar os modelos mais simples, permitir a visualização de seus componentes internos e fornecer explicações claras sobre suas decisões. Já Rudin (2019) defende que o ideal é

adotar, desde o início, modelos que sejam naturalmente compreensíveis para os usuários, evitando a necessidade de justificar posteriormente os resultados de sistemas complexos e de difícil interpretação.

O problema majoritário que atualmente se encontra na discussão sobre direito à explicação tange os parâmetros para sua efetivação, ou seja, quais informações devem ser disponibilizadas ao usuário solicitante, bem como de qual forma deve ser realizada essa explicação. Isso porque, considerando que se trata de uma explicação de código algorítmico, mais comumente relacionada a *machine learning* e, especificamente, a *deep learning*, essas informações geralmente não são acessíveis ao entendimento de pessoas leigas na área.

Até o presente momento inexiste regulamentação para tanto no Brasil, não havendo sequer posicionamento da Autoridade Nacional de Proteção de Dados (ANPD), cuja atuação é prevista nos artigos 55-A e 55-L (incluídos pela Lei nº 13.853/2019) também deve conferir efetividade ao direito à explicação.

Ante toda discussão apresentada, a análise do direito de explicação revela sua centralidade no debate contemporâneo sobre governança algorítmica e proteção de direitos fundamentais. A crescente presença de decisões automatizadas em esferas sensíveis da vida social exige que os cidadãos possam acessar, compreender e eventualmente contestar os critérios adotados por sistemas algorítmicos. No entanto, a ausência de parâmetros regulatórios claros, aliada à complexidade técnica desses sistemas, impõe desafios significativos à efetivação desse direito. Nesse cenário, iniciativas que promovam a explicabilidade desde a concepção dos modelos, aliadas ao fortalecimento de políticas públicas e regulações específicas, são essenciais para garantir que a tecnologia atue em consonância com os princípios democráticos de transparência, justiça e responsabilidade.

3.1 Delimitação conceitual e características do direito de explicação sob a perspectiva da Lei Geral de Proteção de Dados (L. 13.709/2018) e do Regulamento Europeu de proteção de dados (2016/679)

No Brasil, embora a Lei nº 13.709/2018 - a Lei Geral de Proteção de Dados (LGPD) - tenha sido promulgada em 2018, apenas entrou em vigor em 2020. A referida legislação, no decorrer de seus 65 artigos, buscou disciplinar o tratamento

de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado[8].

Dentre os direitos e obrigações definidos pela LGPD, consta o artigo 20, contendo dois importantes direitos, quais sejam: o direito à revisão de decisões totalmente automatizadas e o direito à explicação.

Por outro lado, como se sabe, todo direito se contrapõe à uma obrigação. Assim, em contrapartida ao direito à explicação, tem-se o parágrafo 1º do referido artigo 20, que determina ser dever do controlador fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.

> Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.
> § 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.
> § 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.

É certo que, desde a sua promulgação, a redação deste artigo sofreu alterações. A versão original do *caput* estipulava que o titular dos dados poderia requerer revisão por “pessoa natural”, mas a Medida Provisória nº 869/2018 (convertida na Lei nº 13.853/2019) removeu a expressão “pessoa natural”.

Essa repressão, como explica Mulholland, 2019, implicou, para alguns, a eliminação da possibilidade de revisão, por indivíduos, de decisões automatizadas. Como resultado, tais decisões seriam revistas apenas por meio de outras decisões automatizadas, e não por um ser humano. Por outro lado, pode-se também compreender que a alteração simplesmente omite os detalhes sobre as condições de revisão na LGPD, não proibindo explicitamente a revisão por pessoas físicas,

---

[8] Art. 1º Esta Lei dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

mas tão somente permitindo que solicitações de revisão de decisões automatizadas sejam processadas por sistemas automatizados, em vez de humanos.

Outra questão que surge sobre o texto legal diz respeito ao conceito de *decisão automatizada*. Em que pese o preciosismo da LGPD em trazer em seu texto definição de terminologias técnicas, não há qualquer indicação do que seria, para seus fins, uma decisão automatizada.

Preocupado que essa lacuna fosse capaz de comprometer a proteção pretendida pelo artigo, haja vista haver diversas formas de se tomar decisões automatizadas e, tendo em vista a própria Justificação do PL 4.496/2016, de autoria do Senador Styvenson Valentim (PODEMOS/RN) afirmar que "*algumas* [decisões automatizadas] *são facilmente compreensíveis, como as baseadas em regras ou em algoritmos pré-definidos. Outras, mais sofisticadas e geralmente menos explícitas, aplicam técnicas de aprendizado de máquina (machine learning) ou de inteligência artificial*"), o referido Projeto de Lei, pretende acrescer no artigo 5º da LGPD o inciso XX, definindo decisão automatizada como:

> decisão automatizada: processo de escolha, de classificação, de aprovação ou rejeição, de atribuição de nota, medida, pontuação ou escore, de cálculo de risco ou de probabilidade, ou outro semelhante, realizado pelo tratamento de dados pessoais utilizando regras, cálculos, instruções, algoritmos, análises estatísticas, inteligência artificial, aprendizado de máquina, ou outra técnica computacional.

Atualmente o projeto ainda está em trâmite, não havendo qualquer modificação do texto legal para inclusão do conceito de decisão automatizada.

De toda sorte, para o presente trabalho, fixa-se como ponto de discussão as questões relacionadas ao direito de explicação, ou seja, ao parágrafo primeiro do artigo 20, sem, contudo, descartar a inegável relevância do direito à revisão das decisões totalmente automatizadas.

Desde a primeira lei criada no Brasil com o intuito de proteger seus usuários e passar a regular tal ambiente, o Marco Civil da Internet em 2010, veio em seguida um conjunto regulatório que ficou conhecido como constitucionalismo digital, do qual fazem parte regulações como: a Lei de Acesso à Informação (2016), a Lei Geral da Proteção de Dados (2020), a Lei de Governo Digital (2021), o Marco Legal das Startups (2021). Ainda assim, são notáveis a demora e a deficiência legislativa na

regulamentação das inúmeras consequências da sociedade digital. Sua rápida evolução, contribui para as diversas lacunas presentes na regulação do ciberespaço.

O direito à explicação foi mencionado pela primeira vez no Brasil na Lei nº 12.414/2011 (Lei de Cadastro Positivo), que inclui, entre os direitos do cadastrado, o direito de solicitar ao consulente a revisão de decisão realizada exclusivamente por meios automatizados (art. 5º, VI). A sua inserção no microssistema de proteção de dados pessoais ampliou o campo da aplicabilidade.

Em breve análise axiológica do artigo 20 em comento, da LGPD, evidencia-se a intenção de resguardar os indivíduos contra vulnerabilidades, incluindo aquelas resultantes de decisões arbitrárias das quais não haja possibilidade de apelação ou compreensão dos fundamentos que as embasaram.

Ainda sobre a interpretação da LGPD, importa ressaltar que o artigo 6º da mesma lei obriga que o tratamento de dados observe a boa-fé e os princípios contidos em seus incisos, dentre os quais destaca-se à presente discussão o *princípio da transparência*, do qual o artigo 20 emerge como um desdobramento.

> Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:
> VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;

O relatório substitutivo ao Projeto de Lei 21/2020, elaborado pela Comissão de Juristas do Senado e publicado em dezembro de 2022[9], previu a transparência como medida de governança para aplicações de Inteligência Artificial (IA), seja com relação à interface humano-máquina ou ao uso, clareza e poder informativo das medidas governamentais adotadas no desenvolvimento dos projetos.

Assim, o direito à explicação, no Brasil, é regulamentado especialmente pela Lei Geral de Proteção de Dados, contando também com disposições complementares em leis setoriais, como a Lei do Cadastro Positivo e o Código de

---

[9] BRASIL. Congresso Nacional. Senado Federal. Projeto de Lei n° 21, de 2020: Relatório final da Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial. Diário do Senado Federal, ano 77, n. 204, suplemento n. B, p. 3-915, 9 dez. 2022. Disponível em: https://legis.senado.leg.br/diarios/BuscaDiario?tipDiario=1&datDiario=09/12/2022&paginaDireta=3&indSuplemento=Sim&codSuplemento=B&desVolumeSuplemento=&desTomoSuplemento=&_gl=1*hulqbi*_ga*MjgwNzg1MDUzLjE2MjQ0ODA3Nzk.*_ga_CW3ZH25XMK*MTY5NjUzMDI1MS4yNS4xLjE2OTY1MzA3MDcuMC4wLjA. Acesso em: 30 mar. 2024.

Defesa do Consumidor. A previsão do artigo 20, da LGPD, reúne e organiza uma série de prerrogativas que decorrem não apenas da autodeterminação informativa do titular de dados (art. 2º, II, da LGPD) (Frazão, 2018), mas dos próprios princípios da Lei:

> (i) livre acesso (art. 6º, IV), (ii) qualidade e a clareza dos dados (art. 6º, V), (iii) transparência dos dados, (art. 6o, VI), (iv) a prevenção de danos (art. 6º, VII), (v) não discriminação (art. 6º, IX) e (vi) responsabilização e prestação de contas (art. 6o, X) (Frazão, 2018).

O direito à explicação, portanto, não diz respeito a um mero pedido de acesso a informações. Ele deve ser compreendido, na verdade, como um conjunto de direitos e prerrogativas que o titular pode exercer no caso de decisões automatizadas, "que podem ou não estar destinadas à formação de perfil, indicando a possibilidade de que o direito à explicação possa ocorrer independentemente de ter havido ou não a formação de perfil" (Frajhof, 2022, p. 75).

Deste modo, a LGPD permite que sejam adotadas decisões baseadas exclusivamente em tratamento automatizado de dados pessoais, mesmo que essas decisões afetem os interesses do titular. Nesses casos, o titular pode acionar uma série de medidas, a saber: (i) solicitar a revisão da decisão (conforme disposto no art. 20, *caput*, da LGPD); (ii) ter acesso a informações detalhadas e claras sobre os critérios e processos empregados na decisão automatizada (art. 20, §1º, da LGPD); (iii) requerer, por meio de petição, que a Autoridade Nacional de Proteção de Dados (ANPD) conduza uma auditoria para averiguar possíveis práticas discriminatórias, caso o controlador se recuse a fornecer os dados, alegando segredo comercial ou industrial (art. 20, §2º, combinado com o art. 18, §1º, da LGPD); e (iv) exercer o direito de oposição ao referido tratamento (art. 18, §2º, da LGPD) (Mulholland e Frajhof, 2019) .

Esse conjunto de prerrogativas visa impedir que qualquer pessoa seja submetida, de forma irrestrita e sem garantias, a julgamentos oriundos de decisões completamente automatizadas, garantindo o que alguns autores denominam de "devido processo algorítmico" (Mulholland; Frajhof, 2019).

Nesse contexto, destaca-se, contudo, que nem toda decisão tomada exclusivamente por sistemas automatizados se enquadra automaticamente nas disposições do artigo 20 da LGPD. O rol apresentado no *caput* do artigo deve ser

interpretado como meramente ilustrativo, e não como uma lista fechada de situações (Monteiro, 2018). Isso impõe um papel ativo à Autoridade Nacional de Proteção de Dados (ANPD), ao Judiciário e à doutrina, que passam a ter o encargo de exemplificar e delimitar em quais cenários o direito à explicação e à revisão deve ser assegurado.

Como a LGPD não especifica com exatidão quais tipos de impacto são considerados relevantes para acionar essas garantias, a abrangência da norma acaba sendo bastante ampla. Em geral, basta que a decisão automatizada afete de alguma forma os interesses do titular, especialmente quando envolve a construção de perfis pessoais, profissionais, de consumo, de crédito ou relacionados a características da sua personalidade (Frajhof, 2022).

Já foi mencionado que o artigo 20 da LGPD teve como base o artigo 12 do GDPR - no que se refere à transparência de informação -, os artigos 13 e 14 - sobre as informações que devem ser fornecidas aos titulares quando da coleta e tratamento (direito de notificação) - e o artigo 15 - sobre acesso a dados - e artigo 22 - no que tange a decisões automatizadas. Também, mas não somente, está baseado nos Considerandos da GDPR, que acompanham o Regulamento e foram importantes bases para a criação da norma brasileira. Considerando nosso escopo - o direito à explicação -, ressaltam-se, ainda, os Considerandos 58 - princípio da transparência, 60 - obrigação de informação -, 63 - direito de acesso - e 71 - *profiling*. Observe-se que o principal objetivo dos Considerandos é estabelecer a teleologia fundamental dos direitos - no caso, de informação e de acesso -, mas não as formas como devem ser exercidos.

Apenas a partir de uma leitura conjunta e sistemática destes dispositivos, pode-se extrair o que ficou conhecido como o direito à explicação de decisões algorítmicas, embora ainda não esteja bem definido e delimitado como estes comandos normativos devem ser atendidos. De toda sorte, o termo “explicação” apenas é utilizado no Considerando 71, do GDPR, não se repetindo em nenhum outro momento.

Antes, contudo, importa mencionar a primeira regulação da União Europeia sobre proteção de dados estabelecia que os Estados-Membros reconheciam o direito de um indivíduo a não ser submetido à tomada de decisão exclusivamente

automatizada - vide artigo 15 da Diretiva 95/46/CE[10]. Da mesma forma, no contexto de decisões automatizadas, garantia que o indivíduo conhecesse a lógica subjacente ao tratamento autorizado dos dados que lhe dissessem respeito - vide artigo 12 da mesma Diretiva[11]. Tais dispositivos fomentaram discussões a respeito da existência de um direito ou de uma proibição da decisão automatizada.

Com a homogeneização da proteção de dados nos países da União Europeia, que se deu pela promulgação do Regulamento Geral de Proteção de Dados (GDPR), a questão atinente à tomada de decisões automatizadas passou a ser regulada pelo artigo 22[12], havendo transportado tais questionamentos sobre o caráter permissivo ou proibitivo (Toniazzo *et al* 2021).

---

[10] Artigo 15º. 1. Os Estados-membros reconhecerão a qualquer pessoa o direito de não ficar sujeita a uma decisão que produza efeitos na sua esfera jurídica ou que a afecte de modo significativo, tomada exclusivamente com base num tratamento automatizado de dados destinado a avaliar determinados aspectos da sua personalidade, como por exemplo a sua capacidade profissional, o seu crédito, confiança de que é merecedora, comportamento. 2. Os Estados-membros estabelecerão, sob reserva das restantes disposições da presente directiva, que uma pessoa pode ficar sujeita a uma decisão do tipo referido no nº 1 se a mesma: a) For tomada no âmbito da celebração ou da execução de um contrato, na condição de o pedido de celebração ou execução do contrato apresentado pela pessoa em causa ter sido satisfeito, ou de existirem medidas adequadas, tais como a possibilidade de apresentar o seu ponto de vista, que garantam a defesa dos seus interesses legítimos; ou b) For autorizada por uma lei que estabeleça medidas que garantam a defesa dos interesses legítimos da pessoa em causa.

[11] Artigo 12º. Tomada de decisão individual automatizada, incluindo criação de perfis. Direito de acesso Os Estados-membros garantirão às pessoas em causa o direito de obterem do responsável pelo tratamento: a) Livremente e sem restrições, com periodicidade razoável e sem demora ou custos excessivos: - a confirmação de terem ou não sido tratados dados que lhes digam respeito, e informações pelo menos sobre os fins a que se destina esse tratamento, as categorias de dados sobre que incide e os destinatários ou categorias de destinatários a quem são comunicados os dados, - a comunicação, sob forma inteligível, dos dados sujeitos a tratamento e de quaisquer informações disponíveis sobre a origem dos dados, - o conhecimento da lógica subjacente ao tratamento automatizado dos dados que lhe digam respeito, pelo menos no que se refere às decisões automatizadas referidas no nº 1 do artigo 15º; b) Consoante o caso, a rectificação, o apagamento ou o bloqueio dos dados cujo tratamento não cumpra o disposto na presente directiva, nomeadamente devido ao carácter incompleto ou inexacto desses dados; c) A notificação aos terceiros a quem os dados tenham sido comunicados de qualquer rectificação, apagamento ou bloqueio efectuado nos termos da alínea b), salvo se isso for comprovadamente impossível ou implicar um esforço desproporcionado.

[12] Artigo 22. O titular dos dados terá o direito de não estar sujeito a uma decisão baseada exclusivamente no processamento automatizado, incluindo a criação de perfis, que produza efeitos legais a seu respeito ou que o afete significativamente de forma semelhante. O parágrafo 1 não se aplicará se a decisão for necessária para a celebração ou a execução de um contrato entre o titular dos dados e um controlador de dados for autorizada pela legislação da União ou do Estado Membro à qual o controlador está sujeito e que também estabelece medidas adequadas para proteger os direitos e liberdades e os interesses legítimos do titular dos dados; ou for baseado no consentimento explícito do titular dos dados. Nos casos mencionados nas alíneas (a) e (c) do parágrafo 2, o controlador de dados deve implementar medidas adequadas para proteger os direitos e liberdades e os interesses legítimos do titular dos dados, pelo menos o direito de obter intervenção humana por parte do controlador, de expressar seu ponto de vista e de contestar a decisão. As decisões referidas no parágrafo 2 não devem se basear em categorias especiais de dados pessoais referidas no artigo 9(1), a menos que se apliquem as alíneas (a) ou (g) do artigo 9(2) e que existam medidas adequadas para proteger os direitos e liberdades e os interesses legítimos do titular dos dados.

Lima *et al* (2020, p. 234) afirmam que o Artigo 22 do GDPR contém “norma ostensivamente proibitiva”, na medida que expressamente veda decisões baseadas unicamente em processos automatizados quando estas impactam significativamente a esfera jurídica do indivíduo. A partir dessa previsão, depreende-se a existência de um direito à explicação, que funciona tanto como uma prerrogativa pessoal do titular dos dados quanto como um mecanismo voltado à fiscalização dessas decisões automatizadas[13].

Importa mencionar que o objeto contido no artigo 12 do GPDR já havia sido abordado parcialmente na *Directive 95/46 - Data Protection Directive*, do Parlamento Europeu e do Conselho, de 24 de outubro de 1995, relativa à proteção de pessoas no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados[14].

Na visão de Kuner *et al* (2020, p. 401/402), o artigo 12 do GDPR tem por fundamentação e base política “o exercício eficaz dos direitos de informação e de acesso, em primeiro lugar em benefício das pessoas em causa e, em segundo lugar, em benefício dos responsáveis pelo tratamento”. E segue o mesmo autor:

> A justificativa subsequente para o Artigo 12 é a necessidade de maior clareza e consistência das regras sobre troca de informações entre controladores e titulares de dados em toda a União.
> As regras unificadas têm como objetivo melhorar a compreensão e a confiança do consumidor nas transações transfronteiriças e, ao mesmo tempo, oferecer às empresas a oportunidade de desenvolver apenas um conjunto de procedimentos de conformidade para todo o mercado comum europeu, em vez de ter que investir em soluções diferentes para cada jurisdição.

---

[13] Ao contrário do ilustrado, No ordenamento jurídico brasileiro, por outro lado, o direito à explicação está previsto em situações específicas, especialmente quando: (a) a decisão é tomada de forma inteiramente automatizada, sem intervenção humana; (b) há repercussões relevantes para os interesses do titular; e (c) o objetivo da decisão é a definição de um perfil pessoal.

[14] Artigo 12 - Direito de acesso. Os Estados-Membros garantirão a todo titular de dados o direito de obter do responsável pelo tratamento: (a) sem restrições, a intervalos razoáveis e sem demora ou despesas excessivas: - confirmação de que os dados que lhe dizem respeito estão ou não sendo processados e informações pelo menos sobre as finalidades do processamento, as categorias de dados em questão e os destinatários ou categorias de destinatários aos quais os dados são divulgados, - comunicação a ele, de forma inteligível, dos dados em processamento e de qualquer informação disponível sobre sua origem, - conhecimento da lógica envolvida em qualquer processamento automático de dados que lhe digam respeito, pelo menos no caso das decisões automatizadas referidas no artigo 15 (1); (b) conforme apropriado, a retificação, o apagamento ou o bloqueio de dados cujo processamento não esteja em conformidade com as disposições da presente diretiva, em especial devido à natureza incompleta ou imprecisa dos dados; (c) a notificação a terceiros a quem os dados tenham sido divulgados de qualquer retificação, apagamento ou bloqueio realizado em conformidade com a alínea (b), a menos que isso seja impossível ou envolva um esforço desproporcional.

Ainda segundo o mesmo autor, no ordenamento jurídico europeu, a essência do direito à explicação surge do direito de informação dos próprios dados (autodeterminação informativa), com fundamento primordialmente no direito do consumidor. Inclusive, até hoje, a abordagem dos direitos de informação no artigo 12 do GDPR é muito parecida à legislação consumerista.

Em complemento, para além da legislação geral da União Europeia (GDPR), em matéria de defesa do consumidor, aplicam-se também disposições específicas relativas ao exercício dos direitos de informação em determinados setores. No setor de comunicações eletrônicas, por exemplo, explica Kuner (2020, p. 403), o artigo 11 da Diretiva Serviço Universal (2002/22/CE) confere aos Estados-Membros poderes para aplicar regulamentação que permita aos utilizadores finais e aos consumidores acesso a informações completas, comparáveis e de fácil utilização.

O artigo 12 do GDPR, para Kuner (2020), se justifica diante da necessidade de maior clareza e coerência das regras sobre o intercâmbio de informações entre os responsáveis pelo tratamento e os usuários. Isso porque, sendo a União Europeia uma comunidade composta por diversos países independentes, buscou-se melhorar a compreensão e a confiança dos consumidores nas transações transfronteiriças, dando simultaneamente às empresas a oportunidade de desenvolverem apenas um conjunto de procedimentos de conformidade para todo o mercado comum europeu, em vez de terem de investir em soluções diferentes para cada jurisdição.

Nota-se, portanto, que a questão relacionada à confiabilidade do processamento de dados é transfronteiriça e atinge a sociedade mundial como uma unidade. O princípio da transparência exige que qualquer informação disponibilizada ao usuário seja concisa, facilmente acessível e compreensível, e que seja utilizada uma linguagem clara e simples, além de que sejam utilizadas técnicas de visualização, se for o caso. Essas informações podem ser fornecidas em formato eletrônico, como, por exemplo, quando dirigidas ao público, através de um sítio Web (Kuner, 2021).

As leis de proteção de dados da UE e do Conselho da Europa exigem que o processamento de dados pessoais seja feito de uma forma transparente em relação ao titular dos dados. Segundo o *Handbook on European data protection law* (2018):

> As operações de processamento devem ser explicadas aos titulares dos dados de forma facilmente acessível, garantindo que eles

> entendam o que acontecerá com seus dados. Isso significa que a finalidade específica do processamento de dados pessoais deve ser conhecida pelo titular dos dados no momento da coleta dos dados pessoais. A transparência do processamento exige que seja usada uma linguagem clara e simples. Deve ficar claro para as pessoas envolvidas quais são os riscos, as regras, as salvaguardas e os direitos relativos ao processamento de seus dados pessoais. (Tradução livre).

A partir das reflexões desenvolvidas ao longo da literatura acerca dos artigos mencionados até o presente momento, cabe destacar o apontamento de Frajhof (2022), que observa um movimento que transcende a mera formulação teórica de direitos e se volta à concretude das salvaguardas que devem acompanhar decisões automatizadas. A autora pontua, especificamente, que

> (i) saber quando estão sujeitas a uma decisão algorítmica em que não há intervenção humana, (ii) exercer uma espécie de devido processo algorítmico, com a garantia de salvaguardas que visem proteger seus direitos e liberdades, (iii) conhecer os motivos que embasaram esta decisão, especialmente quando diante do uso de algoritmos de tomada de decisão que se valem de ML.

É certo que, inicialmente, o debate sobre o direito à explicação no contexto do GDPR centrou-se na discussão sobre o tipo e o conteúdo da explicação para decisões totalmente automatizadas, conforme foi delineado pelo regulamento europeu. Em que pese essa questão ainda não ter sido completamente solucionada, o direito à explicação atualmente integra uma discussão mais ampla sobre a prestação de contas e a responsabilidade dos algoritmos, com ênfase especial nos aspectos de explicabilidade e interpretabilidade dos algoritmos de IA.

Isto é, a noção de explicação não se limita mais ao plano individual. Assume, também, um caráter coletivo, voltado à *accountability* das tecnologias utilizadas, tratando-se de uma exigência de transparência sistêmica que demanda, entre outros aspectos, a documentação dos processos de design, modelagem e treinamento dos algoritmos — elementos essenciais para permitir sua posterior interpretação, controle e crítica.

3.1.2 Direito autônomo ou extensão de outros direitos já positivados ou princípio normativo.

A análise da natureza jurídica do direito à explicação exige um exame cuidadoso de sua fundamentação normativa nas legislações de proteção de dados. A questão emerge a partir da análise de situações reais, como por exemplo o já mencionado caso dos professores de Washington, avaliados e, eventualmente, demitidos com base no sistema do algoritmo IMPACT. Confrontados com pontuações arbitrárias e a exigência de detalhes, a resposta que obtiveram foi de que o sistema era "muito complexo" para ser explicado.

Mas não é só. Em Londres, a previsão automatizada de notas prejudicou estudantes de baixa renda a ingressarem no ensino superior, a concessão de crédito imobiliário para locação, a precificação de novos produtos e a definição de fianças em processos criminais são determinadas por sistemas de decisão automatizada que se baseiam, respectivamente, no histórico de crédito, no comportamento de compra online e no histórico do réu (Wachter; Mittelstadt; Russell, 2021).

Os casos narrados evidenciam o problema central da opacidade algorítmica e a urgência de aprofundar a discussão sobre o direito à explicação. No contexto da LGPD e do GDPR, emerge uma questão central: o direito à explicação se trata de um direito autônomo, de um desdobramento instrumental de direitos já positivados ou um princípio? E é essa questão que busca-se responder na presente seção.

É certo que o "direito à explicação" exige dos responsáveis pelo tratamento de dados a obrigação de fornecer uma explicação clara e pertinente sobre o processo decisório de seus sistemas inteligentes (Kim; Routledge, 2018, p. 64/74).

De uma análise mais detida ao texto legal do ordenamento europeu - especialmente artigos 13, 14, 15 e 22 do GDPR, e da Convenção 108+ -, nota-se não existir algo parecido como um direito à explicação sobre uma decisão algorítmica, mas tão somente um direito de informação sobre as funções do sistema (direito de acesso). Essa é a conclusão a que chegaram Wachter *et al*. (2017), sob um olhar positivista, partindo exclusivamente da legislação. Isso porque, ante à não explicitação de um direito, não seria possível exigir uma obrigação dos controladores.

Tal posicionamento sofreu severas críticas, especialmente no que tange a importância das informações úteis/funcionais - que viabilizem o exercício de outros direitos - serem transmitidas a pessoas leigas, preceito este cuja base legal deveria ser interpretada de maneira flexível. Powles e Selbst (2017), por exemplo, aduzem que os fundamentos de defesa da existência positiva do direito à explicação se

encontram justamente no artigo 22 do GDPR, somado ao Considerando 71, sustentando a leitura dos artigos 13, 14 e 15, do mesmo diploma.

Ao fim e a cabo, como elucidado por Frajhof (2022), os autores que discordam de Wachter *et al* reconhecem que as disposições contidas nos artigos 13 a 15 e 22 do GDPR podem, sim, ser compreendidas como a base para um verdadeiro direito à explicação. Para esses, o foco não deveria recair sobre a denominação formal desse direito, mas sim sobre a proteção substancial que ele assegura aos titulares de dados. Defende-se, nesse sentido, uma leitura funcional e adaptável desse direito, capaz de garantir que os indivíduos não apenas compreendam as decisões automatizadas que os afetam, mas também possam exercer de forma plena os direitos previstos tanto no próprio GDPR quanto em tratados e normativas de direitos humanos que abordem o tema da transparência algorítmica.

Tampouco na LGPD há menção expressa a um direito de explicação. Isto é, conforme apontam Franzolin, Monteiro e Laurentiis (2023), as legislações de proteção de dados não consagram o direito à explicação de forma explícita, o que direciona ao entendimento de que o direito à explicação ou a explicabilidade algorítmica atua como um mecanismo necessário à efetivação de outras garantias fundamentais.

Isto é, o direito à explicação se manifesta como um “direito-ferramenta” ou “guarda-chuva”, cuja função é capacitar o indivíduo a exercer prerrogativas já existentes, mas que seriam inócuas diante da opacidade algorítmica.

O seu caráter instrumental evidencia-se, sobretudo, em sua profunda conexão com outros direitos, dos quais deriva, como por exemplo o direito à informação e a autodeterminação informativa.

Nesse contexto, é imperioso reconhecer que a demanda por explicação não surge de um vácuo normativo, mas assenta-se sobre o alicerce do Direito à Informação, já consagrado no ordenamento brasileiro. O direito à explicação deve ser compreendido como uma releitura tecnológica do dever de transparência que remonta ao Código de Defesa do Consumidor (CDC). O CDC foi pioneiro ao estabelecer, em seu artigo 6º, III, o direito básico à informação adequada, clara e precisa sobre os diferentes produtos e serviços, bem como sobre os riscos que apresentem.

Considerando que os algoritmos de IA são frequentemente utilizados na mediação de relações de consumo — como na concessão de crédito ou precificação de serviços —, a opacidade algorítmica viola frontalmente essa diretriz consumerista histórica. Portanto, o direito à explicação na LGPD não é uma invenção isolada, mas a atualização do direito à informação para a era do *Big Data*: já não basta apenas ser informado estaticamente sobre a existência da coleta de dados, mas é imperativo compreender a dinâmica do processamento, ou seja, a rationale de como esses dados são sopesados para gerar decisões.

Essa interpretação ganha força normativa ao se observar os fundamentos da própria LGPD. O artigo 1º da lei declara expressamente que a disciplina da proteção de dados tem como objetivo a proteção dos direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. Assim, a explicação atua como o elo funcional que conecta o processamento automatizado à dignidade humana, impedindo que o indivíduo seja reduzido a um mero objeto de tratamento de dados, destituído de agência e capacidade de contestação.

O direito à informação é transformado na era da IA: já não basta apenas ser informado sobre a coleta de dados, mas é imperativo compreender como esses dados são processados para gerar decisões que afetam a vida do titular. Nesse contexto, a explicação é parte da forma contemporânea e tecnológica de exercer o direito de saber e de contestar.

Por sua vez, a autodeterminação informativa, conceito que empodera o indivíduo a controlar o fluxo de seus dados pessoais, torna-se um direito abstrato sem a explicabilidade. Isto é, sem a compreensão dos critérios e da lógica de uma decisão automatizada, o titular de dados perde a capacidade de questionar e de governar a própria informação, subvertendo a essência da autodeterminação. A explicação, por conseguinte, atua como o elo que vincula o processamento automatizado à capacidade de agência humana, conferindo substância prática a esse direito.

Tendo em vista a ausência de menção expressa ao direito de explicação no ordenamento jurídico, seria este, então, um princípio normativo?

Partindo da definição de Humberto Ávila, que adota a teoria da generalidade para classificar as normas, apontando como único critério de diferenciação o grau de abstração da previsão normativa (1999, p. 167/168), o conceito de princípios pode ser sintetizado como normas imediatamente finalísticas, que orientam (com menor

determinação) a tomada de decisões, mas exigem um alto grau de interpretação e dependem de outras normas para determinar o comportamento adequado. Por outro lado, o autor destaca que regras são normas mais específicas, mediatamente finalísticas, que estabelecem o comportamento devido com maior precisão e, por isso, demandam menos interpretação para sua aplicação.

A definição relevante não reside na intenção do legislador, mas sim na estrutura inerente à norma a ser interpretada. Isso ocorre porque a natureza de um princípio está intrinsecamente ligada aos seus resultados e às ações que dele emanam. Assim, há direitos (normas positivadas) que denotam princípios – como os direitos fundamentais -, ao mesmo tempo em que há denominados princípios (lato sensu) que denotam regras – a legalidade e a anterioridade, por exemplo. Igualmente, há normas que variam entre regras e princípios a depender do caso concreto analisado – como é o caso da igualdade.

Em contrapartida, o posicionamento adotado por Silva (2003, p. 611) é no mesmo sentido de Robert Alex, ou seja, enquanto os princípios expressam deveres *prima facie*, podendo ser realizados em maior ou menor grau, as regras expressam direitos e deveres definitivos, que apenas podem ser cumpridas em sua totalidade, ou descumpridas.

Tradicionalmente, princípios são definidos como as normas mais fundamentais do sistema, ao passo que regras costumam ser definidas como “uma concretização desses princípios”, tendo caráter instrumental e menos fundamental (2003, p. 612).

Quando falamos da concepção de Alexy é necessário mantermos em vista que sua teoria não conceitua princípios considerando a fundamentalidade da norma, não interferindo, portanto, em nada na ideia de este ser um mandamento nuclear do sistema. Deste modo, antes de se discutir a tipologia desenhada por Alexy, é necessário deixar de lado os tradicionais princípios, porquanto partindo de sua fundamentalidade, “não poderiam mais ser consideradas como princípios, devendo ser incluídos na categoria das regras” (2003, p. 612/614).

É certo, entretanto, que a discussão entre os juristas sobre as espécies e classificações das normas de direitos fundamentais é infindável, não chegando “ao menos perto de algum denominador comum”, como menciona Silva (2003, p. 607).

Em atenção à explicabilidade, Araújo (2024) a aborda não como um direito, mas como princípio, fundamental à construção de um marco regulatório que

assegure a responsabilidade e a conformidade dos sistemas de IA. Araújo (2024) autor destaca a natureza de "princípio da explicabilidade" por compreender que essa noção funcionaria como um guia normativo fundamental para a regulação da inteligência artificial, cuja função seria permitir a compreensão, a rastreabilidade e a justificação das decisões automatizadas, garantindo transparência e proteção dos direitos humanos no contexto da IA. Isto é, trataria-se de princípio normativo em razão da abordagem enquanto fundamento normativo que influencia e orienta as práticas, regulamentações e controles da IA, não se configurando apenas como uma simples obrigação técnica, mas como um balizador que harmoniza tecnologia e direitos fundamentais.

Além disso, Araújo (2024, p. 295) situa a explicabilidade como um dos quatro princípios éticos centrais indicados no "Guia Europeu de Orientações para IA de Confiança" (os outros sendo autonomia humana, prevenção de danos e justiça), mostrando seu papel fundamental para a construção de sistemas de IA confiáveis, transparentes e alinhados a direitos fundamentais.

O princípio da explicabilidade, portanto, opera como um parâmetro ético e jurídico que orienta a criação, o desenvolvimento e a regulação da IA, objetivando promover a responsabilidade, a segurança e o respeito à dignidade humana nas decisões tomadas por sistemas automatizados.

Isto é, nessa ótica apresentada pelo Guia Europeu, a explicabilidade não se limita à mera exposição dos códigos de programação, que poderia comprometer segredos industriais, mas transcende à capacidade de fornecer uma explicação acessível, proporcional e adequada ao usuário, de modo a assegurar a compreensão das decisões automatizadas e possibilitar sua a contestação e controle (Araújo, 2024, p. 288).

Assim, tem-se que a partir da análise empreendida, conclui-se que o direito à explicação em sua atual configuração não possui natureza de um direito autônomo e expressamente positivado nas legislações de proteção de dados analisadas. Sua existência, ao contrário, deriva de uma análise hermenêutica e sistêmica de outros direitos humanos e fundamentais já consagrados, como o direito de informação, a autodeterminação informativa e o devido processo legal, elucidando sua impossibilidade de ser tratado como uma regra fechada.

A característica de ser um direito derivado e não uma regra expressa denota sua natureza de princípio, na medida em que princípios são normas jurídicas de

caráter mais abstrato, que servem de base e orientam o sistema jurídico, ao contrário das regras que são mais específicas e diretamente aplicáveis. Ao ser concebido como princípio, o direito à explicação não é imediatamente finalístico, mas sim uma diretriz que orienta a interpretação e a aplicação das normas no contexto das decisões automatizadas. Sua função atual não é a de ser um comando de cumprimento integral, mas a de guiar a governança algorítmica, reforçando a *accountability* e a transparência dos sistemas de inteligência artificial.

3.1.2.1 *Relação da explicabilidade com direitos fundamentais*

A discussão acerca da natureza jurídica do direito à explicação e de sua classificação como princípio ou regra, embora fundamental para a delimitação teórica do tema, encontra seu propósito final na sua capacidade de salvaguardar direitos e garantias fundamentais na era da automação. A opacidade dos sistemas de inteligência artificial não se apresenta como um problema meramente técnico ou de transparência. Na verdade, se traduz em ameaça concreta à proteção dignidade da pessoa humana, da liberdade, autonomia, do livre desenvolvimento da personalidade e a igualdade (Kaminski, 2019; Selbst; Doneda *et al*., 2018).

Nesse contexto, o direito à explicação emerge como um mecanismo de controle e mitigação de riscos, cuja instrumentalidade se revela essencial para assegurar que o avanço tecnológico ocorra em consonância com os preceitos éticos e jurídicos que sustentam o Estado Democrático de Direito. Isto é, é intrínseca a relação entre a explicabilidade e a efetivação desses direitos, de sorte que sua ausência compromete a agência do indivíduo, a justiça social e a previsibilidade das decisões.

Embora o direito à explicação não seja expressamente positivado de forma autônoma na LGPD e na GDPR, sua instrumentalidade assume uma função de suma importância na proteção de direitos fundamentais e humanos, como a dignidade da pessoa humana, a igualdade e o devido processo legal. A opacidade intrínseca aos sistemas de IA, ou a problemática da "caixa-preta", representa uma ameaça não apenas à transparência, mas aos alicerces de um sistema jurídico-social justo e equitativo.

A explicabilidade é um pilar para a proteção de direitos fundamentais mais amplos e de cunho constitucional, como a dignidade da pessoa humana, a igualdade

e o devido processo legal. No que tange à dignidade, a impossibilidade de entender uma decisão que afeta a vida de um indivíduo pode ser interpretada como uma forma de desumanização. Ao ser tratado como um dado em um processo obscuro, o ser humano é reduzido a um objeto, e não a um sujeito de direitos. A explicação, por sua vez, restaura a agência e a autonomia, permitindo que a pessoa compreenda e reaja, reafirmando sua condição de sujeito.

A isonomia e a não discriminação são igualmente afetadas pela opacidade algorítmica. Os algoritmos, ao serem treinados com dados que refletem preconceitos sociais, podem replicá-los e até intensificá-los de forma invisível. A ausência de transparência nos algoritmos da Amazon, por exemplo, que evidenciou um viés de gênero, é um caso paradigmático. O direito à explicação é, portanto, a garantia fundamental para que a discriminação algorítmica possa ser identificada e contestada, servindo como uma ferramenta para que se fiscalizem os critérios de decisões automatizadas e se exponham vieses injustos.

De plano, ressalta-se que o princípio da dignidade da pessoa humana é - como certamente deveria ser deveria ser - o cerne axiológico deste debate. A exigência da transparência com base em direitos individuais – como a dignidade da pessoa, a liberdade, a autonomia e o desenvolvimento da personalidade – é uma forma de atender às preocupações sobre a justificação e legitimidade de certas decisões, conforme apontado por Kaminski (2019). Além disso, o exercício desses mesmos direitos possibilita, justifica e instrumentaliza outras reivindicações por explicações.

Partindo dessas premissas, observa-se que a eficácia do direito à explicação está profundamente ligada ao reconhecimento da proteção de dados pessoais como um direito fundamental, consagrado constitucionalmente.

O tratamento de dados pessoais de indivíduos tem sido elevado ao patamar de direito fundamental em diversas legislações internacionais (Pinheiro, 2020). Além disso, Bioni (2019) enfatiza que, por constituírem direitos da personalidade, os dados pessoais ultrapassam o escopo restrito da privacidade.

O Brasil, como explica Franzolin *et al* (2025, p. 75), consagra os direitos à proteção de dados e à privacidade como direitos fundamentais na Constituição da República Federativa do Brasil (CRFB/88), especificamente, o Artigo 5º, listando a proteção da "intimidade" e da "vida privada" nos incisos X, XI e XII. Esses incisos tratam da inviolabilidade de casas e do sigilo de correspondência, reforçando a

proteção da esfera privada. Nada obstante, a Lei Geral de Proteção de Dados (LGPD) tem como objetivo proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural.

O foco central da Lei Geral de Proteção de Dados (LGPD) é o resgate da dignidade dos titulares de dados e de seus direitos básicos, conforme Frazão (2018), com destaque para a autodeterminação informativa. Por sua vez, Monteiro (2018) ressalta que esse tipo de legislação não se restringe à privacidade, abrangendo um conjunto mais amplo de direitos.

O direito de proteção de dados, enquanto um direito fundamental, possui duas vertentes: a dimensão subjetiva e a dimensão objetiva (Mendes, 2014, p. 176). A dimensão subjetiva garante ao indivíduo o direito à autodeterminação informativa e um espaço de liberdade e privacidade, permitindo o controle sobre o fluxo de suas informações pessoais e assegurando meios de proteção contra ameaças ou intervenções. Contudo, essa proteção não é absoluta, devendo ser ponderada em relação a direitos e interesses de terceiros, exigindo uma análise de necessidade e proporcionalidade entre a finalidade do tratamento de dados e o direito à sua proteção (Mendes, 2014, p. 177-178). Por outro lado, a dimensão objetiva (dever de proteção estatal) exige uma ação do Estado para concretizar e delimitar esses direitos, promovendo mecanismos e procedimentos para o tratamento legítimo de dados pessoais e, simultaneamente, abstendo-se de interferir na liberdade e privacidade do titular (Mendes, 2014, p. 176). Essa perspectiva dupla — individual e coletiva — está ligada à *accountability* algorítmica binária.

A relação da opacidade das decisões algorítmicas e dos direitos humanos e fundamentais também se escancara ante à afetação direta da personalidade e da integridade moral dos indivíduos, impondo restrições à sua autonomia e liberdade de escolha. Nesse aspecto, mencionando Selbts e Barocas (2016, p. 1120), Frajhof (2022, p. 172) explica que o direito à explicação é considerado um "bem em si" (*inherent good*), por possibilitar a participação da pessoa no processo decisório, permitindo a avaliação das categorizações e da representação feitas sobre ela, o conhecimento do uso de decisões automatizadas e a compreensão de como seus aspectos pessoais foram considerados pelo algoritmo. E, ainda, prossegue lecionando que esse acesso à informação é crucial por fornecer os elementos necessários para que o indivíduo tome decisões informadas, como solicitar a revisão de uma decisão algorítmica ao verificar, por exemplo, que seu histórico de crédito

atual não foi considerado, o que demonstra o valor funcional e instrumental desse direito, já que o acesso à informação se configura como pressuposto essencial para o exercício da liberdade e da autonomia, condições-chave para o exercício de outros direitos (Selbst; Barocas, 2016; Selbst; Powles, 2017).

O acesso à informação sobre o processo de decisão automatizada é, portanto, um pressuposto para o exercício da liberdade e da autonomia.

Já no que se refere à autodeterminação informativa, em específico, este direito subjetivo é definido como a capacidade do indivíduo de manter o controle sobre suas próprias informações e determinar a maneira de construir sua própria esfera particular.

Como definido por Rodotà (2008, p. 15 *in* Frajhof, 2022, p. 41), no âmbito da formação de perfis (perfilização ou *profiling*), “o direito de manter o controle sobre suas próprias informações e de determinar a maneira de construir sua própria esfera particular”.

A opacidade inerente às técnicas de inteligência artificial e *profiling* ameaça gravemente a autodeterminação informativa, vez que o indivíduo geralmente desconhece os critérios e procedimentos pelos quais seus perfis são construídos e as decisões são tomadas, perdendo o controle sobre o fluxo de seus dados pessoais, concedendo ao titular dos dados apenas uma falsa ideia de controle. O direito à explicação surge, assim, como um mecanismo para assegurar o controle efetivo sobre os dados pessoais (Costa, 2021, p. 67).

> O regime jurídico conferido pelo RGPD não se concretizará plenamente se ao titular dos dados não for garantido o controlo efetivo sobre os seus dados pessoais. A necessidade de assegurar a sua autodeterminação informacional convoca a implementação de mecanismos de transparência capazes de agir sobre múltiplas camadas: da (in) visibilidade do tratamento, aos dados tratados, às regras a que o seu funcionamento obedece e às razões que fundamentam as decisões, conditio sine qua non para que os meios de tutela consagrados sejam efetivos. Assim se compreende que a linha argumentativa da doutrina maioritária que se propõe diluir a dificuldade de regulamentação desta nova realidade assente essencialmente em apelar à transparência e ao reconhecimento de um direito a obter uma explicação sobre decisões automatizadas (COSTA, 2021, p. 67).

Demais disso, um dos maiores riscos éticos e jurídicos do uso de IA é o potencial discriminatório dos algoritmos. As decisões algorítmicas, longe de serem neutras, podem replicar e até reforçar os preconceitos existentes na sociedade,

violando os princípios constitucionais da igualdade e não discriminação (Requião, 2022, p. 4).

O direito à explicação se mostra, portanto, ferramenta primária para identificar e investigar a ocorrência de discriminações e vieses codificados nos algoritmos, permitindo ao indivíduo avaliar se o tratamento de dados automatizado deixou de promover adequadamente o direito à proteção de dados ou se houve discriminação. Nesse ínterim, a LGPD, em seu Art. 20, § 2º, prevê que a auditoria pela Autoridade Nacional de Proteção de Dados (ANPD) visa precisamente verificar aspectos discriminatórios no tratamento automatizado

As decisões destinadas a definir perfil pessoal, profissional, de consumo e de crédito, por exemplo, possuem potencial de anular a capacidade de perceber as nuances do indivíduo, resultando em categorizações errôneas ou inexatidão dos dados que perpetuam preconceitos. A explicação fornece o subsídio para a correção de eventuais inacurácias e o combate a este potencial lesivo.

Assim, a garantia de que o indivíduo possa compreender e contestar as decisões automatizadas que o afetam não é apenas uma exigência de transparência, mas um imperativo ético e jurídico para assegurar que a tecnologia se harmonize com os alicerces de um Estado Democrático de Direito, reafirmando o ser humano como o centro do processo decisório.

3.1.2.2 Princípios da transparência, accountability e governança algorítmica no âmbito do direito de explicação

A eficácia do direito à explicação ultrapassa a esfera da proteção individual, consolidando-se como um pilar essencial para a governança algorítmica democrática e responsável. A relação entre a explicabilidade e os princípios de justiça, prestação de contas e transparência é resumida no acrônimo FAT (*Fairness, Accountability, Transparency*) (CASEY *et al.*, 2019, p. 148). O FAT busca evitar que sistemas de IA produzam resultados discriminatórios, cometam injustiças ou causem erros e danos, por meio da conferência de justiça/igualdade (*fairness*), prestação de contas/responsabilização (*accountability*) e transparência (*transparency*) (FRAJHOF, 2022).

A transparência, segundo Turilli e Floridi (2009) não é um princípio ético, mas uma condição próética, um meio valioso para fins éticos.

Fundamentalmente, a transparência é definida em termos da disponibilidade de informação, às condições de acessibilidade e à forma como a informação pode apoiar de forma pragmática ou epistémica o processo de tomada de decisão do utilizador (Turilli e Floridi, 2009, p. 106), sendo a acessibilidade e a compreensibilidade os seus componentes essenciais. Contudo, informações sobre a funcionalidade dos algoritmos são frequentemente mantidas intencionalmente pouco acessíveis. Algoritmos proprietários são resguardados por segredo para garantir vantagem competitiva (Glenn e Monteith, 2014; Kitchin, 2016; Stark e Fins, 2013), segurança nacional (Leese, 2014) ou por razões de privacidade, o que demonstra que a transparência pode contrapor-se a outros ideais éticos, notavelmente a autonomia das organizações e a privacidade dos titulares dos dados.

Neste contexto, observa-se uma luta de poder entre os interesses dos titulares dos dados em obter transparência e a viabilidade comercial dos processadores (Granka, 2010). A abertura da estrutura algorítmica poderia facilitar manipulações mal-intencionadas dos resultados de pesquisa, ou «manipulação do sistema», sem beneficiar o utilizador comum sem conhecimentos técnicos (Granka, 2010; Zarsky, 2016), chegando a ameaçar a subsistência de negócios em setores como a negociação de alta frequência. Apesar disso, os titulares mantêm um interesse legítimo em compreender como as decisões baseadas em dados que os afetam são criadas e influenciadas. Esta disputa é agravada pela assimetria da informação e por um desequilíbrio no conhecimento e no poder de decisão que claramente favorece os processadores de dados (Tene e Polonetsky, 2013a: 252).

Para que a transparência seja efetiva, a informação deve ser, além de acessível, compreensível (Turilli e Floridi, 2009). Os esforços para desvendar a opacidade algorítmica enfrentam o desafio técnico significativo de tornar os processos complexos de tomada de decisão acessíveis e inteligíveis. Este obstáculo é evidenciado pelo problema persistente da interpretabilidade em algoritmos de aprendizagem automática (*machine learning*), que sublinha a dificuldade em superar a opacidade inerente a muitos sistemas (Burrell, 2016; Hildebrandt, 2011).

Segundo os mesmos autores, a origem o termo *accountability* é empregado para se referir à responsabilização, prestação de contas, fiscalização ou sanção, bem como à *answerability* (Augusto; Rizzardmonta ao contexto do direito público, caracterizado pela exigência de prestação de contas dos Entes públicos e pela fiscalização de seus atos, seja pelo próprio Estado e/ou por seus cidadãos.

Outrossim, os mesmos sentidos são, por vezes, atribuídos à “transparência”, o que leva à comum confusão entre os termos, que passam a ser utilizados por alguns como sinônimos.

Entretanto, ao contrário de sinônimos, a transparência se mostra como mecanismo facilitador da *accountability*, sendo um dos elementos que devemos demandar do governo e incentivar na indústria (Diakopoulos, 2016).

O direito à explicação atua como a manifestação mais concreta e funcional do princípio da transparência algorítmica. Enquanto a transparência é o valor que exige a abertura do sistema, a explicação é a ação jurídica que o viabiliza, traduzindo o conceito abstrato em um mecanismo exigível. A transparência é, portanto, um pressuposto para a verificação dos algoritmos, a fim de avaliar a legitimidade e legalidade dos seus resultados.

Contudo, como mencionado, a transparência não se limita à abertura indiscriminada do código-fonte. A mera disponibilização do código não é suficiente, pois "visualizar o código e o seu funcionamento interno não significa entendê-lo" (Ananny; Crawford, 2018, p. 978). O desafio reside na opacidade intrínseca de sistemas complexos, onde a abertura de elementos técnicos (como parâmetros e abstração de dados) ainda torna "a avaliação da conformidade jurídica um problema" (Frazão; Gottenauer, 2020, p. 57).

A propósito, a disponibilização do código-fonte é insuficiente, pois "visualizar o código e o seu funcionamento interno não significa entendê-lo". É necessário prover uma tradução da linguagem matemática para a linguagem natural, o que Frazão e Gottenauer (2020) consideram o papel da explicabilidade.

Nesse sentido, o direito à explicação tem a pretensão de servir como uma ponte conciliatória, promovendo a tradução da linguagem técnico-matemática para a linguagem natural, essencial para a compreensão jurídica.

Nesse ponto, Pereira (2021) expõe o tema sob o prisma de IA aplicada às plataformas de interação social:

> A inteligibilidade de sistemas de ML [machine learning] não deve consistir necessariamente em uma descrição precisa e detalhada de como os algoritmos funcionam, especialmente porque tal forma de fornecimento de informações pode ser inútil para o usuário final de uma plataforma de mídia social que recebe desinformação e que pretende aprender mais sobre como as informações são direcionadas para sua conta (PEREIRA, 2021, p. 97).

Em breve síntese, Acioly (2023, p. 235) apresenta que Angelov *et al*. (2021) definem transparência algorítmica como o potencial de um sistema de IA ser compreendido (oposto de *blackbox*), interpretabilidade como a capacidade de fornecer interpretações humanas e explicabilidade como a interface de explicação entre ser humano e sistema de IA.

O princípio da *accountability* (prestação de contas) é fortalecido diretamente pela explicabilidade. A tese de Frajhof (2022) destaca que a explicabilidade se insere em uma discussão mais ampla de prestação de contas, devendo atender a uma perspectiva de mérito e de procedimento.

Conforme observado por Kaminski (2019), a *accountability* deve assumir diferentes formas para atender a objetivos de substância e procedimento, o que envolve fiscalização e monitoramento, análise por especialistas, além do conhecimento e envolvimento do público. A transparência, embora necessária, não é suficiente, pois são necessários outros mecanismos para avaliar a prestação de contas e a responsabilização de um algoritmo de tomada de decisão. Sem uma explicação que forneça o rastro da decisão, torna-se quase impossível determinar a responsabilidade por um dano causado por um sistema de IA. A explicabilidade, ao fornecer o fundamento da decisão, viabiliza a auditoria, a responsabilização dos *players* (desenvolvedores, *data controllers*) e cria um sistema de vigilância que assegura a justiça e a conformidade.

A conjugação desses princípios, FAT, demonstra que o direito à explicação é um elemento central para a governança algorítmica. Ele é o pilar que permite *insights*, contestações e, sobretudo, a prestação de contas, sendo a transparência um pressuposto para a viabilização dos outros dois. Essa interconexão consolida o direito à explicação como um elemento essencial para garantir que a IA seja desenvolvida e utilizada de maneira democrática e responsável, lastreada na proteção dos direitos fundamentais.

Significa dizer que essa conjugação desses direitos e princípios, como o direito de informação, a autodeterminação informativa e a *accountability*, consolida o direito à explicação como um elemento central para uma governança algorítmica que seja democrática e responsável, pautada não apenas na eficiência, mas também na proteção de direitos fundamentais

3.2 Classificações e limitações do direito à explicação

O direito à explicação, como visto, surge como uma salvaguarda essencial contra a opacidade algorítmica, todavia, evidentemente sua operacionalização exige delimitações, especialmente no concernente a sua abrangência (conteúdo e tipos), ao momento de sua aplicação e também aos limites da titularidade e exigibilidade.

Por essa razão, a literatura jurídica e técnica tem mapeado diversas classificações para conferir sentido útil à explicação, visto que o que constitui uma "explicação adequada" varia de acordo com o contexto, o destinatário, e a finalidade que se busca atingir (Frajhof, 2022, p. 17).

3.2.1 *Quanto à abrangência do direito de explicação*

A classificação do direito à explicação quanto à sua abrangência refere-se ao conteúdo que deve ser revelado e é fundamental para determinar sua utilidade prática e jurídica. Ademais, a abrangência do direito à explicação refere-se tanto ao tipo de informação a ser fornecida quanto à finalidade que essa informação deve cumprir.

É consenso que a mera abertura do código-fonte (transparência) não é suficiente para a compreensão do sistema por um ser humano (Sardinha, 2022, p. 278).

Isto é, o direito de explicação abrange a funcionalidade geral do sistema - lógica, árvores de decisão, modelos, critérios e estruturas - (Franzolin *et al*, 2023, p. 09) e as decisões específicas do algoritmo (Wachter, et. al., 2017, p.76-78). Exige-se o esclarecimento sobre a lógica e funcionalidade geral do sistema automatizado, bem como a fundamentação individualizada de cada decisão (Wachter, et. al., 2017, p.76-78).

Significa dizer que o escopo da explicabilidade deve incidir sobre múltiplos níveis de informação para mitigar a opacidade inerente aos sistemas de inteligência artificial.

3.2.1.1 Explicações Centradas no Modelo (Model-Centric Explanation – MCE) ou Funcionalidade do Sistema (Global) - macro explicabilidade - abrangência da explicação sobre a funcionalidade geração - conteúdo mínimo requerido - lógica e critérios

Esta primeira modalidade foca na lógica global e nos parâmetros do sistema, possuindo uma função eminentemente preventiva. A exigência é de transparência em relação à arquitetura do algoritmo e aos seus objetivos. Franzolin (2023, p. 9), citando Wachter, et. al. (2017, p. 76-78), descreve essa abrangência como a explicação a respeito "*[...] da lógica do sistema, das árvores de decisão, modelos predefinidos, critérios e estruturas de classificação*".

Inclusive, como já abordado anteriormente, o artigo de Wachter, *et al*. (2017), sustenta que o que se pode derivar do articulado da GDPR (Artigos 13, 14 e 15) é, no máximo, um "direito a ser informado" sobre a funcionalidade do sistema, e não um direito irrestrito à explicação da decisão em si. A aplicação prática dessa abrangência se materializa nas obrigações de documentação e nas Avaliações de Impacto (como o RIPDP ou o AIA), visando a mitigação *ex ante* de riscos.

Tanto o GDPR quanto a LGPD estabelecem um conteúdo mínimo que deve ser disponibilizado pelo controlador de dados:

No GDPR, baseado nos Arts. 13, 14 e 15, exige-se o fornecimento de "informações úteis relativas à lógica subjacente", bem como a importância e as consequências previstas de tal tratamento para o titular dos dados (Costa, 2021, p. 67/68). Já a LGPD, com fundamento no Art. 20, § 1º, determina-se que o controlador forneça "informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada", resguardados os segredos comercial e industrial (Lima, 2020, p. 232; Sardinha, 2022, p. 275).

A obrigação de informar "critérios e procedimentos" na LGPD pode ser interpretada como a exigência de informações sobre como o algoritmo opera e sua funcionalidade, sem necessariamente obrigar a divulgação de elementos técnicos aos titulares (FRAJHOF, 2022, p. 197).

Em suma, esse modelo de explicação, chamado de Explicações Centradas no Modelo (Model-Centric Explanation – MCE) ou Funcionalidade do Sistema (Global) focam em como o modelo foi desenvolvido, metadados de treinamento, métricas de comportamento (acertos e falhas), e como os dados de entrada

influenciam o resultado geral. Este tipo de explicação se assemelha às explicações *ex ante*. A interpretabilidade global no contexto da IA Explicável (XAI) tem sua funcionalidade direcionada à descrição do comportamento do modelo, o que inclui a compreensão das categorias de dados necessárias.

Para tanto, os autores indicam que as MCE devem incluir (i) informações sobre a maneira como o modelo foi desenvolvido, a intenção no processo de modelagem, o tipo de modelo, e os parâmetros utilizados; (ii) metadados do treinamento, como resumo das estatísticas, descrição da qualidade dos dados utilizados para treinar o modelo, onde e como os dados foram obtidos, e quais seriam os dados de saída ou as classificações que foram realizadas; (iii) métricas de comportamento, o que envolve apresentar resultados sobre a habilidade de predizer dados nunca antes analisados, as falhas e acertos; (iv) explicações que possam ser compreendidas por um humano sobre como certos dados de entrada influenciaram ou são transformados em dados de saída, e (v) informação sobre o processo, tais como se o modelo foi testado e treinado para situações "inesperadas e não desejadas", e como ele deveria se comportar nestas hipóteses (Edwards; Veale, 2017, pp. 55-56).

Para que as MCE (Medidas de Conformidade Explicáveis) sejam eficazes, os autores sugerem que elas contemplem os seguintes pontos (Edwards; Veale, 2017, pp. 55/56):

a) *Informações sobre o Desenvolvimento do Modelo:* Detalhes sobre como o modelo foi criado, a intenção do processo de modelagem, o tipo de modelo utilizado e os parâmetros definidos.

O controlador deve esclarecer a intenção no processo de modelagem, o tipo de modelo e os parâmetros utilizados. Exemplo prático: O banco deve informar que utiliza um modelo de Árvore de Decisão ou Regressão Logística, desenvolvido com o objetivo específico de calcular a probabilidade de inadimplência (risco de calote) nos próximos 12 meses, priorizando a segurança financeira da instituição.

b) *Metadados de Treinamento:* Um resumo das estatísticas, descrição da qualidade dos dados usados no treinamento, a origem e o método de obtenção desses dados, além dos dados de saída ou das classificações realizadas.

Devem ser apresentados resumos estatísticos, a descrição da qualidade, origem e método de obtenção dos dados, bem como as classificações realizadas. Exemplo prático: Informar que o sistema foi treinado com uma base de dados de 500.000 clientes, contendo informações coletadas internamente (histórico de conta) e externamente (bureaus de crédito como SPC/Serasa), abrangendo dados demográficos, renda mensal e histórico de pagamentos dos últimos 5 anos.

c) *Métricas de Comportamento:* Apresentação dos resultados que demonstrem a capacidade do modelo de prever dados inéditos, incluindo a taxa de falhas e acertos.

Envolve apresentar resultados sobre a acurácia do sistema, sua habilidade de predizer dados nunca analisados, e suas taxas de falhas e acertos. Exemplo prático: O controlador deve demonstrar que o modelo possui uma taxa de acerto de 85% na previsão de bons pagadores, mas esclarecer que a taxa de falsos positivos (negar crédito indevidamente a um bom pagador) é de 10%, permitindo ao usuário entender a confiabilidade estatística da decisão.

d) *Explicações Humanamente Compreensíveis:* Descrições que permitam a um humano entender como dados de entrada específicos são transformados em dados de saída ou como influenciam o resultado.

Descrições sobre como certos dados de entrada são transformados em dados de saída (peso das variáveis). Exemplo prático: Esclarecer, em linguagem natural, que no modelo global do banco, a variável atraso em pagamentos recentes têm peso de 50% na decisão final, enquanto a idade do solicitante influencia apenas 5%, demonstrando quais fatores são preponderantes para a recusa ou aprovação do crédito.

e) *Informação Processual:* Dados sobre se o modelo foi testado e treinado para lidar com situações "inesperadas e não desejadas" e qual é o comportamento esperado para o modelo nesses cenários.

Dados sobre se o modelo foi testado para situações inesperadas e não desejadas e como deve se comportar. Exemplo prático: Informar como o algoritmo lida com situações atípicas, como a ausência total de histórico de crédito (o problema do *cold start*) ou mudanças econômicas bruscas (como uma pandemia), esclarecendo se, nessas hipóteses, o sistema é programado para negar o crédito automaticamente ou encaminhar para revisão humana (Edwards; Veale, 2017, pp. 55-56).

3.2.1.2 Explicações Centradas no Sujeito (Subject-Centric Explanation – SCE) ou Decisão Específica (Local) - micro explicabilidade - abrangência da explicação sobre a decisão específica - explicação como valor funcional e instrumental

Esta segunda modalidade, a Explicações Centradas no Sujeito (Subject-Centric Explanation – SCE) ou Decisão Específica (Local) trata-se de uma micro explicabilidade, reativa e individualizada, incidindo sobre a motivação da decisão específica que afeta o titular dos dados. Nesse sentido, a micro-explicabilidade exige a fundamentação, razões e circunstâncias que resultaram no desfecho individual (Wachter, et. al., 2017, p. 76-78).

Refere-se ao *rationale*, "*às razões e circunstâncias individuais de uma decisão automatizada específica, e.g. o peso atribuído às características, casos definidos por máquina e regras de decisões específicas, informações sobre grupos de referência ou perfil*" (Wachter et al., 2017, p. 6).

A interpretabilidade local descreve como um sistema chegou a um resultado, a partir de um *input* específico (Confalonieri *et al*., 2020; Phillips *et al*., 2021). Tais explicações são úteis para que os usuários afetados compreendam o porquê foram afetados.

Para que o escopo seja eficaz, ele deve ir além da mera descrição dos dados de entrada. Conforme a tese de Frajhof (2022) e a análise de Edwards e Veale (2017), a explicação deve ser acionável. O titular precisa de informações que o capacitem a tomar uma ação prática, seja contestar a decisão ou alterar seu comportamento para obter um resultado favorável.

Para estes autores, a abrangência ideal deve contemplar explicações contrafactuais, esclarecendo o que o indivíduo deveria ter feito para obter um

resultado diferente, e fatores críticos, indicando o peso e a relevância de cada variável individual na decisão. As explicações contrafactuais são vistas como altamente eficazes, pois auxiliam o titular a tomar uma ação futura e a contestar a decisão.

No cenário apresentado, o direito à explicação transcende a mera descrição técnica, possuindo um valor funcional e instrumental (Frajhof, 2022, 172/173; Selbst; Barocas, 2016; Selbst; Powles, 2017). Uma explicação adequada deve ser útil (*meaningful*) para que o titular possa:

a) Exercer seus direitos assegurados pela LGPD ou GDPR;
b) Contestar a decisão;
c) Avaliar a legitimidade e legalidade do processo e dos resultados algorítmicos, verificando erros ou discriminações;
d) Adaptar seu comportamento para obter um resultado distinto no futuro.

De acordo com Frajhof (2022, p. 153/154), as SCEs são explicações limitadas a pequenas partes do sistema, variando conforme a pergunta e respondendo com informações apenas dos dados de entrada. Diferente de Wachter *et al*. (2017), Edwards e Veale defendem que esse tipo de explicação pode ser fornecido tanto antes quanto depois da decisão, exigindo apenas a apresentação dos dados usados para responder à questão.

As Explicações Contrafactuais Esparsas (SCEs) podem ser classificadas em quatro tipos (Edwards; Veale, 2017, p. 58): (i) baseada na sensibilidade (sensitivity-based), que busca identificar quais os dados de entrada a serem alterados para modificar a decisão; (ii) baseada no caso (case-based), que visa encontrar dados de treinamento similares ao do indivíduo; (iii) baseada na demografia (demographic-based), que busca sujeitos com tratamento semelhante; e (iv) baseada no comportamento (performance-based), que visa informações sobre a confiança estatística da decisão, incluindo casos similares com resultados esperados diferentes.

Na perspectiva de Doshi-Velez e Kortz (2017), embora as explicações SCEs sejam menos eficazes para questões de regularidade processual, elas são benéficas para os usuários afetados. Contudo, existem limitações técnicas significativas: modelos muito complexos, com um grande volume de dados de entrada, podem inviabilizar a geração de explicações causais. Além disso, explicações locais, que

cobrem apenas partes do sistema, podem ser insuficientes para explicar casos atípicos (*outliers*). Por fim, usuários com perfis muito diferentes da média podem não receber explicações adequadas sobre a lógica fundamental do modelo.

Nunes (2023) reforça que o uso da IA Explicável (XAI) é a ferramenta necessária para atender a esse escopo, transformando a complexidade do *machine learning* em informação inteligível que justifique a decisão individual.

Em resumo à matéria abordada neste tópico, vale conferir Acioly (2023, p. 236):

> Ademais, a interpretabilidade global tem sua funcionalidade direcionada à descrição do comportamento do modelo, o que incluir a compreensão das categorias de dados que são necessárias e a descrição detalhada do funcionamento do (CONFALONIERI et al., 2020; PHILLIPS et al., 2021). A interpretabilidade local, por sua vez, descreve na ótica individual como um sistema chegou a dado resultado, a partir de um input específico (CONFALONIERI et al., 2020; PHILLIPS et al., 2021). A utilidade das abordagens de inteligibilidade varia de acordo com a necessidade a qual se destinam (PEREIRA, 2021), ponderando os interesses que melhor se adequam ao caso concreto, a partir de uma dinâmica de sopesamento (KROLL et al., 2017).

Como visto, a classificação quanto à abrangência da explicação muito se relaciona com o momento de aplicação do direito, que pode se dar antes ou depois da tomada da decisão automatizada. Assim, cumpre prosseguir com a discussão nesse aspecto.

3.2.2 *Quanto ao momento de aplicação*

A análise da explicabilidade algorítmica transcende a mera discussão sobre sua natureza jurídica, exigindo a classificação de seu momento de aplicação no ciclo de vida de um sistema de Inteligência Artificial. Essa segmentação, fundamental para a operacionalização regulatória, divide a exigência de transparência em explicações *ex ante* e *ex post* em relação ao processo de tomada da decisão automatizada.

A classificação temporal é crucial para determinar o tipo de informação que pode ser exigida, distinguindo-se a informação prévia (funcionalidade) da justificativa posterior (decisão específica) (Toniazzo, 2021, p. 62/63; Frajhof, 2022, p. 70).

3.2.2.1 Ex ante

A explicação *ex ante* refere-se à demanda por inteligibilidade que antecede a tomada de decisão automatizada específica, possuindo uma função eminentemente preventiva. É a explicação fornecida antes de uma decisão automática ocorrer.

Conforme Franzolin *et al* (2024, p. 74), uma explicação prévia pode ser solicitada antes de uma decisão automatizada, mas aborda apenas a funcionalidade geral do sistema, já que a justificativa de uma decisão específica só é conhecida após sua efetivação. Essa modalidade, portanto, está necessariamente relacionado o conteúdo à funcionalidade geral do sistema (MCE), à lógica, aos critérios e às consequências previstas, já que a decisão individual ainda não foi produzida, implicando na divulgação dos parâmetros, das variáveis de treinamento e dos limites de operação do modelo algorítmico (Toniazzo, 2021, p. 61; Franzolin, 2023, p. 74; Costa, 2021, p. 69).

No GDPR, os deveres de notificação (Arts. 13 e 14) se restringem a este momento *ex ante* (Edwars; Veale; 2017, p. 51-52)

Nesse cenário, a principal função pretendida é de garantir que a arquitetura do sistema esteja em conformidade com os princípios éticos e legais desde a fase de *design*, possibilitando a auditoria sistêmica e a mitigação proativa de riscos de discriminação e injustiça. Na prática regulatória, essa exigência se materializa nas obrigações de documentação e nas Avaliações de Impacto à Proteção de Dados (DPIAs/RIPDs), como elementos de governança algorítmica e transparência institucional.

3.2.2.2 Ex post

Em contraposição, a explicação *ex post* é de natureza reativa e ocorre após a concretização de uma decisão automatizada que gerou consequências sobre o indivíduo. Segundo Wachter *et. al*. (2017, p. 7678), a explicação *ex post* pode abranger tanto a funcionalidade geral do sistema - como a explicação *ex ante* - quanto às questões relativas à decisão específica tomada em relação ao titular dos dados (o fundamento, as razões e circunstâncias da decisão específica - SCE (Frajhof, 2022, p. 69; Costa, 2021, 70).

Esse tipo de explicabilidade é o que confere substância ao direito de contestação e ao exercício da autodeterminação informativa do indivíduo, permitindo que o afetado compreenda os motivos pelos quais foi negado um crédito ou o porquê teve um pedido indeferido. A aplicação da explicação oferecida *ex post* colabora para que a "composição e funcionamento do sistema se torne mais compreensível para o leigo, atenuando o tipo de nebulosidade que surge da falta de conhecimento técnico exigido pela tecnologia" (Franzolin, 2023, p. 10).

Entretanto, Almada (2019) ressalta que, embora as técnicas de explicação busquem apresentar o algoritmo na forma mais simples possível, essa redução de complexidade pode não ser suficiente para transformar o sistema em algo plenamente compreensível para qualquer indivíduo. Essa limitação técnica não anula o dever de explicar, mas aponta para a necessidade de que a legislação determine padrões de explicabilidade adequados ao receptor e proporcionais ao risco da decisão, o que a Lei da IA Europeia e o PL 2338/2023 brasileiro buscam endereçar. Desse modo, a dicotomia *ex ante*/*ex post* serve para demonstrar a exigência de uma explicabilidade multinível na proteção dos direitos fundamentais.

Aliás, o debate em comento deve ser compreendido dentro de uma discussão mais abrangente de *accountability* (prestação de contas) de sistemas de inteligência artificial (Kaminski, 2019). Por conseguinte, a garantia e a efetivação do direito à explicação exigem uma abordagem abrangente e diversificada, que se manifesta na exigência de documentação, justificação e monitoramento do desenvolvimento, da implementação e da aplicação desses sistemas, em momentos que antecedem (*ex ante*) e sucedem (*ex post*) a tomada de decisão. Essa perspectiva integradora reconhece que a explicabilidade não é um evento isolado, mas sim um processo contínuo e essencial para assegurar a responsabilidade e a conformidade ética e legal da IA.

3.2.2.3 Quanto ao momento da explicação nos ordenamentos jurídicos

A questão sobre se o GDPR consagra um direito à explicação *ex post* (sobre a decisão específica) é objeto de intenso debate.

Wachter et al (2017) argumentam em favor de uma tese restritiva, cuja interpretação dos artigos relevantes do GDPR (13, 14 e 15) não preveem a explicação *ex post* de decisões específicas.

Nos Arts. 13 e 14, as informações são fornecidas no momento da coleta e se limitam à funcionalidade. Mesmo no Art. 15 (Direito de Acesso), a expressão "consequências previstas" sugere informações orientadas para o futuro, indicando que apenas a funcionalidade do sistema deve ser explicada, e não a justificativa da decisão específica já tomada (Wachter, et. al., 2017, p. 83; Costa, 2021, p. 70; Franzolin, 2025. p. 80).

Por outro lado, a tese ampliativa, defendida por Selbst e Powles, sustenta que o Art. 15, o qual trata do direito de acesso, pode ser invocado a qualquer momento, inclusive após a decisão, tornando plausível a exigência de uma explicação *ex post* sobre o *rationale* da decisão específica. Além disso, o Considerando 71 (embora não vinculativo) reforça o direito de obter uma explicação sobre a decisão tomada (Costa, 2021, p. 71).

O Artigo 20 da LGPD não especifica o momento em que a explicação pode ser exigida, não se atendo à discussão temporal do GDPR. Contudo, a interpretação sistemática da LGPD (e o princípio da transparência) sugere que a explicabilidade deve ser garantida tanto *ex ante* (funcionalidade do sistema) quanto *ex post* (fundamentos da decisão) (Toniazzo, 2021, p. 65; Freitas, 2020, p. 79).

3.3 Quanto à titularidade e exigibilidade do direito de explicação

Relativamente à titularidade do direito à explicação, segundo Frajhof (2022, p. 204/205), este é conferido ao titular dos dados e pode ser exercido por ele mesmo ou por seu representante (art. 18, § 3º, da LGPD), diretamente ao controlador, sem ação judicial. Para sua efetiva aplicação, o controlador deve estar estruturado para receber pedidos e notificar o titular sobre a coleta e uso de dados em decisões automatizadas. A inobservância desses deveres pode configurar aplicação inadequada do direito.

Já no que se refere à exigibilidade, caso o controlador se recuse a fornecer as informações, o titular pode ajuizar ação judicial competente, inclusive pleiteando a inversão do ônus da prova (art. 42, § 2º, da LGPD). Na referida ação, o magistrado pode determinar o compartilhamento, sob sigilo judicial, de documentos que permitam compreender o algoritmo. Nada obstante, frise-se ser cabível também ação coletiva (art. 22, da LGPD).

O não atendimento da determinação de explicação do algoritmo pode gerar

dever de indenizar (art. 42, da LGPD) e, em ação coletiva, a determinação de implementação de medidas de transparência e *accountability*. Mas não é só. A recusa em fornecer explicações (art. 20, § 1º) pode levar à demanda de auditoria perante a ANPD (art. 20, § 2º).

Alternativamente à via judicial, o titular dos dados ou seu representante legal podem apresentar petição à ANPD (art. 18, § 1º). Neste caso, pode ser exigida uma gama maior de documentos (*ex ante* e *ex post*), sendo que o não atendimento ao direito à explicação pode resultar na aplicação de sanções administrativas pela ANPD, variáveis conforme a gravidade (art. 52, da LGPD).

Destaca-se que, embora a LGPD se refira ao titular dos dados pessoais, o Projeto de Lei nº 2338/2023 (Marco Legal da IA) propõe uma abordagem mais ampla e abrangente. Este PL permite que a pessoa afetada possa solicitar explicação sobre qualquer sistema de IA, independentemente do nível de risco oferecido, e sem se limitar a casos que gerem efeitos jurídicos relevantes.

Da observação dos ordenamentos jurídicos estudados, é possível notar a existência de uma diferença fundamental na natureza da proteção concedida pela legislação europeia e brasileira, que impacta a exigibilidade do direito. Isto é, o Artigo 22 do GDPR trata a decisão automatizada que produza efeitos jurídicos ou afete significativamente o titular, como uma proibição geral, a não ser que se enquadre em exceções (necessidade contratual, autorização legal ou consentimento explícito) (Toniazzo, 2021, p. 61 e 68). O direito à explicação, tal qual abstraído do Considerando 71, é, portanto, uma salvaguarda adicional implícita, garantida nas hipóteses de exceção (Lima, 2020, p. 234).

Outrossim, na LGPD (Art. 20), não há proibição da decisão automatizada, mas confere ao titular o direito a solicitar a revisão e impõe ao controlador o dever de fornecer informações claras e adequadas sobre os critérios e procedimentos (Lima, 2020, p. 232; Toniazzo, 2021, p. 62). A explicação é uma condição de eficácia do princípio da transparência (Toniazzo, 2021, p. 65).

3.4 Aspectos de limitação jurídica existente nos ordenamentos jurídicos quanto ao exercício do direito à explicação

Como vimos, a consagração do direito à explicação é um esforço regulatório para mitigar os riscos e a assimetria informacional inerentes à opacidade dos algoritmos de Inteligência Artificial (IA). No entanto, a sua implementação prática

encontra óbices significativos que podem ser categorizados em duas vertentes principais: limitações jurídicas, decorrentes de conflitos com outros direitos legalmente tutelados, e limitações intrínsecas à própria natureza e complexidade da tecnologia.

3.4.1 *Limitação jurídica do direito à explicação*

A tentativa de impor transparência irrestrita aos sistemas algorítmicos colide com o arcabouço protetivo de outros bens jurídicos no ordenamento, em particular aqueles relacionados à atividade econômica e à inovação.

3.4.1.1 Sigilo industrial e propriedade intelectual

A principal limitação legal à exigibilidade e abrangência do direito à explicação reside na proteção ao segredo comercial e industrial. Nesse sentido, a proteção da propriedade intelectual atua como um contraponto à transparência irrestrita dos mecanismos de tratamento de dados.

Esta proteção visa garantir a competitividade do negócio, protegendo *know-how*, metodologias de produção, e fórmulas, o que inclui os algoritmos e códigos-fonte que compõem o "*core business*" da empresa.

Dessa limitação, como visto, surge uma opacidade que pode ser considerada bastante vantajosa para os agentes econômicos que controlam o código, uma vez que o sigilo, juridicamente amparado, protege seus interesses (Frazão; Gottenauer, 2020, p. 53).

Na verdade, Frank Pasquale (2015), ao cunhar o termo "Black Box Society", argumenta que a opacidade e a falta de transparência não são meramente características intrínsecas, mas sim o resultado da ação deliberada de agentes econômicos ou estatais, que utilizam estratégias jurídicas (como a proteção do segredo de negócios) para manter a opacidade e proteger sua valorosa propriedade intelectual.

No GDPR, o Considerando 63 esclarece que o direito de o titular de dados conhecer a lógica subjacente não deverá prejudicar os direitos e liberdades de terceiros, "*incluindo o segredo comercial ou a propriedade intelectual e, particularmente, o direito de autor que protege o software*". Contudo, o mesmo

Considerando determina que "*essas considerações não deverão resultar na recusa de prestação de todas as informações ao titular dos dados"*.

A LGPD, por sua vez, ao consagrar o princípio da transparência, estabelece a ressalva de que as informações fornecidas devem observar os segredos comercial e industrial (Art. 6º, VI). Da mesma forma, o dever do controlador de fornecer informações claras e adequadas sobre os critérios e procedimentos de decisões automatizadas (Art. 20, § 1º) também resguarda o sigilo.

Observa-se, portanto, que a proteção jurídica concedida ao sigilo industrial e à propriedade intelectual tem sido frequentemente utilizada como uma justificativa para a ausência de explicações. Paradoxalmente, significa dizer que o direito pode contribuir para a opacidade.

A proteção legal conferida aos algoritmos sob o argumento do segredo empresarial e/ou propriedade intelectual tem sido alvo de severas críticas por parte de pesquisadores (Watcher; Mittelstadt, 2019; Pasquale, 2015; Frazão, 2021). Isso ocorre porque essa defesa legal tem sido usada para impedir a investigação do artefato e dos fundamentos de suas decisões.

Diante desse dilema, Pasquale (2015, p. 142) propõe a transparência qualificada (*qualified transparency*), em que a explicação, embora não seja pública e irrestrita, deve ser disponibilizada a auditores confiáveis e agências reguladoras para verificar a legalidade e a ausência de discriminação. Para o autor, esta medida respeitaria os interesses das partes envolvidas, ou seja, dos agentes econômicos e daqueles impactados pelas decisões dos algoritmos.

A propósito, outros autores endossam tal raciocínio (Scherer, 2016; Zarsky, 2013), sob fundamento de que a abertura do código nestes termos não implicaria em abertura indiscriminada do conteúdo ao público em geral

A proteção ao segredo comercial e industrial, a qual é mencionada treze vezes na LGPD, gera tensão com a proteção de dados pessoais, inclusive no princípio da transparência (art. 6º, VI), que exige informações claras, mas ressalva os segredos. Essa restrição visa equilibrar interesses econômicos com os direitos dos titulares, entretanto não pode justificar a ausência de explicações sobre algoritmos, como tem sido usada por agentes econômicos como justificativa legal para negar o compartilhamento de qualquer informação relativa ao algoritmo.

Isto é, a doutrina sustenta que o segredo empresarial não pode ser considerado absoluto e deve ceder diante de direitos fundamentais e interesses

sociais relevantes, como a proteção contra discriminação e a garantia do devido processo (Frazão, 2021).

Por fim, insta salientar que, na tentativa de equilibrar o conflito em comento, a LGPD prevê que, em caso de recusa do controlador baseada no segredo comercial ou industrial, o titular poderá solicitar à Autoridade Nacional de Proteção de Dados (ANPD) a realização de auditoria para verificar a ocorrência de aspectos discriminatórios no tratamento automatizado (§ 2º do Art. 20). O não atendimento do direito à explicação pode gerar o dever de indenizar e acarreta a aplicação de sanções administrativas pela ANPD.

É importante notar que a LGPD limita o direito de requerer auditoria, pois a possibilidade de a ANPD realizá-la é discricionária, não obrigatória, e está restrita à hipótese de recusa por segredo comercial/industrial. Outras situações, como investigar a veracidade da explicação, seriam igualmente merecedoras de auditoria.

3.4.1.2 Limitação intrínseca - complexidade técnica da IA

Mesmo que a barreira jurídica do segredo industrial seja superada, a natureza intrínseca de certos sistemas de IA impõe obstáculos técnicos que comprometem a viabilidade de uma explicação compreensível, transformando a explicabilidade em um problema não apenas legal, mas também epistemológico e técnico.

Como se sabe, a complexidade técnica é uma das três formas de opacidade identificadas na literatura de Jenna Burrel (2016). Esta modalidade é característica de algoritmos avançados, como os de *machine learning* e *deep learning*, especialmente aqueles que utilizam redes neurais.

Em regra, decorre do processamento de um volume incomensurável de dados (*big data*) e da capacidade de o algoritmo de resolver problemas de otimização - ou seja, de adaptar ou readaptar sua lógica de decisão interna à medida que aprende com os dados de treinamento. Tal desenvolvimento torna o sistema “inescrutável", dificultando a compreensão humana do processo decisório.

Essa característica se manifesta no fato de seu funcionamento ser invisível para o titular dos dados, além de que, “*por vezes, é também opaco para quem o concebeu*” (Cf. EuropeAN DATA PROTECTION SUPERVISOR, Opinion 3/2018..., cit., p. 9 e Giovanni Sartor (autor); Parlamento Europeu, Artificial Intelligence, p. 6 *in* Costa, 2021, p. 45). Nesse sentido:

> Nestas hipóteses, os algoritmos inteligentes são inescrutáveis porque a complexidade das suas regras desafia a compreensão humana, mesmo a mais especializada. Esta nuvem que paira sobre os algoritmos entorpece a sua fiscalização e controlo e dificulta a própria conceção de um regime legal verdadeiramente adequado. A opacidade do algoritmo é o primeiro véu que deve ser levantado para a correção de todos estes problemas. Os resultados incorretos, injustificados e injustos geram múltiplas ameaças à pessoa, entre as quais a manipulação (COSTA, 2021, p. 45).

Em suma, os aspectos de limitação jurídica do direito à explicação são definidos pelo *trade-off* entre a necessidade de transparência para a proteção da personalidade humana e a proteção legal do segredo de negócios, somada ao desafio técnico de traduzir a lógica de sistemas complexos de *machine learning* para uma linguagem que seja inteligível e útil ao cidadão.

3.4.2 O devido processo no âmbito privado e a explicabilidade algorítmica

Diante da capacidade de decisões algorítmicas causarem impactos nocivos e discriminatórios na vida das pessoas, o direito à explicação se consolida como uma garantia instrumental e funcional indispensável para assegurar a tutela da personalidade humana.

O Devido Processo Algorítmico (DPA) é um conceito que busca traduzir as garantias processuais do Estado de Direito para o contexto das decisões automatizadas e que tem sido defendida especialmente na literatura norte-americana, a qual, a fim de evitar uma tirania de decisões automatizadas, salienta a importância de se garantir um devido processo algorítmico, (Kaminski, 2019b; Citron; Pasquale, 2014; Crawford; Shultz, 2014).

Na prática, a explicabilidade atua como a condição *sine qua non* para que o titular possa exercer o controle efetivo sobre seus dados pessoais e contestar decisões.

As fontes identificam que o DPA deve abranger um conjunto de salvaguardas procedimentais essenciais, quais sejam:

3.4.2.1 *Notificação e ciência*

O sujeito deve ser notificado de que está sujeito a uma decisão algorítmica e ter conhecimento de que o direito de não sujeição lhe é conferido. Este dever de notificação e informação decorre dos artigos 13º e 14º do GDPR, que exigem informações ex ante (antes da decisão) relativas à lógica subjacente e às consequências previstas de tal tratamento.

3.4.2.2 *Manifestação e revisão humana*

O titular deve ter a oportunidade de manifestar o seu ponto de vista e obter intervenção humana. A intervenção humana é vista como crucial para criar e manter a confiança, e deve ser substancial para descaracterizar a decisão como exclusivamente automatizada, evitando a "fabricação" de envolvimento humano ineficaz.

3.4.2.3 *Contestação*

O direito de contestar a decisão impõe, no mínimo, a obrigação de reavaliar o resultado. O direito à explicação é, portanto, o substrato necessário para a contestabilidade da decisão algorítmica.

3.5 Análise propositiva dos horizontes regulatórios para a opacidade algorítmica

Considerando a contínua discussão que permeia o avanço da regulação da inteligência artificial e o enfrentamento da opacidade algorítmica - que cada vez mais se mostra latente e necessária, como abordado ao longo deste trabalho.

Para tanto, a exigência de um esforço propositivo, que transcenda a mera descrição da tecnologia, é imperativo. Este capítulo se dedica a analisar as fronteiras e as tensões jurídicas na busca pela efetivação do direito de explicação, examinando o diálogo complexo entre a Lei Geral de Proteção de Dados e o modelo apresentado pelo Regulamento Geral de Proteção de Dados europeu, adentrando ao real papel da discussão sobre a transparência como solução e os direcionamentos regulatórios mais recentes na Europa e no Brasil.

3.5.1 A (im) possibilidade de adequação do direito à explicação no diálogo entre direito interno e legislação internacional (ênfase na regulação europeia)

Como mencionado no capítulo introdutório deste trabalho, escolheu-se a comparação da legislação brasileira com a europeia em razão do vanguardismo regulatório da UE (GDPR, AI Act), que estabelece tendências globais e influencia diversas regiões, entre elas o Sul Global. Contudo, de proêmio elucidou-se a intenção de se proceder a análise comparativa considerando o impacto dessas normas nos países periféricos - que se tornam consumidores de tecnologias e padrões regulatórios do Norte Global, reforçando assimetrias e uma lógica neocolonial.

A fim de evitar a presunção de que as experiências legislativas dos países do norte global seriam de alguma forma mais representativas ou importantes que as de outras localidades, propôs-se que o estudo comparativo partisse da ponderação de aspectos epistemológicos e metodológicos. Assim, a análise seguiu a partir da aplicação dos filtros apresentados por Hirschl (2014), visando não apenas analisar um modelo dominante, mas também questioná-lo a partir da reflexão sobre como as regulações afetam a soberania digital e a autonomia regulatória do Sul Global, bem como destacando sua falta de protagonismo na definição dessas normas.

Dito isto, a presente análise transcende a mera dimensão jurídico-formal para adentrar na dimensão ética e social da crescente influência dos sistemas de inteligência artificial.

3.5.2 Cumprimento dos Parâmetros Metodológicos de Hirschl

3.5.2.1 Análise contextual

De plano, a comparação do direito à explicação entre a LGPD e o GDPR exige o reconhecimento de que os sistemas jurídicos operam em contextos distintos.

Ao destacar a incomensurável influência das políticas públicas europeias em outros países, em especial quanto à regulação de novas tecnologias, Bradford (2020) nomeou como “Efeito Bruxelas”. Este refere-se à vanguarda regulatória global assumida pela União Europeia, que passa a influenciar outras jurisdições ao redor do mundo.

Dentre os influenciados, surge o Brasil, como nação do Sul Global, com particularidades culturais, políticas e sociais, mas que herda e adapta este modelo europeu.

No que se refere especificamente ao direito de explicação, conforme visto no capítulo 2 (“Fontes heterônomas sobre a proteção de dados e informações da pessoa”), sua análise contextual nos ordenamentos jurídicos em debate revela a diferença na natureza do direito:

O GDPR, apesar de estabelecer uma proibição geral de o titular de dados ser submetido a decisões exclusivamente automatizadas que o afetem significativamente (artigo 22º/1), o direito à explicação, tal como discutido no GDPR, é extraído implicitamente, a partir de uma interpretação sistemática dos Artigos 13, 14 e 15, e serve como uma salvaguarda para as exceções à proibição.

Na LGPD, por sua vez, o legislador brasileiro não proibiu o titular de ser submetido à decisão automatizada, mas sim oportunizou um conjunto de direitos decorrentes desse tratamento, mas confere direitos ao titular, como a explicação subjacente ao direito de revisão, os quais atuam como os principais mecanismos de proteção (Toniazzo; Barbosa; Ruaro, 2021, p.65). Entretanto, assim como a legislação europeia, a LGPD é omissa sobre o que constitui uma decisão baseada exclusivamente no tratamento automatizado.

Para suprir essa lacuna, o Projeto de Lei n° 4.496, de 2019 propõe que decisões automatizadas são aquelas baseadas em regras, algoritmos ou inteligência artificial que oferecem pouca ou nenhuma compreensão adicional sobre o tema (Brasil, 2019).

3.5.2.2 Diversidade de Fontes

Relativamente à diversidade das fontes de pesquisa, o rigor metodológico é demonstrado *in casu* ao não se limitar aos textos legais originais (GDPR e LGPD), mas incorporar uma diversidade de fontes que se estendem para a *hard law* (em especial o PL 2338/2023) e a *soft law* (Considerandos, orientações de Grupos de Trabalho, e literatura multidisciplinar).

De fato, essa diversidade é crucial para abordar a opacidade algorítmica, que não é apenas um desafio jurídico (segredo comercial), mas também um problema técnico (inescrutabilidade de *deep learning*) e social (analfabetismo técnico). A

análise, portanto, se apoia em disciplinas como a ciência da computação (XAI/explicabilidade) e as Ciências Sociais (colonialismo de dados).

3.5.2.3 Prática de “Cherry-Picking” (seleção de casos)

Em cumprimento à prática de “cherry-picking”, utilizando o princípio de seleção de "casos mais semelhantes", a metodologia comparativa se concentra na dialeticidade entre a LGPD e o GDPR, na medida em que ambas as leis são diplomas gerais de proteção de dados.

Em relação aos "casos mais diferentes", o filtro é cumprido em razão da abordagem sobre o direito à explicação e a natureza da proteção. Isto é, considera-se a hipótese da pesquisa: a existência do direito à explicação em ambos, mas com aplicação e interpretação significativamente variáveis.

Nesse sentido, o *cherry-picking* é direcionado e sistemático, focado na tensão do conteúdo da explicação:

(a) Conteúdo *ex post*

Observa-se que, apesar da LGPD exigir informações sobre "critérios e procedimentos" utilizados para a decisão, sua redação não é explícita quanto à necessidade de conhecimento *ex post* e as consequências, diferentemente do que ocorre nos no GDPR, nos artigos 13, 14 e 15 (Toniazzo, 2021, 65).

(b) Controvérsia temporal

O debate europeu concentra-se, especialmente, em saber se os referidos artigos 13, 14 e 15 exigem informações úteis *ex ante* (funcionalidade geral do sistema) ou se permitem informações *ex post* (sobre a decisão específica). Nesse aspecto, a tese mais restritiva (*Wachter et al.*) argumenta que o direito à explicação de uma decisão individual só está no Considerando 71 (não vinculativo), defendendo a inexistência de um direito explícito à explicação *ex post*.

3.5.3 Reflexões Axiológicas

A análise comparativa se completa com a dimensão ética e a reflexão axiológica, essencial para combater a imposição acrítica de padrões jurídicos do Norte Global.

Nesse sentido, a regulação deve se atentar à colonialidade do poder e ao colonialismo de dados. As *Big Techs*, sediadas no Norte Global, utilizam o sigilo (opacidade intencional) e a propriedade intelectual para manter a assimetria informacional. Essa racionalidade centrada em dados busca impor uma percepção de neutralidade que é, na verdade, a expressão da colonialidade.

Ademais, ressalta-se a exigência axiológica da explicabilidade, portanto, é a chave para proteger os direitos fundamentais como o devido processo legal e a igualdade, além de assegurar a dignidade da pessoa humana e a autodeterminação informativa. Negar a explicação em nome do segredo comercial ou da inescrutabilidade técnica é manter a subordinação do Sul Global, que se encontra na condição de colônia de dados.

Em resumo, constata-se, portanto, que o cerne da assimetria regulatória reside na natureza jurídica do direito conferido ao titular dos dados.

Isso porque, o GDPR (22º do GDPR) estabelece, em regra, uma proibição de o titular dos dados ser submetido a uma decisão tomada exclusivamente com base em tratamento automatizado. As exceções a essa proibição (como consentimento explícito ou necessidade contratual) desencadeiam a aplicação de salvaguardas. A existência do Direito à Explicação é, para muitos, extraída de uma interpretação sistemática e teleológica do diploma, sendo um direito implícito (corolário do princípio da transparência e do direito de acesso). O direito de oposição/não sujeição não é absoluto, mas o Artigo 22º exige medidas adequadas, como o direito de obter intervenção humana, manifestar o ponto de vista e contestar a decisão.

Por outro lado, a LGPD adota abordagem distinta, não proibindo que o titular de dados seja submetido à decisão automatizada. Em vez disso, o Artigo 20º consagra um direito à revisão da decisão automatizada, sendo a explicação o principal mecanismo de proteção e um conjunto de prerrogativas que decorrem da autodeterminação informativa

No tocante ao conteúdo, persiste a controvérsia sobre a densidade da explicação, o GDPR exige "informações úteis relativas à lógica subjacente", o que, para a tese restritiva, se limita a informações *ex ante* (sobre a funcionalidade geral do sistema).

Diferentemente, a LGPD exige "informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados", uma redação que, embora similar ao GDPR, não traz explicitamente a necessidade de conhecimento sobre a lógica subjacente.

Por fim, a inadequação da LGPD em face da complexidade algorítmica e dos padrões internacionais é mais evidente no que tange à revisão humana. O Artigo 20º da LGPD sofreu um veto presidencial que suprimiu a obrigatoriedade de que a revisão fosse realizada por "pessoa natural". A justificativa do veto era a de que essa exigência contrariava o interesse público e inviabilizaria os modelos de negócios de muitas empresas, além de impactar negativamente a oferta de crédito.

Essa ausência de vedação expressa à revisão por não-humano torna o direito à explicação questionável. O Artigo 22º do GDPR, em contraste, prevê o direito de obter intervenção humana como salvaguarda. A limitação da LGPD é vista como um empecilho e óbice para o pleno exercício do direito à explicação.

3.5.4 Transparência como solução à opacidade algorítmica

A transparência é amplamente reconhecida como um meta-princípio e um pilar basilar na regulação da inteligência artificial (Acioly, 2023, p. 246), atuando como pressuposto para a concretização dos princípios da Prestação de Contas (*accountability*) e Justiça (*fairness*) (Frajhof, 2022, p. 86). A Lei Geral de Proteção de Dados (LGPD) define transparência como a "garantia, aos titulares, de informações claras, precisas e facilmente acessíveis".

No entanto, a análise propositiva demonstra que a mera transparência, embora necessária, é insuficiente, exigindo outros mecanismos para avaliar a prestação de contas e a responsabilização de algoritmos de tomada de decisão (Ananny e Crawford, 2018, p. 984).

A explicação de decisões algorítmicas difere da transparência: esta requer a abertura de informações do sistema de IA, enquanto aquela visa detalhar a influência de fatores no resultado do algoritmo (Doshi-Velez; Kortz, 2017, p. 6). Ambas buscam viabilizar a prestação de contas e a responsabilização dos desenvolvedores e usuários do algoritmo.

A eficácia do direito à explicação, portanto, depende da identificação da natureza da opacidade a ser combatida.

A opacidade não é um fenômeno singular, mas um desafio multifacetado, que exige soluções técnicas e jurídicas (Frajhof, 2022, p. 95). Como já explicado anteriormente, a opacidade é classificada em três formas principais: o sigilo intencional (para proteger o segredo de negócio e propriedade intelectual, o mais difícil de contornar), o analfabetismo técnico (dificuldade do leigo em compreender o código), e a opacidade intrínseca ou inescrutabilidade dos modelos de aprendizado de máquina, que geram resultados não-intuitivos e complexos até mesmo para seus desenvolvedores.

No que se refere à opacidade deliberadamente promovida por agentes econômicos ou estatais. Essa opacidade é mantida por estratégias jurídicas (o que (Pasquale, 2015, p. 6/7, chama de *legal secrecy*) para proteger o segredo comercial e industrial e a propriedade intelectual do código. A proteção ao segredo corporativo é o limiar mais difícil de traçar, mas é constantemente utilizada para justificar a recusa de informações e impedir o acesso ao código para avaliação e investigação.

Para a opacidade por analfabetismo técnico, que resulta da incapacidade de o público em geral compreender a linguagem computacional, a mera apresentação do código-fonte é inútil para o usuário final, pois o código é uma habilidade especializada. Esta barreira exige esforços de longo prazo, como educação e tradução da linguagem matemática para a linguagem natural.

No caso da opacidade algorítmica intrínseca (inescrutabilidade), esta é a forma mais desafiadora. Oriunda da complexidade inevitável e inescrutabilidade de algoritmos de Aprendizado de Máquina (ML) e Aprendizado Profundo (Deep Learning), na qual o algoritmo altera sua estrutura comportamental e pode gerar resultados complexos, volumosos, heterogêneos e não-intuitivos. Esta opacidade intrínseca dificulta a compreensão do processo decisório até mesmo para os desenvolvedores.

3.5.5 Parâmetros de transparência - ser humano

O debate sobre a explicabilidade algorítmica frequentemente parte da premissa de que o julgamento humano é transparente ou, no mínimo, mais explicável do que a máquina (Frajhof, 2022, p. 149). Contudo, torna-se imperativa a reflexão sobre o referencial comum de mensuração: se um sistema é tido como mais opaco que o outro, qual a medida usada para aferir o primeiro?

Isto é, a opacidade da mente humana, muitas vezes ignorada, é um ponto crucial de comparação. A fundamentação escrita pelo ser humano não é necessariamente o racional exato que se passou em sua mente. Isso ocorre porque os humanos racionalizam e apresentam justificativas ex post para explicar suas ações, sendo que os reais motivos que desencadeiam o pensamento ou comportamento não são acessíveis à inspeção (Cushman, 2020; Mercier; Sperber, 2017; Klein, 2018; Kahneman, 2021).

Em sua obra “Ruído: Uma falha no julgamento humano”, Kahneman (2021) fornece o referencial comparativo necessário para quantificar as falhas inerentes ao julgamento humano, desafiando a percepção de que os humanos são intrinsecamente justos e transparentes. Segundo o autor, o erro no julgamento humano é dividido em dois componentes distintos: viés (*bias*) e ruído (*noise*).

O viés, segundo o autor, é um erro sistemático e previsível, um desvio consistente na mesma direção (e.g., uma balança sempre marcando 200g a mais), enquanto o ruído refere-se à variabilidade indesejada, dispersão aleatória e inconsistente do resultado quando se espera uma resposta próxima.

Como exemplo, Kahneman (2021, p. 18) apresenta estudos mostram que decisões judiciais são influenciadas por fatores irrelevantes e emocionais. Juízes tendem a ser mais lenientes no início do dia ou após pausas, sendo "juízes com fome mais inclementes"[15]. Decisões são mais severas às segundas-feiras após derrotas do time do coração do juiz, afetando desproporcionalmente réus negros[16]. Preferências esportivas[17] e até o aniversário do réu (com mais leniência nessa data, observada na França) também influenciam[18]. Além disso, o clima impacta: temperaturas mais altas reduzem as chances de asilo, sugerindo que é melhor ter a audiência em um dia ameno[19].

---

[15] Shai Danziger, Jonathan Levav e Liora Avnaim-Pesso, “Extraneous Factors in Judicial Decisions”, Proceedings of the National Academy of Sciences of the United States of America, v. 108, n. 17, 2011, pp. 6889- 92.

[16] Ozkan Eren e Naci Mocan, “Emotional Judges and Unlucky Juveniles”, American Economic Journal: Applied Economics, v. 10, n. 3, 2018, pp. 171-205.

[17] Daniel L. Chen e Markus Loecher, “Mood and the Malleability of Moral Reasoning: The Impact of Irrelevant Factors on Judicial Decisions”, SSRN Electronic Journal, 21 set. 2019, p. 170. Disponível em: <http://users.nber.org/dlchen/papers/Mood_and_the_Malleability_of_Moral_Reasoning.pdf>.

[18] Daniel L. Chen e Arnaud Philippe, “Clash of Norms: Judicial Leniency on Defendant Birthdays”, 2020. Disponível em : <https://ssrn.com/abstract=3203624>.

[19] Anthony Heyes e Soodeh Saberian, “Temperature and Decisions: Evidence from 207,000 Court Cases”, American Economic Journal: Applied Economics, v. 11, n. 2, 2018, pp. 238-65.

Nota-se, portanto, que o ruído é um problema de sistemas organizacionais. Em profissões em que os julgadores deveriam ser intercambiáveis (juízes, analistas de seguros, médicos, recrutadores), o ruído leva à injustiça, pois pessoas em situações similares recebem tratamentos e resultados drasticamente diferentes - o que é um grave problema aos direitos humanos.

Através dos experimentos expostos no livro *How Humans Judge Machines*, Hildalgo *et al* (2021) concluem que, em suma, julgamos pessoas por motivação e máquinas por resultado, o que explica a diferença nas métricas.

Isto é, os humanos, ao julgarem, focam na coerência da narrativa e na sensação interna de conclusão. O julgamento avaliativo é ruidoso porque as pessoas discordam em como traduzir a gravidade do ultraje em uma escala de punição. Tendemos a julgar pessoas pela motivação (o processo complexo, ainda que ruidoso e enviesado). Já as máquinas são julgadas pela precisão do resultado. O algoritmo pode ser "mais confiavelmente enviesado" do que o humano, pois codifica e replica preconceitos passados (como o viés em planos de saúde que preveem custos ao invés de doenças). No entanto, os algoritmos são, em certo sentido, mais transparentes do que os seres humanos, pois podem ser submetidos a escrutínio (auditoria) para verificar se discriminam, o que é muito mais difícil de fazer com o julgamento humano, muitas vezes inconscientemente enviesado.

Embora a transparência seja essencial, precisamos de um referencial comum: se questionamos a opacidade de um, qual a medida para o outro? A fundamentação escrita nem sempre reflete o racional da pessoa, conforme observa-se dos exemplos acima descritos.

3.5.5.1 A falácia da transparência

A linha argumentativa predominante apela à transparência, com a presunção de que ela é a panaceia para os problemas éticos (Mittelstadt, 2016, p. 6). Contudo, a crítica mais robusta, defendida por autores como Edwards & Veale (2017, p. 67), aponta que confiar nos direitos individuais de explicação corre o risco de criar uma "falácia da transparência" (similar à falácia do consentimento).

A ideia ingênua de que a transparência resolve o problema ignora o fato de que a disponibilização do código-fonte não se converte adequadamente em conhecimento e explicação. Nesse sentido, Ferguson (2017, p. 137/138): “*the issue*

[...] is not the transparency of the algorithm [...] but the transparency of how the program is explained to the public and, of course, what is done with the information”.

Ainda, a transparência pode ser ineficaz se a informação não for "significativa" (*meaningful*). A opacidade pode ser ofuscada pela tentativa deliberada de dissimulação, ou pela apresentação de informações excessivas que sobrecarregam o usuário, tornando o algoritmo, paradoxalmente, ainda mais opaco (Pasquale, 2015, p. 6/7).

Por fim, Floridi (2023, p. 99) interpreta que a transparência irrestrita pode permitir que indivíduos manipulem o sistema, a partir do conhecimento sobre a fonte dos dados ou a métrica. Assevera que, em contextos algorítmicos, a transparência total pode gerar problemas éticos, sobrecarregando usuários e tornando o algoritmo mais opaco (Kizilcec 2016, Ananny e Crawford 2018). Ademais, assevera que o foco excessivo na transparência pode prejudicar a inovação e desviar recursos que seriam melhor empregados em segurança, desempenho e precisão (Danks e London 2017, Oswald 2018, Ananny e Crawford 2018, Weller 2019).

Para Floridi (2023, p. 99), a opacidade pode ser mais útil em certos contextos, como no sigilo de votos ou em leilões de serviços públicos. O autor ressalta que mesmo em contextos algorítmicos, a transparência total pode gerar problemas éticos, como sobrecarregar o utilizador com informações, tornando o algoritmo mais opaco (Ananny e Crawford 2018, Kizilcec 2016). Além disso, assevera que o foco excessivo na transparência pode prejudicar a inovação e desviar recursos que poderiam ser usados para melhorar a segurança, o desempenho e a precisão (Danks e London 2017, Oswald 2018, Ananny e Crawford 2018, Weller 2019).

3.5.6 Direcionamentos legislativos atuais no Brasil e Europa

O avanço na regulação de sistemas de Inteligência Artificial (IA) tem se caracterizado pelo distanciamento de abordagens baseadas apenas em *soft law* em direção a modelos de regulação baseada em risco. Este movimento regulatório adota uma perspectiva antropocêntrica que disciplina a atividade tecnológica com foco nos riscos que podem advir dela, garantindo a curadoria constante dos direitos humanos.

A regulação europeia, como mencionado, há muito tem assumido a vanguarda regulatória, servindo de inspiração para as formulações de outros países, incluindo as brasileiras.

Assim, o Modelo Baseado em Risco do AI Act, lançado por meio do Regulamento (UE) 2024/1689, estabeleceu a primeira lei abrangente sobre IA. Nela, adotou-se uma taxonomia de riscos que fundamenta suas obrigações de governança. As aplicações de Risco Inaceitável (e.g., técnicas subliminares manipuladoras) representam uma ameaça clara aos valores fundamentais da UE e são proibidas. As de Alto Risco (utilizadas em setores críticos como infraestrutura, educação, emprego, administração da justiça) exigem regras específicas e mais rígidas de governança e gestão de riscos. Já as de Risco Limitado ou Mínimo impõem obrigações de transparência proporcionais (e.g., *chatbots*). Diferentemente do GDPR, no qual o direito à explicação é controverso e extraído de maneira implícita, o AI Act inova ao prever, de forma explícita e categórica, o direito à explicação.

O Artigo 86º (1) do AI Act consagra o direito individual de obter explicação, mas o faz de forma restrita: aplica-se a qualquer pessoa afetada por um sistema de IA de alto risco que produza efeitos jurídicos ou afete significativamente a pessoa (e.g., repercussões negativas em sua saúde, segurança ou direitos fundamentais) - - a pergunta aqui fica sendo “o que é um impacto significativo?”.

Nesse caso, o conteúdo da explicação deve ser "claro e pertinente sobre o papel do sistema de IA no processo de tomada de decisão e sobre os principais elementos da decisão tomada". Este dispositivo foge da ideia de funcionalidade geral do processo decisório automatizado e foca no substancial e *ex post*. Os provedores de sistemas de IA de alto risco têm a obrigação de submeter seus sistemas a procedimentos de avaliação da conformidade antes da colocação no mercado (AI Act, Art. 16(f) e 43). O AI Act também exige que esses sistemas sejam projetados para garantir transparência suficiente, permitindo que os implementadores interpretem os resultados (AI Act, Art. 13).

Doutro bordo, no Brasil, o PL 2338/2023 segue o modelo europeu de regulação baseada em risco, buscando disciplinar a IA com o Projeto de Lei nº 2.338, de 2023 (PL 2338/2023), que assumiu a dianteira no cenário legislativo.

Aprovado pelo Senado Federal em dezembro de 2024 (Toniazzo; Barbosa; Ruaro, 2025, p. 81), adota o paradigma antropocêntrico, ampliando o âmbito de

proteção. Nele, o Art. 3º (VI) contempla expressamente transparência, explicabilidade, inteligibilidade e auditabilidade como princípios aplicáveis ao desenvolvimento e uso de sistemas de IA.

Ainda, o referido PL estabelece direitos para as pessoas afetadas por sistemas de IA, que estão relacionados com o princípio da explicabilidade. Ademais, diferentemente do AI Act, o PL 2338/2023 é mais abrangente, permitindo que a solicitação de explicações (Art. 5º, II) seja usufruída de forma irrestrita por qualquer pessoa afetada por qualquer sistema de IA, independentemente do nível do risco que ofereça, e não limita a solicitação aos casos de efeitos jurídicos relevantes.

O dever de informação (*ex ante*) está previsto no Art. 7º, conferindo o direito a receber, previamente, informações claras e adequadas, incluindo as categorias de dados pessoais utilizados e as medidas de segurança (Brasil, 2023, Art. 7º).

Quanto ao conteúdo da explicação (*ex post*), o Art. 8º dispõe que a pessoa afetada pode solicitar explicação com informações sobre critérios e procedimentos utilizados, abrangendo fatores que afetam a decisão específica, o que inclui a racionalidade e a lógica do sistema, o grau e o nível de contribuição da IA, os critérios para a tomada de decisão e, quando apropriado, a sua ponderação, aplicados à situação da pessoa afetada (Brasil, 2023, Art. 8º). Este detalhamento busca dar maior concretude à explicabilidade, visando encurtar a assimetria informacional.

No que se refere à governança e Avaliação de Impacto Algorítmico (AIA), o PL 2338/2023, ao classificar o sistema como de alto risco, estabelece a obrigatoriedade de elaborar uma Avaliação de Impacto Algorítmico (AIA) (Brasil, 2023, Art. 22). A AIA é um Instrumento de Cognição mais amplo do que o Relatório de Impacto à Proteção de Dados Pessoais (RIPD), pois não se limita a dados pessoais, mas abarca a programação algorítmica e o aprendizado de máquina (Lemos et al., 2023).

A finalidade da AIA é delinear a responsabilização, identificar possíveis danos e dispor medidas de mitigação. No PL 2338/2023, o Art. 24 prevê que a AIA deve registrar os riscos conhecidos e previsíveis, os benefícios, a lógica do funcionamento do sistema e o processo e resultado dos testes para verificação de impactos a direitos (Brasil, 2023, Art. 24). A AIA é um processo iterativo contínuo, executado ao longo de todo o ciclo de vida dos sistemas de alto risco (Brasil, 2023, Art. 25). Há uma conexão com a LGPD, pois o AIA e o RIPD demonstram compatibilidade

funcional com a necessidade de inteligibilidade e explicabilidade do sistema, atuando como mecanismos de prestação de contas *ex ante* para a prevenção contra o potencial discriminatório dos algoritmos.

Demais disso, a Resolução 615/2025 também exige a publicação do sumário público da Avaliação de Impacto Algorítmico para soluções de alto risco.

Posto isto, o direcionamento regulatório atual demonstra que a explicabilidade é um princípio fundamental, sendo o AI Act e o PL 2338/2023 os diplomas normativos que o estabelecem como direito individual. Contudo, a principal diferença reside no escopo de aplicação: o AI Act (UE) é restritivo, limitando o direito à explicação aos casos de IA classificada como alto risco e que gere efeitos jurídicos relevantes, enquanto o PL 2338/2023 (Brasil) é amplo, permitindo a solicitação de explicação por qualquer pessoa afetada por qualquer sistema de IA, independentemente do nível de risco. A eficácia desses direcionamentos, no entanto, dependerá da capacidade de as estruturas de governança corporativa e as autoridades reguladoras (como a ANPD) operacionalizarem os instrumentos de transparência *ex ante* (AIA/RIPD) e garantirem a tradução de conceitos técnicos em explicações inteligíveis (Fjeld et al., 2020, p. 43).

É certo, portanto, que embora a regulação brasileira (LGPD) e o projeto de Marco Legal da IA (PL 2338/2023) demonstrem um horizonte de maior abrangência em relação ao direito de explicação (abrangendo qualquer pessoa afetada por qualquer sistema de IA, no PL 2338/2023, independentemente do nível de risco), a adequação do direito é comprometida pela fragilidade processual (ausência de revisão humana obrigatória na LGPD) e pela vagueza do conteúdo exigido ("critérios e procedimentos"), que pode não ser suficiente para desvendar a inescrutabilidade algorítmica e assegurar a justiça.

CONCLUSÕES

A presente dissertação dedicou-se a examinar o direito à explicação nas decisões automatizadas, empregando a técnica do direito comparado entre os ordenamentos jurídicos do Brasil e da União Europeia. A relevância deste estudo reside no impacto transformador e nos riscos inerentes à inteligência artificial, que, enquanto tecnologia de propósito geral, exige um arcabouço regulatório alinhado à transparência, responsabilidade e equidade. a questão central é como traduzir a opacidade dos algoritmos de *machine learning* (ML) em informação inteligível, garantindo o devido processo e a proteção dos direitos fundamentais.

A opacidade do algoritmo é um problema nuclear que se manifesta desde a invisibilidade de sua existência até a inescrutabilidade de seu funcionamento. Sistemas de IA, especialmente os modelos probabilísticos de ML e as redes neurais profundas (DNNs), são inerentemente complexos, o que dificulta a compreensão dos critérios de decisão, mesmo pelos desenvolvedores.

A pesquisa demonstrou que a mera abertura do código-fonte é inócua e insuficiente, configurando uma "falácia da transparência". O direito à explicação, portanto, deve ser compreendido como um esforço de transparência qualificada e inteligibilidade.

O objetivo é traduzir a complexidade técnica em formatos compreensíveis ao homem-médio. A explicabilidade é a interface entre o sistema inteligente e o ser humano, atuando para fornecer o "porquê" oculto nas decisões. As explicações devem ser úteis e funcionais, permitindo que o titular dos dados possa contestar a decisão ou tomar uma ação para mudar sua situação.

É certo, entretanto, que o direito à explicação deve ser contextualizado pela ciência cognitiva, o qual aponta as explicações humanas frequentemente como racionalizações *ex post facto*, e que o julgamento humano é inerentemente ruidoso, carecendo de consistência. A complexidade e a variabilidade indesejada (*noise*) do julgamento humano tornam o algoritmo, se bem desenhado, mais confiável em termos de consistência.

A investigação evidenciou que o direito de explicação não se concretiza apenas pela postulação do direito individual (*ex post*), mas exige uma estrutura robusta de governança de IA e de dados (*ex ante*), com o intuito de viabilizar a prestação de contas (*accountability*).

Nesse sentido, documentos como o Relatório de Impacto à Proteção de Dados (RIPD) na LGPD e a Avaliação de Impacto Algorítmico (AIA), exigida no PL 2.338/2023 para sistemas de alto risco, são essenciais. Eles agem como instrumentos de prevenção de riscos e rastreabilidade, fornecendo subsídios para as explicações sobre critérios e procedimentos utilizados.

A governança deve ser pautada no vetor de prevenção, assegurando o monitoramento e auditoria contínuos em todas as fases do ciclo de vida da IA. A implementação de práticas razoáveis e eficientes de governança pode servir como fator atenuante para a responsabilidade jurídica das corporações.

A constatação central é que a complexidade da governança de IA é equivalente à da própria tecnologia. Portanto, não existe um modelo único ou uma estrutura padronizada de governança que seja plausível para lidar com o espectro de arquiteturas e impactos da IA. Defende-se a criação de um ambiente ou subconjunto especializado de governança corporativa, que atue como função orquestradora para integrar as quatro forças complementares (leis formais, normas sociais, regras de mercado e tecnologia) e traduzir os princípios éticos gerais (como transparência e justiça) em políticas e protocolos objetivos.

Em suma, a dissertação conclui que, considerando os marcos normativos propostos (LGPD e GDPR), não existe um direito à explicação positivado, tratando-se de um direito extraído da interpretação de outros direitos e princípios. É inquestionável, entretanto, que o direito à explicação é um direito fundamental de controle democrático, especialmente em um mundo mediado por algoritmos. Sua efetividade exige a transposição de princípios abstratos em práticas operacionais, o que só é possível através da implementação de uma governança de IA flexível, multidisciplinar e rigorosa, capaz de conciliar o avanço tecnológico com a inegociável proteção dos direitos humanos.

REFERÊNCIAS BIBLIOGRÁFICAS

ALMADA, Marco. Human intervention in automated decision-making: toward the construction of contestable systems. In: ICAIL '19: SEVENTEENTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW. Proceedings [...]. Nova York: ACM Press, 2019. Disponível em: https://dl.acm.org/doi/10.1145/3322640.3326699. Acesso em 25/09/2025.

ANANNY, Mike; CRAWFORD, Kate. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society. 2018, v. 20(3), pp. 973-989. https://journals.sagepub.com/doi/10.1177/1461444816676645. 01.05.2021.

ANDRADE, Otávio Morato de. Governamentalidade algorítmica: democracia em risco? 1. ed. São Paulo: Dialética, 2022; RUSSEL, Stuart; NORVIG, Peter. Artificial intelligence: a modern approach. New Jersey: Prentice-Hall, 1995

ARAÚJO, Alexandre. 2024, Principle of explicability: Regulatory challenges on artificial intelligence

AUGUSTO, Luís Gustavo Henrique; RIZZARDI, Maira Martinelli. Accountability segundo os Ministros dos Tribunais Superiores do Judiciário Brasileiro. Apresentação de trabalho no XXIII Encontro Nacional do Conpedi - Direito e administração Pública. Florianópolis, 2014. Disponível em: https://www.researchgate.net/publication/280626416_Accountability_segundo_os_Ministros_dos_Tribunais_Superiores_do_Judiciario_Brasileiro_Accountability_according_to_the_Ministers_of_the_Superior_Courts_of_the_Brazilian_Judiciary> Acessado em 19.06.2021.

ÁVILA, Humberto Bergmann. **A distinção entre princípios e regras e a redefinição do dever de proporcionalidade.** Revista de Direito Administrativo, Rio de Janeiro, 215:151-179, jan./mar. 1999. Disponível em: <http://bibliotecadigital.fgv.br/ojs/index.php/rda/article/download/47313/45714>. Acesso em: 10 jun. 2020.

BARBOSA, Xênia de Castro. BREVE INTRODUÇÃO À HISTÓRIA DA INTELIGÊNCIA ARTIFICIAL. **Jamaxi**, v. 4, n. 1, p. 1-13, 2020.

BIONI, Bruno Ricardo. Proteção de dados pessoais: a função e os limites do consentimento. Rio de Janeiro: Forense, 2019. E-book.

BECK, César; BOFF, Murilo Manzoni; PIAIA, Thami Covatti. Lei Geral de Proteção de Dados e a revisão de decisões automatizadas: os mecanismos de regulação baseados em uma inteligência artificial ética. Revista Eletrônica Direito e Política, Programa de Pós-Graduação Stricto Sensu em Ciência Jurídica da UNIVALI, vo 17, no 2, 2o quadrimestre de 2022. Disponível em: https://periodicos.univali.br/index.php/rdp - ISSN 1980-7791. DOI: https://doi.org/10.14210/10.14210/rdp.v17n2.p509-546

BRASIL. Câmara dos Deputados. *Projeto de Lei 21/2020*. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928 . Acesso em: 1 abr. 2025.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Dispõe sobre o Marco Civil da Internet. Diário Oficial da União, Brasília, 24 abr. 2014. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em 01 mai. 2025

BRASIL. Congresso Nacional. Senado Federal. Projeto de Lei n° 21, de 2020: Relatório final da Comissão de Juristas responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial. Diário do Senado Federal, ano 77, n. 204, suplemento n. B, p. 3-915, 9 dez. 2022. Disponível em: https://legis.senado.leg.br/diarios/BuscaDiario?tipDiario=1&datDiario=09/12/2022&paginaDireta=3&indSuplemento=Sim&codSuplemento=B&desVolumeSuplemento=&desTomoSuplemento=&_gl=1*hulqbi*_ga*MjgwNzg1MDUzLjE2MjQ0ODA3Nzk.*_ga_CW3ZH25XMK*MTY5NjUzMDI1MS4yNS4xLjE2OTY1MzA3MDcuMC4wLjA . Acesso em: 30 mar. 2024.

BRADFORD, Anu. The Brussels Effect: How the European Union Rules the World. Oxford: Oxford University Press, 2020.

BURREL, J. How the machine ‘thinks:’ Understanding opacity in machine learning algorithms. **Big Data & Society**, v. 3, n. 1, p. 1–12, 2016.

DIAKOPOULOS, Nicholas. Accountability in algorithmic decision making. Communications of the ACM, vol. 59(2), 2016, pp. 56–62. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/2844110> . Acesso em 25.09.2025.

DONEDA, Danilo; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes de. Considerações sobre Inteligência Artificial, ética e autonomia pessoal. Pensar – Revista de Ciências Jurídicas, v. 23, p. 1-17, 2018.

EUROPEAN PARLIAMENT. European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)). Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html . Acesso em: 1 abr. 2025.

EUROPE, European Union Agency For Fundamental Rights And Council Of. **Handbook on European data protection law**. 2018. Disponível em: https://fra.europa.eu/en/publication/2018/handbook-european-data-protection-law-2018-edition . Acesso em: 11 jun. 2024.

FRAJHOF, Isabella Zalcberg. *Direito à Explicação e Proteção de Dados Pessoais nas Decisões por Algoritmos de Inteligência Artificial.* 2022. 224 f. Tese (Doutorado) -

Curso de Direito, Departamento de Direito, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2022. Disponível em: https://doi.org/10.17771/PUCRio.acad.60965. Acesso em: 25 nov. 2024.

FRAZÃO, Ana. O direito à explicação e à oposição diante de decisões totalmente automatizadas. Revista JOTA, 05 de dezembro de 2018. Disponível em https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/o-direito-a-explicacao-e-a-oposicao-diante-de-decisoes-totalmente-automatizadas Acessado em 11 mai 2025.

FRAZÃO, Ana; NUKHOLLAND, Caitlin. *Inteligência artificial e direito*: ética, regulação e responsabilidade. 2. ed. São Paulo: Revista dos Tribunais, 2019. Edição do Kindle.

FRAZÃO, Ana; GOETTENAUER, Carlos. O jogo da imitação jurídica: o direito à revisão de decisões algorítmicas como um mecanismo para a necessária conciliação entre linguagem natural e infraestrutura matemática. In: TEPEDINO, Gustavo; SILVA, Rodrigo da Guia (Orgs.). O Direito Civil na era da Inteligência Artificial. São Paulo: Thomson Reuters Brasil, 2020, pp. 45-64.

FREITAS, Juarez; FREITAS, Thomas Bellini. Direito e inteligência artificial: em defesa do humano. 1. Reimpressão. Belo Horizonte: Fórum, 2020.

HARARI, Yuval. Homo Deus: uma breve história do amanhã. Tradução de Paulo Geiger. 1. ed. São Paulo: Companhia das Letras, 2016, p. 343.

HIRSCHL, Ran. *Comparative matters: the renaissance of comparative constitutional law*. Oxford: Oxford University Press, 2014.

HOFFMANN-RIEM, Wolfgang. Teoria geral do direito digital: transformação digital: desafios para o direito/Wolfgang Hoffmann-Riem – 2ed. - Rio de Janeiro:Forense, 2022.

IRIS - INSTITUTO DE REFERÊNCIA EM INTERNET E SOCIEDADE. *Inteligência Artificial no Brasil: a Estratégia Brasileira de Inteligência Artificial (EBIA) e o Projeto de Lei nº 21/2020*. Disponível em: https://irisbh.com.br/inteligencia-artificial-no-brasil-a-estrategia-brasileira-de-inteligencia-artificial-ebia-e-o-projeto-de-lei-no-21-2020/ . Acesso em: 1 abr. 2025.

KAHNEMAN, Daniel. SIBONY, Olivier. SUNSTEIN, Cass R. Ruído: Uma falha no julgamento humano. Editora Objetiva. 2021.

KAMINSKI, Margot E.. Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability. *Southern California Law Review*, Vol. 92, No. 6, 2019 1529 U of Colorado Law Legal Studies Research Paper No. 19-9.

KIM, Tae Wan; ROUTLEDGE, Bryan R. Informational Privacy, A Right to Explanation, And Interpretable AI. Em: PROCEEDINGS - 2018 2ND IEEE SYMPOSIUM ON

PRIVACY-AWARE COMPUTING, PAC 2018, Anais [...]: Institute of Electrical and Electronics Engineers Inc., p. 64–74, 2018.

KNUTH, D. E. (1968). The Art of Computer Programming, Vol. 1: Fundamental Algorithms. Addison-Wesley.

KUNER, Christopher (ed.). *The EU General Data Protection Regulation (GDPR)*: a commentary. Oxford: Oxford University Press, 2020.

LIMA, Taisa Maria Macena de; SÁ, Maria de Fátima Freire de. Inteligência artificial e Lei Geral de Proteção de Dados Pessoais: o direito à explicação nas decisões automatizadas. **Revista Brasileira de Direito Civil**, Belo Horizonte, v. 26, n. 04, p. 227-246, 00 out. 2020. Instituto Brasileiro de Direito Civil - IBDCivil. http://dx.doi.org/10.33242/rbdc.2020.04.011. Disponível em: https://rbdcivil.ibdcivil.org.br/rbdc/article/download/584/425. Acesso em: 11 maio 2025.

LIPTON, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3), 31-57.

MATEAS, M.; MONTFORT, N. A box, darkly: Obfuscation, weird languages, and code aesthetics. In: *Proceedings of the 6th Annual Digital Arts and Culture Conference*, Copenhagen, Denmark, 2005. p. 1-3

MCCARTHY, John, et al. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955.

MCCARTHY, J. What is AI? / Basic Questions. Disponível em: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html#:~:text=What%20is%20artificial%20intelligence%3F,methods%20that%20are%20biologically%20observable. 1 abr. 2025.

MEDON, Felipe. *Inteligência Artificial e Responsabilidade Civil*: Autonomia, Riscos e Solidariedade. 2. ed. São Paulo: JusPodivm. 2022.

MICHAELIS. Opacidade. In: *Michaelis Dicionário Brasileiro da Língua Portuguesa*. [S. l.]: Melhoramentos. Disponível em: https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=opacidade. Acesso em: 1 maio 2025.

MITTELSTADT, Brent Daniel; ALLO, Patrick; TADDEO, Mariarosaria; WACHTER, Sandra; FLORIDI, Luciano. The ethics of algorithms: mapping the debate. *Big Data & Society*, [S. l.], v. 3, n. 2, p. 1–21, jul./dez. 2016. Disponível em: https://doi.org/10.1177/2053951716679679 . Acesso em: 1 maio 2025.

MOLNAR, Christoph. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. 2021.

MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados do Brasil?. Instituto Igarapé. Artigo Estratégico 39, dezembro de 2018, p.

10. Disponível em: https://igarape.org.br/wp-content/uploads/2018/12/Existe-um-direito-a-explicacao-na-Lei-Geral-de-Protecao-de-Dados-no-Brasil.pdf. Acessado em 10 mai 2025.

MORAIS, Fausto Santo de. O uso da inteligência artificial na repercussão geral: desafios teóricos e éticos. Revista de Direito Público, Brasília, v. 18, n. 100, p. 306-326, 2021. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6001/pdf . Acesso em: 18 jun. 2023

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z. Inteligência artificial e a lei de proteção de dados pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de machine learning. In: FRAZÃO, Ana. MOULHOLLAND, Caitlin (Coord.). Inteligência artificial e direito: ética, regulação e responsabilidade. São Paulo: Thomson Reuters (Revista dos Tribunais), 2019. p. 265-290.

NEGNEVITSKY, Michael. Artificial intelligence: A guide to intelligent systems. Harlow: Addison-Wesley, 2004.

NEWMAN, Maxwell Herman Alexander. Alan Mathison Turing, 1912-1954. Biographical memoirs of fellows of the Royal Society. Publicado em 1º nov. 1955. Portal The Royal Society Publishing. Disponível em: http://rsbm.royalsocietypublishing.org/content/roybiogmem/1/253.full.pdf . Acesso em: 29 mar. 2024

NILSSON, N. The quest for artificial intelligence: a history of ideas and achievements. Cambridge: Cambridge University Press, 2009.

NOVAIS, Paulo; FREITAS, Pedro Miguel. *Inteligência artificial e regulação de algoritmos. Brasília e Bruxelas*: Governo Federal Brasileiro e Comissão Europeia, 30 de maio de 2018, p. 9 e 16. (Relatório Diálogos União Europeia-Brasil). Disponível em: https://etica.uazuay.edu.ec/sites/etica.uazuay.edu.ec/files/public/49f7d3_Intelig%C3%AAncia%20Artificial%20e%20Regula%C3%A7%C3%A3o%20de%20Algoritmos.pdf . Acesso em: 1 abr. 2025.

NUNES, Dierle José Coelho; ANDRADE, Otávio Morato de. O uso da inteligência artificial explicável enquanto ferramenta para compreender decisões automatizadas: possível caminho para aumentar a legitimidade e confiabilidade dos modelos algorítmicos? *Revista Eletrônica do Curso de Direito da UFSM*, v. 18, n. 1, e69329, 2023. DOI: 10.5902/1981369469329. Disponível em: https://www.ufsm.br/revistadireito. Acesso em: 1 maio 2025

O'NEIL, Cathy. *Algoritmos de destruição em massa: como o Big Data aumenta a desigualdade e ameaça a democracia*. Tradução de Rafael Abraham. 1. ed. São Paulo: Editora Rua do Sabão, 2021.

OTTERLO, Martijn van. A machine learning view on profiling. In: HILDEBRANDT, Mireille; GUTWIRTH, Serge (ed.). *Profiling the European citizen: cross-disciplinary*

perspectives. Abingdon: Routledge, 2008. p. 41–76. Disponível em: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203427644-4/machine-learning-view-profiling-martijn-van-otterlo. Acesso em: 1 maio 2025.

PASQUALE, Frank. Black box society: The secret algorithms that control money and information. [S. l.: s. n.], 2015. 311 p. ISBN 9780674970847.

PINHEIRO, Patrícia Peck. Proteção de dados pessoais: comentários à Lei n.o 13.709/2018 (LGPD) — 2. Ed. São Paulo: Saraiva Educação, 2020. p. 20-21. E-book.

PIRES, Thatiane Cristina Fontão; SILVA, Rafael Peteffi da. A responsabilidade civil pelos atos autônomos da inteligência artificial: notas iniciais sobre a resolução do Parlamento Europeu. Revista Brasileira de Políticas Públicas, v. 7, n. 3, p. 239-254, dez. 2017. Disponível em: https://www.publicacoesacademicas.uniceub.br/RBPP/article/view/4951 . Acesso em: 23 jan. 2020

RUDIN, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215.

RUSSEL, S. J.; NORVIG, P. Artificial Intelligence: a modern approach. 3ª ed. New Jersey: Prentice Hall, 2009

SANDVIG, Christian et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms. In: Annual Meeting of the International Communication Association, Seattle, maio 2014. Disponível em: https://www.semanticscholar.org/paper/Auditing-Algorithms-%3A-Research-Methods-for-on-Sandvig-Hamilton/b7227cbd34766655dea10d0437ab10df3a127396?p2df . Acesso em: 01 mai. 2025.

SARDINHA, Adriane Nascimento Celestino. Direito à informação e à revisão de decisões informatizadas: uma análise da LGPD. *Res Severa Verum Gaudium,* v. 6, n. 2, Porto Alegre, p. 257-286, mai. 2022.

SELBST, Andrew D.; BAROCAS, Solon. Intuitive Appeal of Explainable Machines. 87 Fordham L. Rev. 1085, 2018.

_____, Andrew D.; POWLES, Julia. Meaningful information and the right to explanation. *International Data Privacy Law*, Oxford, v. 7, n. 4, p. 233–242, 2017. Disponível em: https://academic.oup.com/idpl/article/7/4/233/4762325. Acesso em: 1 maio 2025.

SHANE, Janelle. Você parece uma coisa e eu te amo: como a inteligência artificial funciona e por que está fazendo do mundo um lugar mais estranho/ Janelle Shane; trad. Daniel Salgado – Rio de Janeiro: Alta Books, 2022.

SILVA, Luís Virgílio Afonso da. **Princípios e regras: mitos e equívocos acerca de uma distinção**. Revista Latino-Americana de Estudos Constitucionais, Belo Horizonte, Ano: n.º 1, jan./jun. 2003. Disponível em: <https://constituicao.direito.usp.br/wp-content/uploads/2003-RLAEC01-Principios_e_regras.pdf>. Acesso em: 10 jun. 2020

STANFORD ENGINEERIN. A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE. Disponível em: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html . Acesso em: 1 abr. 2025.

SURDEN, H. (2014). Machine learning and law. Washington Law Review, v. 89, nº 1, mar. 2014, p. 87-115. Disponível em: https://digitalcommons.law.uw.edu/wlr/vol89/iss1/5/ . Acesso em: 18 jun. 2023.

TEIXEIRA, Tarcísio; CHELIGA, Vinicius. Inteligência Artificial: Aspectos Jurídicos. 3. ed. Salvador: JusPodivm, 2021

TONIAZZO, Daniela Wendt; BARBOSA, Tales Schmidke; RUARO, Regina Linden. O DIREITO À EXPLICAÇÃO NAS DECISÕES AUTOMATIZADAS: uma abordagem comparativa entre o ordenamento brasileiro e europeu. *Revista Internacional Consinter de Direito*, Porto, p. 55-69, 21 dez. 2021. Semestral. Disponível em: http://dx.doi.org/10.19135/revista.consinter.00013.00. Acesso em: 11 maio 2025.

VILLANI, Cédric. Donner uns sens à li'intelligence artificielle: pour une stratégie nationale et européenne. Paris, 2018. Disponível em: https://medias.viepublique.fr/data_storage_s3/rapport/pdf/184000159.pdf . Acesso em: 18 jun. 2023.

WACHTER, Sandra; MITTELSTADT, Brent; A Right to Reasonable Interferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *LawArXiv,* 12 out. 2018. Disponível em <https://osf.io/preprints/lawarxiv/mu2kf>. Acesso em 13 out. 2024.

WACHTER, Sandra; MITTELSTADT, Brent; RUSSELL, Chris. Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law. West Virginia Law Review, Morgantown, v.123, n. 3, p.1-51, 2021.