

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE DIREITO
CURSO DE GRADUAÇÃO EM DIREITO

CAMILA CRISTINA DA SILVA

**AS DIMENSÕES DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS
TOMADAS POR MÁQUINAS: análise das leis de proteção de dados e propostas de
regulação de inteligência artificial no Brasil e União Europeia**

2024

BRASÍLIA

CAMILA CRISTINA DA SILVA

AS DIMENSÕES DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS
TOMADAS POR MÁQUINAS: análise das leis de proteção de dados e propostas de regulação
de inteligência artificial no Brasil e União Europeia

Monografia apresentada, no âmbito do Curso de Graduação em Direito, à Faculdade de Direito (FD) da Universidade de Brasília (UnB) como requisito para a obtenção do título de Bacharel em Direito.

Orientadora: Professora Doutora Ana Frazão Vieira de Mello.

2024

BRASÍLIA

*“Fui feliz porque não pedi coisa nenhuma,
Nem procurei achar nada,
Nem achei que houvesse mais explicação
Que a palavra explicação não ter sentido nenhum.”¹*

¹ CAEIRO, Alberto. **Se eu morrer de novo**. Disponível em: <http://arquivopessoa.net/textos/986>. Acesso em: 16 de mai. de 2024.

Às mulheres da minha vida: minha mãe Nádia, minha irmã Amanda e minha avó Eliezita (*in memoriam*) — que partiu ao longo dessa graduação e levou consigo um pedaço de mim — por acreditarem, sempre, que eu era capaz. E, hoje, sei que tudo posso.

AGRADECIMENTOS

Meu primeiro agradecimento não poderia ser direcionado a ninguém que não fosse a minha mãe, Nádia. Esse espaço e essas palavras são insuficientes para descrever a imensidão do apoio, da gratidão e do amor que recebi até o dia de hoje e receberei até meu último dia, porque tudo em mim é composto por você, mãe. Não posso medir a exaustão dos seus ossos e músculos todos os dias sendo pai, mãe, dona de casa e servidora pública concomitantemente. Não posso jamais mensurar o que deixou de possuir, fazer, sentir e viver para que eu andasse o que nunca pôde andar. Não posso contar as vezes que o meu estudo foi prioridade. Nada contabiliza todas as suas noites mal dormidas e todas as manhãs cedo amanhecidas para que pudesse não somente ter oportunidades inestimáveis, como também ter o privilégio de ter escolhas. Obrigada por me ensinar e mostrar, diariamente, que o conhecimento é o caminho para uma vida plena. Esse trabalho não é só meu. É seu também. Um sonho que um dia você teve e a vida não lhe permitiu, mas você me deu todas as condições para realizá-lo: formar em direito na UnB. Sou filha de uma mulher negra nordestina que venceu, apesar de tudo. A sua força eu nunca terei. Mas eu aprendi a tentar ter todos os dias, principalmente para chegar até aqui. Te amo, mãe.

A Deus, por ter me encontrado esse ano. Encontrei Nele a figura de Pai.

Agradeço imensamente, com todo o meu ser, à minha irmã, Amanda. Por todas as vezes que me lembrou que o mundo existe fora das telas de um computador e me lembrou que viver é para além disso. Por sempre me garantir momentos de paz e descontração nas piores fases de toda essa trajetória. Eu não tenho somente uma irmã, eu tenho uma amiga, em uma relação que nos une por sangue, compreensão e amor na mais pura forma. Obrigada por me ver e ouvir quando ninguém mais poderia, preocupando-se com o que na vida, realmente importa. Te amo mais do que você pode imaginar e mensurar. Agradeço também ao meu irmão Matheus pela compreensão silenciosa de podermos ser os nossos lados mais “nerds”.

Não poderia deixar de agradecer a todos os laços que construí nesses cinco anos nos concretos brutos da FD. Lore e Maria, obrigada por serem meu trio, meu pilar e minhas melhores amigas nesses anos, eu as amo e as levo comigo. Ana Lu, obrigada por ser um presente de Deus na minha vida no fim da corrida, a sua fé, força e luz me inspiram. Agradeço também às outras mulheres inspiradoras que marcaram a graduação e não ao caso, compõem um grupo que

denominamos de “as mais lindas da turma”: Ana Bia, Clara, Carol, Lu Gatto, Louisy e Sofia. Obrigada por fazerem os anos mais doces. Que a nossa amizade se estenda para além dos corredores de tijolinhos. Agradeço também aos meus amigos que a FD me deu que tornaram a jornada mais leve, inspiradora e divertida, sendo verdadeiros queridos. Gabriel, obrigada por ser o melhor amigo que a faculdade e Deus poderiam me presentear: o que Ele une, nada mais separa. Dezan, agradeço por ouvir todos os surtos e áudios infundáveis, sendo um porto seguro quando sei que ninguém mais poderia escutar. Barradas, Quezado e Trajano, obrigada por cada conversa e risada. Rafa Krauspenhar, Fabi Berçott e Rafa Muller, obrigada por serem inspirações e parceiros nesses momentos. Sou mais feliz e completa por conhecer todos vocês.

Agradeço também àqueles amigos que vieram antes da FD e fizeram parte de quem fui e estou sendo, perdurando, embora tempo e distância tentassem nos afastar (e, felizmente, falharem). Ana Luisa, Aninha e Gigi obrigada por serem as melhores amigas e irmãs que eu poderia jamais ter, por me apoiarem, serem meu lar até hoje depois de tantos anos. Nobre, agradeço por ser meu mais antigo e melhor amigo, além de ser uma das minhas pessoas favoritas, admiro você como ninguém. Israel, Sophia, Toin e Tonhão, obrigada por serem verdadeiros amigos para todos os momentos nesta caminhada e me concederem lembranças tão memoráveis. Para aqueles que vieram no meio para o final do caminho, há um espaço especial também, pois não é preciso tempo para o que era predestinado: obrigada Ana e Isaque.

Não posso deixar de fazer o meu agradecimento ao único lugar que trabalhei ao longo do curso e aos meus chefes que me deram a chance de me envolver, desenvolver e trabalhar na área que eu me interesse e amo hoje. Obrigada Silva Prado Roscoe Bessa, Bruno e Alan pela paciência, lições, capacitações inestimáveis e o despertar de interesse pelo compliance.

Por fim, agradeço às minhas inspirações na academia. À Professora e minha orientadora Ana Frazão, que, embora possa não ser capaz de mensurar o seu impacto e influência sobre todas as garotas recém ingressantes da FD-UnB, inspira-nos a ser uma mulher, jurista, pesquisadora e professora como ela: destemida e sábia. Ao Laboratório de Políticas Públicas e Internet (LAPIN) e à Cynthia, por me acolher como pesquisadora e confiar em mim para somar ao movimento da sociedade civil em temas tão pertinentes à tecnologia. Ao Laboratório de Governança e Regulação da Inteligência Artificial (LIA-IDP), à professora Tainá Junquillo e à Giovanna Milanese por me darem a oportunidade de estar em um ambiente de pesquisa privilegiado, e serem inspirações feminina

RESUMO

Com a crescente e polêmica discussão regulatória de sistemas de inteligência artificial que traz à tona relevantes *tradeoffs* entre inovação tecnológica e garantia de direitos fundamentais, a presente monografia tem como objetivo investigar as múltiplas dimensões do direito à explicação em decisões automatizadas, com foco nos contextos jurídicos brasileiro e europeu. A partir do estudo das características técnicas da explicabilidade de sistemas de inteligência artificial (IA), discute-se a viabilidade e os limites de fornecer explicações claras e compreensíveis sobre decisões tomadas por algoritmos. Além disso, explora-se a importância existencial desse direito em uma contemporaneidade moldada pelo fenômeno do capitalismo de vigilância, relacionando-o com institutos principiológicos do livre desenvolvimento da personalidade e a autodeterminação informativa. Fundamentalmente, esta monografia examina o panorama legislativo, em perspectiva comparada das abordagens da Lei Geral de Proteção de Dados (LGPD) do Brasil e do Regulamento Geral sobre a Proteção de Dados (RGPD) da União Europeia, bem como as propostas de regulamentação específicas de IA, incluindo o recente AI Act europeu e os projetos de lei em tramitação no Brasil. Concluiu-se que a explicabilidade é crucial para garantir a transparência, *accountability* e a correção de vieses discriminatórios em sistemas de IA em um modelo regulatório que integra proteção de dados em aplicações de inteligência artificial e que, embora o direito à explicação não seja previsto expressamente em leis de proteção de dados de ambos os ordenamentos jurídicos, as regulamentações de IA - PL 21/2020, PL 2338/2023, PL 210/2024 e recente relatório final do PL 2338/2023 e AI Act - adicionam novas camadas à discussão.

Palavras-Chave: Proteção de Dados; Decisão Automatizada, Direito à Explicação; Explicabilidade; IA explicável.

ABSTRACT

With the growing and controversial regulatory discussion on artificial intelligence systems that brings to light relevant trade-offs between technological innovation and the guarantee of fundamental rights, this dissertation aims to investigate the multiple dimensions of the right to explanation in automated decisions, focusing on the Brazilian and European legal contexts. By studying the technical characteristics of the explainability of artificial intelligence (AI) systems, the feasibility and limits of providing clear and understandable explanations for algorithmic decisions are discussed. Additionally, the existential importance of this right in a contemporary society shaped by the phenomenon of surveillance capitalism is explored, relating it to the principled institutes of the free development of personality and informational self-determination. Fundamentally, this dissertation examines the legislative landscape, in a comparative perspective of the approaches of Brazil's General Data Protection Law (LGPD) and the European Union's General Data Protection Regulation (GDPR), as well as specific AI regulation proposals, including the recent European AI Act and the bills under consideration in Brazil. It is concluded that explainability is crucial to ensuring transparency, accountability, and the correction of discriminatory biases in AI systems within a regulatory model that integrates data protection in AI applications and that, although the right to explanation is not expressly provided for in data protection laws of both legal systems, AI regulations - PL 21/2020, PL 2338/2023, PL 210/2024 and the recent final report of PL 2338/2023 and AI Act - add new layers to the discussion.

Keywords: Data Protection, Automated Decision, Right to Explanation, Explainability, Explainable AI

SUMÁRIO

SUMÁRIO.....	9
CONSIDERAÇÕES INICIAIS.....	10
1. A DIMENSÃO TÉCNICA DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS.....	16
1.1. A NATUREZA E AS NUANCES DOS ALGORITMOS DE APRENDIZADO DE MÁQUINA.....	16
1.2. DECISÕES AUTOMATIZADAS: O CONCEITO DE DECIDIR NO CONTEXTO DA AUTOMAÇÃO.....	27
1.3. EXPLICABILIDADE DECISÓRIA: UTOPIA TÉCNICA OU PRESSUPOSTO DE TRANSPARÊNCIA E ACCOUNTABILITY?.....	36
1.4. O SEGREDO DE EMPRESA COMO ELEMENTO LIMITANTE DA EXPLICABILIDADE DE SISTEMAS ALGORÍTMICOS AUTOMATIZADOS.....	50
2. A DIMENSÃO EXISTENCIAL DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS.....	54
2.1. O DIREITO À EXPLICAÇÃO COMO DECORRÊNCIA DO DEVIDO PROCESSO LEGAL E DEVIDO PROCESSO INFORMACIONAL.....	59
2.2. O INDIVÍDUO NO CAPITALISMO DE VIGILÂNCIA E NA SOCIEDADE DO ALGORITMO.....	61
2.3. O DIREITO À EXPLICAÇÃO COMO EFETIVAÇÃO DO LIVRE DESENVOLVIMENTO DA PERSONALIDADE E DA AUTODETERMINAÇÃO INFORMATIVA.....	68
2.4. O DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS COMO MECANISMO DE COMBATE À DISCRIMINAÇÃO ALGORÍTMICA.....	74
3. A DIMENSÃO LEGISLATIVA DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS.....	80
3.1. O DIREITO À EXPLICAÇÃO NO REGULAMENTO GERAL SOBRE A PROTEÇÃO DE DADOS DA UNIÃO EUROPEIA.....	81
3.2. O DIREITO À EXPLICAÇÃO PARA ALÉM DOS ARTS. 13, 14, 15 E 22 DO REGULAMENTO SOBRE A PROTEÇÃO DE DADOS.....	94
3.3. O DIREITO À EXPLICAÇÃO NO CÓDIGO DE DEFESA DO CONSUMIDOR E NA LEI GERAL DE PROTEÇÃO DE DADOS BRASILEIRA.....	104
3.4. O DIREITO À EXPLICAÇÃO NO EU AI ACT E NOS PLs DE REGULAMENTAÇÃO DE IA NO BRASIL.....	112
3.4.1. O Direito à Explicação em Decisões Automatizadas no EU AI ACT.....	113
3.4.2. O Direito à Explicação em Decisões Automatizadas nos PL nº 21/2020, 210/2024 e 2338/2023.....	122
CONSIDERAÇÕES FINAIS.....	133
REFERÊNCIAS.....	136

CONSIDERAÇÕES INICIAIS

Cada vez mais, máquinas decidem sobre nós: sobre quem somos, quem podemos vir a ser e quais oportunidades teremos acesso. Essas decisões impactam significativamente em nossas vidas, com efeitos jurídicos evidentes e ampla diversidade de aplicações: concessão de benefícios sociais e crédito, processos seletivos, sistemas de reconhecimento facial² e até mesmo ferramentas no sistema de justiça para avaliar a nossa capacidade de cometer crimes e eventual reincidência. O caso estadunidense *State vs. Eric Loomis* é uma retrato verossímil de uma máquina, denominada COMPAS (“*Correctional Offender Management Profiling for Alternative Sanctions*”)³ criada para avaliação da possível reincidência de suspeitos nos processos penais no estado de Wisconsin. Loomis fora condenado a seis anos de prisão por ter sido categorizado como um indivíduo de alta periculosidade por um algoritmo matemático. A fórmula se baseia nas respostas de um formulário respondidas pelos condenados, mas os critérios são desconhecidos e explicações não são dadas sobre como se chega a determinada pontuação de um preso, pois a empresa responsável pelo sistema alega não poder compartilhá-los devido à proteção do instituto jurídico do segredo de negócios.⁴

O caso esbarra diretamente em princípios e garantias constitucionalmente previstos em ordenamentos jurídicos democráticos diante do fato de que a ausência de explicabilidade em modelos algorítmicos gerar um conflito direto com os princípios constitucionais da recorribilidade, ampla defesa, contraditório e devido processo legal, como o próprio Eric fundamentou em seu recurso.⁵ Ademais, discussões raciais são trazidas à tona por pesquisadores e jornalistas que acompanham o COMPAS diante da constatação de que minorias étnicas recebem pontuações maiores comparativamente a pessoas brancas. Assim, suscita-se: como lidar, juridicamente, com máquinas decidindo e influenciando decisões com efeitos jurídicos diretos na vida das pessoas?

² BENTES, Anna et al. **Para além da proteção de dados: uma coletânea**. São Paulo: Data Privacy Brasil Ensino, 2023, p. 148.

³ “Perfil de Gerenciamento Corretivo de Infratores para Sanções Alternativas”, em tradução livre.

⁴ MAYBIN, Simon. Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA. **BBC News Brasil**, 31 de out. 2016. Disponível em: <https://www.bbc.com/portuguese/brasil-37677421>

⁵ JUNQUILHO, Tainá. **Inteligência Artificial no Direito: limites éticos**. Salvador: Juspodivm, 2022, p. 82.

Para além de negação de acesso a garantias fundamentais, as decisões automatizadas podem ser discriminatórias por atribuírem pesos desproporcionais a determinados critérios e características individuais durante o processo decisório. Pode-se visualizar exemplificadamente no seguinte caso: ao se postar imagens com dimensões demasiadamente verticais ou horizontais na linha do tempo do Twitter (atualmente “X”), o algoritmo da plataforma identificava, automaticamente, a parte mais relevante dentre os milhões de pixels para mostrá-la ao usuário. Trata-se de uma rede neural de aprendizado profundo treinada a fim de identificar destaques em imagens para identificar rostos de pessoas, textos, entendidos como “objetos de interesse”, sendo abastecido com vastas bases de dados visuais, visando instruí-lo sobre o que constitui um objeto relevante. Tal sistema é configurado para detectar automaticamente áreas de importância nas imagens, empregando aprendizado de máquina para reconhecer padrões e características específicas que definem esses elementos. No entanto, usuários da rede social começaram a notar um estranho padrão na conduta da arquitetura neural. Quando fotos de indivíduos eram postadas, privilegiava-se marcadores físicos de homens brancos. Se houvesse uma pessoa branca e uma negra na imagem, o algoritmo sempre considerava relevante e destacava o tom de pele mais claro, e, de forma similarmente discriminatória, quando se postava fotos de mulheres, a máquina focava em partes sexuais como pernas, coxas e seios.⁶

A partir da ampla divulgação desses resultados, o público da plataforma, mormente, representantes de grupos minoritários, começaram a exigir explicações ao Twitter: como uma plataforma dessa magnitude e impacto poderia permitir algoritmos nitidamente preconceituosos? Quais dados de treinamento estavam sendo utilizados? Quais critérios estavam sendo considerados como os mais relevantes para a análise das fotos? Qual o peso que se dava para marcadores identitários de grupos privilegiados? Em resposta, a rede social reconheceu o erro e explicou que, ao examinarem determinados aspectos que permitiam a explicabilidade do sistema, os parâmetros da rede neural foram treinados com vieses de grupos majoritários, sob a ótica masculina e branca, conforme diferenças demográficas de usuários. Por esse erro evidente, iriam retirar a ferramenta e permitir que os próprios usuários definam onde as suas imagens postadas seriam cortadas. Infere-se a partir deste e outros casos a hipótese de que a automatização de

⁶ G1. Twitter exclui algoritmo de recorte de imagem por viés contra minorias. **Portal G1**, 19 de mai. de 2021. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2021/05/19/twitter-exclui-algoritmo-de-recorte-de-imagem-por-vies-contra-minorias.ghtml>

decisões não somente afeta a esfera individual de cada pessoa, como também a dimensão coletiva de determinado grupo de indivíduos, principalmente aqueles marginalizados socialmente, como mulheres e pessoas pretas.⁷

Desse modo, ao se optar por máquinas realizarem julgamentos decisórios, há necessariamente uma escolha por um processo padronizado por volumes de dados, que, em última instância, podem e irão refletir preconceitos inerentes à realidade social sem uma análise complexa de parâmetros subjetivos para fundamentar a decisão⁸, o que gera perpetuação de assimetrias de poder reflexas nos dados de entrada do treinamento, nas metodologias utilizadas na programação e nas acepções dos desenvolvedores. Isso ocorre devido à mudança paradigmática de algoritmos passarem a decidir ao invés de humanos: as máquinas priorizam aspectos quantitativos em detrimento de elementos subjetivos qualitativos, muitas vezes, sem haver clareza dos pesos aplicados a determinados parâmetros.⁹

Situações como a do Twitter e a de Eric Loomis não são exceções nem raridades, são, na realidade, cada vez mais recorrentes em uma contemporaneidade moldada pela automação algorítmica. A tecnologia vinculada à sociedade da informação permite a coleta e tratamento de dados por parte de *big techs* e governos que revelam quem nós somos (ou a faceta de nós que os interessa), e, assim, há uma dinâmica de estruturação de dados pessoais em grandes bancos de dados¹⁰ com o intuito de revertê-los em tomadas de decisões políticas, sociais e econômicas¹¹ sobre determinado indivíduo e coletividade. Por conseguinte, a nossa individualidade, expressada em nossos gostos, medos, sonhos, pensamentos, vozes e rostos, tornou-se a matéria-prima de uma economia pautada na hipervigilância por meio do uso do *Big Data* e, portanto, nossas pegadas digitais são rastreáveis, registráveis, e acima de tudo, lucrativas.¹²

⁷ G1. Twitter exclui algoritmo de recorte de imagem por viés contra minorias. **Portal G1**, 19 de mai. de 2021. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/2021/05/19/twitter-exclui-algoritmo-de-recorte-de-imagem-por-vies-contra-minorias.ghtml>

⁸ FRAZÃO, Ana. Obstáculos para a consideração de questões éticas nos julgamentos algorítmicos. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial**, São Paulo: Almedina, 2023, p. 37.

⁹ SOLOVE, Daniel J. MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. **Fordham Law Review**, vol. 92, 6 de nov. de 2023, pp. 3-4.

¹⁰ BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, p. 34

¹¹ MENDES, Laura Schertel. Série IDP - Linha de pesquisa acadêmica - **Privacidade, proteção de dados e defesa do consumidor: linhas gerais de um novo direito fundamental**. 1ª Edição. São Paulo: Editora Saraiva, 2014, p. 22.

¹² VÉLIZ, Carissa. **Privacidade é Poder: porque e como você deveria retomar o controle de seus dados**. São Paulo: Editora Contracorrente, 2021, p. 23.

Diante do exposto, em um cenário de acelerado desenvolvimento e lançamento de inteligências artificiais, os modelos de negócio com base em dados (*data-driven models*) se tornaram não somente a realidade empresarial, como o futuro da mudança digital.¹³ Urge, em consonância a disrupção originada pela utilização em massa de IAs, um debate aprofundado acerca dos seus impactos diretos e indiretos em garantias fundamentais e os instrumentos para mitigação desses riscos, e para tal, torna-se indispensável compreender como inteligências artificiais funcionam, como chegam a determinados resultados e como vieses e discriminações algorítmicas são refletidas no processo decisório. Para isso, baseia-se no fundamento de que a regulamentação de IA deve ser orientada pela centralidade da pessoa humana: a ética e filosofia da tecnologia devem ser antropocêntricas, partindo do pressuposto que beneficiar pessoas é compreender que somos intrinsecamente distintos.¹⁴

Assim, a presente monografia leva em consideração que a ausência de explicações claras sobre como determinadas decisões são produzidas pelos algoritmos permite práticas discriminatórias e esbarra em conceitos como devido processo legal, autodeterminação informativa e livre desenvolvimento da personalidade. Por consequência lógica, considera a explicabilidade da automação decisória um meio de usuários se oporem e corrigirem determinações injustas ou incorretas, possibilitando, como resultado, o exercício de outros direitos como os de revisão, intervenção humana, contestação e correção; o cumprimento de princípios de *accountability* e transparência; e, em última instância, seria a expressão do devido processo legal (e devido processo informacional), do livre desenvolvimento da personalidade humana e da autodeterminação informativa em uma sociedade algorítmica moldada por tentativas de controle e manipulação de pessoas.

Diante de tal cenário disruptivo, o presente trabalho pretende, portanto, examinar as dimensões de um direito à explicação em decisões automatizadas no ordenamento jurídico brasileiro e no europeu. Assim, dimensioná-lo será, em outros termos, destrinchar os limites técnicos da explicabilidade das tecnologias responsáveis pela decisão automatizada; fundamentar e defender a sua existência ao explorar os conceitos basilares da doutrina do direito digital; e, com esse arcabouço teórico, investigar as suas origens e desdobramentos doutrinários e

¹³ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial, São Paulo**: Almedina, 2023. pp. 74-75.

¹⁴ ZANATTA, Rafael A.F. A centralidade da pessoa humana entre a filosofia e a regulação. **Jota**, 6 de dez, de 2023. <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/a-centralidade-da-pessoa-humana-entre-a-filosofia-e-a-regulacao-06122023>

legislativos no Brasil e na União Europeia, apresentando as principais particularidades de cada sistema normativo.

As decisões automatizadas podem ser realizadas tanto por meros algoritmos quanto por um complexo sistema de inteligência artificial alimentado por técnicas de aprendizado de máquina e aprendizado profundo¹⁵, sendo o foco da presente monografia a segunda possibilidade. Em ambos os casos, as tecnologias são alimentadas por dados, e, em uma perspectiva prática, há uma tendência a envolver o processamento em massa daqueles de natureza pessoal.¹⁶ A interoperabilidade regulatória entre o paradigma de proteção de dados com o da inteligência artificial é uma conexão necessária, haja vista a necessidade de assegurar os direitos dos titulares de dados, inclusive como uma forma de limitar riscos a direitos fundamentais e assegurar uma abordagem ética no desenvolvimento desses modelos.¹⁷

Por conseguinte, para uma análise completa do objeto de estudo desta monografia, há de se abordar normativos de proteção de dados e propostas de regulamentação de inteligência artificial não separadamente, mas como regulações complementares uma à outra. Tal abordagem decorre da premissa estabelecida de que as leis de proteção de dados podem ser pressupostos basilares da regulação da IA¹⁸, e, inevitavelmente, um modelo regulatório brasileiro ou europeu para disciplinar a inovação de IA deve considerar um robusto panorama legislativo previamente estabelecido na disciplina de proteção de dados.¹⁹

Nessa senda, qualquer análise doutrinária e legislativa do direito à explicação em decisões automatizadas deve ser precedida de um estudo acerca das limitações técnicas da explicabilidade de automação de processos decisórios, sendo a abertura e capítulo inicial da presente monografia. Para tal, é de suma relevância compreender como operam os algoritmos e as técnicas de inteligência artificial, a exemplo do aprendizado de máquina e a sua subespécie, o

¹⁵ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 2.

¹⁶ MACLURE, Jocelyn. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. **Mind and Machines**, v.31, 23 de ago. 2021, p. 421–438

¹⁷ MULHOLLAND, Caitlin; GOMES, Rodrigo Dias de Pinho. Inteligência Artificial e seus principais desafios para os programas de compliance e as políticas de proteção de dados. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas. **Compliance e Políticas de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2021, pp. 164-169.

¹⁸ BIONI, Bruno Ricardo; LUCIANO, Maria. O princípio da precaução na regulação de inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada? In: FRAZÃO, Ana; MULHOLLAND, Caitlin (org.). **Inteligência Artificial e Direito - Ética, Regulação e Responsabilidade**. São Paulo, Thomson Reuters Brasil, 2019, p. 207.

¹⁹ ROSAS, Eduarda Moraes Chacon. **Inteligência artificial: regulação ética a partir das regras de proteção de dados pessoais**. Rio de Janeiro: Editora GZ, 2023, pp. 7-8.

aprendizado profundo. Assim, ao entendê-las, é possível abordar os limites do quanto as decisões tomadas por tais tecnologias são explicáveis ou não, a partir de questionamentos como: a opacidade algorítmica é inerente à tecnologia ou um artifício para escapar de imposições regulatórias? Quais são os limites da explicabilidade sob o ponto de vista técnico? Discriminar é essencial para a tecnologia até que ponto? Máquinas podem decidir? Sobre qualquer coisa ou existem linhas que nunca devem ser cruzadas para inteligências artificiais decidirem sobre?

Dessa forma, a partir da premissa de que direito de se obter explicações em uma decisão ou conjunto de decisões automatizadas é possível, ainda que com determinadas limitações técnicas inerentes aos algoritmos e aos sistemas de inteligência artificial, há de se questionar: por que tal direito deveria existir? De onde decorre a sua necessidade? Quais são os fenômenos sociais da contemporaneidade que o tornam um elemento essencial para a efetivação de outras garantias fundamentais e quais são essas garantias? Esses são os questionamentos que buscarão ser respondidos no segundo capítulo por meio dos conceitos de capitalismo de vigilância, sociedade da informação, autodeterminação informativa e livre desenvolvimento da personalidade.

O terceiro e último capítulo propõe-se a produzir uma comparação histórica, legislativa e doutrinária entre o paradigma do direito à explicação em decisões automatizadas no Brasil e na União Europeia. A análise se concentra na interpretação e aplicação desse direito à luz da Lei Geral de Proteção de Dados Pessoais (LGPD), dos projetos de regulamentação de IA (PL 21/2020, PL 2338/2023, PL 210/2024 e recente relatório final do PL 2338/2023) no Brasil e do Regulamento Geral sobre a Proteção de Dados (RGPD) e AI Act recém-aprovado na União Europeia. A avaliação revela um panorama similar decorrente do “Efeito Bruxelas”, com discrepâncias notáveis na abordagem e concretização desses direitos em ambas as esferas jurisdicionais, principalmente no que se refere à sua existência ou não em cada ordenamento jurídico. Enquanto a LGPD ainda está em seus estágios iniciais de fiscalização e interpretação pelos órgãos judiciários, o RGPD possui uma história mais extensa e sólida de aplicação, com autoridades de proteção de dados mais robustas e atuantes do que a recém-nascida Autoridade de Proteção de Dados (ANPD) brasileira, ainda em seus primeiros passos de *enforcement*. Além disso, as novas propostas de regulamentação de inteligência artificial inserem novas camadas à discussão para além das leis de dados, com novas previsões para o direito à explicação em circunstâncias de alto risco. O intuito desta análise comparativa é, portanto, contribuir para a

compreensão das ramificações práticas e conceituais de tal direito diante do atual cenário de produção, discussão e aprovação de regulações de inteligência artificial a nível global.

A metodologia utilizada para alcançar o objetivo da pesquisa é a de levantamento bibliográfico e pesquisa documental, especificamente com o uso de obras e autores consolidados academicamente na área jurídica de proteção de dados pessoais e inteligência artificial no contexto brasileiro e europeu. O intuito é examinar e responder a problematização apresentada por meio de uma pesquisa exploratória, descritiva e crítica. Assim, pretende-se, em última instância, defender a existência de um direito à explicação no contexto de tratamento de dados pessoais em modelos de inteligência artificial, em prol de uma regulação centrada na pessoa humana, não obstante as limitações técnicas e legais impostas por correntes de autores e pelas próprias legislações do ordenamento jurídico do Brasil e da União Europeia.

1. A DIMENSÃO TÉCNICA DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS

As recentes discussões regulatórias acerca de inteligência artificial trazem à tona recorrentes desencontros referentes aos conceitos que permeiam a tecnologia, haja vista ser um campo científico repleto de termos restritos ao entendimento daqueles que estão cotidianamente em contato. Há uma evidente confusão entre a utilização de “algoritmos”, “IA”, “aprendizado de máquina” e “aprendizado profundo”, pois são frequentemente utilizados como sinônimos entre si, mas se diferem fundamentalmente em aspectos técnicos. O intuito é destrinchar as nuances dessas definições para compreender, posteriormente, o fenômeno da explicabilidade.

1.1. A NATUREZA E AS NUANCES DOS ALGORITMOS DE APRENDIZADO DE MÁQUINA

Um algoritmo básico é comumente comparado, para fins didáticos, a uma receita de bolo, com um passo a passo sequencial que determina como um computador deve operar a fim de executar uma tarefa simples e delimitada.²⁰ Para isso, os dispositivos computacionais são construídos com bilhões de chaves - transistores- ligáveis e desligáveis a bilhões de vezes em um

²⁰ HORA, Nina da. *My News Explica! Algoritmos*. São Paulo: Edições 70, 2023, pp. 19-20.

único segundo que podem ser traduzidos em *bits* de informação conforme o seu estado: ligado (1) e desligado (0). A base dos computadores, portanto, é a matemática, no entanto, as protagonistas não são grandes operações aritméticas com equações e números inimagináveis, em verdade, são complexas estruturas de raciocínio lógico resumidas em operações de “E”, “OU” e “NÃO” que assumem a liderança em um sistema computacional. Desse modo, o algoritmo mais simples existente é girar a chave de um transistor, o segundo é combinar dois *bits* de informação e assim sucessivamente até atingir um nível de complexidade de trilhões de *bits* para atividades comuns de navegação na internet.²¹

Entretanto, a analogia de reduzir algoritmos a receitas culinárias é simplória até demais e não engloba as exigências conceituais necessárias para compreendê-los de forma mais profunda como um objeto de pesquisa. As instruções de uma receita não são precisas, afinal, as medições costumam ser estabelecidas por critérios subjetivos como “xícaras de chá” ou “colheres de sopa”, o que, inevitavelmente, permite que os resultados possam variar entre uma boa comida e um verdadeiro desastre. Em contrapartida, os algoritmos exigem instruções não somente exatas, mas também isentas de ambiguidade, e, com ordenações iguais, geram os mesmos resultados²², sendo este o grande diferencial da disrupção algorítmica: previsibilidade.

Ademais, os algoritmos são heterogêneos. Não há, embora seja um campo de pesquisa em desenvolvimento, um algoritmo central capaz de executar múltiplas atividades. Para o professor especialista em interação humano-máquina da *Northwestern University*, Nicholas Diakopoulos, existem quatro representações algorítmicas amplamente encontradas nas principais aplicações de IA. O primeiro se refere aos algoritmos de ranqueamento, ordenação ou priorização. São aqueles construídos com métricas e parâmetros classificadores por meio de valorações que determinam o que deve ir para o topo da lista, ou para o final.²³ A aplicação mais intuitiva é o motor de busca do Google que seleciona para o usuário quais são as páginas mais relevantes conforme as palavras-chave.²⁴ O autor traz como exemplo do funcionamento e riscos desse modelo algorítmico um famoso caso de ranqueamento de professores elaborado pelo Departamento de Educação de Nova York para avaliar a contribuição dos educadores no

²¹ DOMINGOS, Pedro. **O Algoritmo Mestre**. São Paulo, Novatec, 2017, p. 20.

²² DOMINGOS, Pedro. **O Algoritmo Mestre**. São Paulo, Novatec, 2017, pp. 22-23.

²³ DIAKOPOULOS, Nicholas. **Algorithmic accountability reporting: on the investigation of black boxes**. [S. l.]: Tow Center for Digital Journalism, 2014, p. 5.

²⁴ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 29.

aprendizado de cada aluno, em que organizações da sociedade civil, por meio de um requisito semelhante ao requisito de acesso da Lei de Acesso à Informação (LAI) no ordenamento jurídico brasileiro, obtiveram o ranking e as notas de cada professor. Contudo, análises das estatísticas liberadas identificaram uma correlação de somente 24% entre o rendimento de diferentes estudantes com os respectivos professores, trazendo à tona discussões sobre vieses e ruídos²⁵, as quais serão devidamente exploradas nos tópicos seguintes da presente monografia.

O segundo método são os algoritmos de categorização, comumente confundidos com o modelo anterior. Com efeito, categorizar significa, em vocábulos simples, classificar um elemento em uma classe específica com propriedades que a distinguem de outras classes. A confusão ocorre, por conseguinte, devido à utilização comum de algoritmos de priorização como uma etapa inicial para o uso de algoritmos de classificação. Diakopoulos exemplifica tal aplicação com o sistema de identificação automática de conteúdo do Youtube, o Content ID, o qual identifica e categoriza os vídeos carregados por usuários na plataforma através de um parâmetro de direitos autorais de músicas.²⁶ Em geral, esses métodos algoritmos também são o escopo de processamento de perfis, pois são aqueles capazes de classificar um indivíduo de acordo com inúmeras características associadas a quem são, como traços físicos, gostos pessoais, gênero, gostos pessoais, histórico de compras, idade, geolocalização e outros aspectos pessoais à pessoa.²⁷

Analogamente, o terceiro modelo se refere aos algoritmos de associação, criados com o intuito de descobrir e determinarem correlações entre inúmeros dados de entrada, sendo reiteradamente aplicados em sistemas de detecção de fraudes²⁸. No cenário brasileiro, o Tribunal de Contas da União utiliza o modelo de IA denominado como Alice (acrônimo para “Análise de Licitações e Editais”), responsável por avaliar, mediante a coleta de dados de processos licitatórios e contratos públicos publicizados no Diário Oficial da União e no sistema Comprasnet — o sistema de registro de compras governamentais — possíveis indícios de fraudes com o

²⁵ DIAKOPOULOS, Nicholas. **Algorithmic accountability reporting: on the investigation of black boxes**. [S. l.]: Tow Center for Digital Journalism, 2014, p. 5.

²⁶ DIAKOPOULOS, Nicholas. **Algorithmic accountability reporting: on the investigation of black boxes**. [S. l.]: Tow Center for Digital Journalism, 2014, p. 6.

²⁷ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 28.

²⁸ DIAKOPOULOS, Nicholas. **Algorithmic accountability reporting: on the investigation of black boxes**. [S. l.]: Tow Center for Digital Journalism, 2014, p. 7.

cruzamento de múltiplas informações inferidas e analisadas.²⁹ Outras aplicações comuns são os algoritmos de recomendação de plataformas de streaming e compras online como Spotify, Amazon e Youtube, em que identificam padrões de interesse do consumidor por meio do histórico de navegação e utilização dos aplicativos, e, a partir disso, sugerem opções semelhantes para futuros consumos.³⁰

Por fim, ressaltam-se os modelos algorítmicos de filtragem que, intuitivamente, peneiram os inputs entre o que deve ser considerado relevante ou descartável³¹. Esse algoritmo, hodiernamente, pode ser encontrado como grande caso prático de aplicação nas redes sociais. Quando as linhas do tempo de plataformas como Facebook, Instagram e Twitter (ou “X”) mostram determinado conteúdo ao usuário, na verdade, estão filtrando conforme padrões de comportamento já identificados, por exemplo, por algoritmos de associação.³² Desse modo, é relevante não entender cada um dessas modelações algorítmicas isoladamente, mas sim como engrenagens associadas que operam conjuntamente e dependem necessariamente uma das outras para um sistema de maior magnitude funcione devidamente.

Nessa senda, a inteligência artificial explora os modelos de algoritmos para além de sequências lógicas ordenadas e combina esses variados tipos de modelos algorítmicos em prol de um determinado objetivo previamente especificado no desenvolvimento. Trata-se de uma tecnologia que não é inovadora por si só, com origem que remonta à década de 1950 e com ciclos de progresso e retrocesso nesse meio tempo, à medida que os pioneiros do conceito de IA buscavam, à época, uma meta bastante delimitada: desenvolver máquinas que reproduzem a inteligência humana.³³ A história da IA pode, então, ser compreendida por meio de ondas de modelos com propósitos e técnicas distintas. A primeira onda é conhecida pela inteligência artificial simbólica, um campo baseado em computadores programados por códigos de regras

²⁹ GOMES, Helton Simões. Como as robôs Alice, Sofia e Monica ajudam o TCU a caçar irregularidades em licitações. **Portal G1**. Disponível em: <https://g1.globo.com/economia/tecnologia/noticia/como-as-robos-alice-sofia-e-monica-ajudam-o-tcu-a-cacar-irregularidades-em-licitacoes.ghtml>; Acesso em 12 de mar. de 2024.

³⁰ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 29.

³¹ DIAKOPOULOS, Nicholas. **Algorithmic accountability reporting: on the investigation of black boxes**. [S. l.]: Tow Center for Digital Journalism, 2014, p. 8.

³² MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 29.

³³ LEE, Kai-fu. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos**. Rio de Janeiro: Globo Livros, 2019, p. 19.

lógicas de ordem “se X, então Y”, e assim, nesses modelos as melhorias no desempenho da IA ocorrem por meio do ajuste ou incremento humano de informações na entrada de dados no algoritmo.³⁴ Essa fase também foi denominada como “sistemas de especialistas” ao fazer uso de entrevistas e perguntas produzidas a pessoas com expertise nas áreas de aplicação da IA, em uma tentativa de codificar a sabedoria humana em tomadas de decisões.³⁵

As principais modalidades da segunda onda são o aprendizado de máquina, comumente conhecido na sua expressão na língua inglesa como *machine learning*, e a sua subárea, o aprendizado profundo, e da mesma maneira, conhecido como *deep learning* ou *deep neural networks* - *DNNs*. O aprendizado de máquina aprende com experiências observáveis e não observáveis no mundo fático e se refere a um conjunto de técnicas que automatizam a aprendizagem algorítmica por meio de um treinamento com grandes volumes de dados, recebendo este modelo a denominação de *data-driven* por estes serem os motores operacionais.³⁶ À medida que IAs simbólicas necessitavam de milhares e milhões de linhas de código para conseguirem traduzir ordenações para tarefas complexas, uma IA aprendiz pode estabelecer o mesmo objetivo e alcançá-lo com muito menos, como somente centenas de linhas.

Hodiernamente, o aprendizado de máquina é predominante em cada interação tecnológica de um indivíduo na vida cotidiana. Desde uma simples busca de pesquisa na internet com recomendações personalizadas por máquinas, seleção e classificação de emails como *spams*, sistemas antifraude de bancos em operações de crédito e carros auto dirigíveis até tecnologias utilizadas em significativos projetos de progresso científico como satélites geoestacionários, sequenciadores de DNA e aceleradores de partículas. Pequenas e grandes decisões do dia-a-dia são influenciadas pelo *machine learning*, a exemplo de algoritmos indicação de livros e filmes, de identificação de possíveis terroristas, de escolha de melhores currículos em vagas de emprego e universidades e seleção das ações mais adequadas a um investidor. A probabilidade de uma

³⁴ BOUCHER, Philip. **Artificial Intelligence: how it works? Why does it matter and what we can do about it?** European Parliamentary Research Service, 2020, p. 2. Disponível em: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU\(2020\)641547_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf). Acesso em 5 de mar de 2024.

³⁵ LEE, Kai-fu. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos.** Rio de Janeiro: Globo Livros, 2019, pp. 20-21.

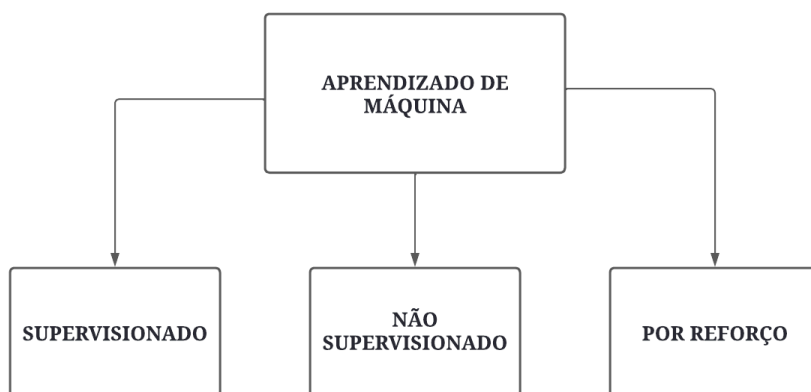
³⁶ BOUCHER, Philip. **Artificial Intelligence: how it works? Why does it matter and what we can do about it?** European Parliamentary Research Service, 2020, p. 2. Disponível em: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU\(2020\)641547_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf). Acesso em 5 de mar de 2024.

pessoa usar um dispositivo com processamento computacional como computadores, celulares e tablets e estar em contato com uma técnica de aprendizado de máquina é altíssima.³⁷

Essa forma de treinamento algorítmico pode ocorrer de três formas principais: (a) aprendizado supervisionado; (b) aprendizado não supervisionado e; (c) aprendizado por reforço). O supervisionado é treinado de forma emparelhada e rotulada, isto é, são inseridas informações de entrada correspondentes às informações de saída, sendo um dos melhores exemplos a classificação de emails como spams, pois a IA é treinada com modelos que são considerados descartáveis e os que não são. O não supervisionado se diferencia justamente no fato de não haver rotulagem, e, assim, a máquina precisa, por contra própria, inferir padrões com os dados que são fornecidos, como é o caso de IAs de algumas IAs de reconhecimento de objetos, recebendo imagens de pedras, papéis e tesouras e sendo responsável por identificar as diferenças entre eles.³⁸

Por fim, o aprendizado por reforço se assemelha ao que humanos conhecem como tentativa e erro ao ser treinado para tomar decisões com base nos dados de saída que melhor se adequem a um objetivo específico. Cada ação positiva é reforçada à medida que ações negativas são descartadas, em um sistema de recompensa. O principal exemplo real para essa técnica é o marketing personalizado para cada usuário a partir de suas informações de navegação.³⁹

Figura 1 - Tipos de aprendizado de máquina



Fonte: esquema elaborado pela autora.

³⁷ DOMINGOS, Pedro. **O Algoritmo Mestre**. São Paulo, Novatec, 2017, pp. 9-12.

³⁸ AWS. **Aprendizado supervisionado versus não supervisionado**. Disponível em: <https://aws.amazon.com/pt/compare/the-difference-between-machine-learning-supervised-and-unsupervised/>. Acesso em 5 de mar de 2024.

³⁹ AWS. **O que é aprendizado por reforço?** Disponível em: <https://aws.amazon.com/pt/what-is/reinforcement-learning/>. Acesso em 5 de mar de 2024.

Em suma, um algoritmo não necessariamente é parte de um sistema de IA, mas obrigatoriamente uma IA faz uso de algoritmos para operar plenamente. Por conseguinte, é uma máquina que aprende com as informações de entrada, incrementa o desempenho com o tempo, e, ao invés de seguir seguir instruções específicas delimitadas previamente, descobre padrões e conexões de forma independente, e, assim, é mais flexível e lida com dados mais complexos do que os modelos algorítmicos básicos⁴⁰. Pode-se dizer, portanto, que “o machine learning é algo novo em nossas vidas: é uma tecnologia que constroi a si própria (...) um algoritmo de aprendizado é como um mestre artesão: cada uma de sua produção é diferente e adaptada primorosamente às necessidades do cliente”⁴¹.

O aprendizado de máquina é, acima de tudo, um ativo econômico poderoso na era da informação, pois o poder se encontra em quem tem os melhores algoritmos aprendizes e a maior quantidade de dados para treiná-los de modo otimizado⁴². Torna-se, em síntese, um ciclo de retroalimentação em que dados entram, máquinas aprendem, mais consumidores são captados, mais clientes significa mais dados, mais dados são modelos melhores e assim sucessivamente, em uma corrida comercial. A verdade é que esse modelo de negócios era inevitável com o surgimento da internet e rios de informações transbordando incessantemente, pois nela “a natureza humana é raspada, arrancada e tomada para o projeto de mercado de um novo século”⁴³. Dessa maneira, empresas que dominam os seus nichos mercadológicos possuem como um dos relevantes diferenciais melhores algoritmos de aprendizado, como é o caso da Netflix e o seu aparente interminável catálogo de filmes e séries, haja vista serem inúmeras e variadas opções de gêneros. Sem bons aprendizes para entender os gostos de cada cliente, analisá-los e recomendar outros conteúdos semelhantes da forma mais eficiente e verossímil, a plataforma se tornaria um caos de aleatoriedade.⁴⁴

As redes neurais de aprendizado profundo são uma subespécie do *machine learning* por ser uma forma de tentativa de replicar o cérebro humano: são formadas por camadas de dados transcritos e analisados, assim como a arquitetura cerebral é composta por camadas de neurônios, a IA neural busca estruturar máquinas por meio de camadas artificiais (*layers*). As entradas (*inputs*), portanto, são as camadas aparentes, com fatores observáveis pelos programadores. No

⁴⁰ HORA, Nina da. *My News Explica! Algoritmos*. São Paulo: Edições 70, 2023, pp. 33-34.

⁴¹ DOMINGOS, Pedro. *O Algoritmo Mestre*. São Paulo, Novatec, 2017, p. 12.

⁴² DOMINGOS, Pedro. *O Algoritmo Mestre*. São Paulo, Novatec, 2017, p. 12

⁴³ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Intrínseca, 2021, p.105.

⁴⁴ DOMINGOS, Pedro. *O Algoritmo Mestre*. São Paulo, Novatec, 2017, p. 30.

meio do caminho, existem as camadas de processamento, intrinsecamente ocultas e não observáveis, sendo estas, a grande questão da ausência de explicabilidade nesse modelo de IA. Por fim, toda essa complexa estrutura matemática é codificada para gerar respostas de saída (*outputs*). Assim, redes com mais camadas (maior profundidade) tendem a apresentar resultados mais promissores no que se refere a eficiência e acurácia de resultados.⁴⁵

Tal arquitetura é um reflexo de uma variedade de fatores determinantes para o design da DNN, como o tamanho e o tipo de dados de entrada, instrumentos de classificação e a capacidade singular de manusear *Big Data* para extrair padrões por meio de filtros capazes de lidar com a volatilidade e complexidade dos dados reais. O grande diferencial das redes neurais decorre justamente de ser capaz de transfigurar a não linearidade do mundo fático e humano em informação estatística.⁴⁶

A exemplo disso, tem-se as IAs de identificação de imagens, como ocorrem nos testes de *CAPTCHA* em que os usuários devem reconhecer objetos e animais, e, assim, essa rede neural possui a capacidade de avaliar milhões de pixels para que o modelo funcione apropriadamente.⁴⁷ Trata-se de uma sigla em na língua inglesa *Completely Automated Public Turing Test to Tell Computers and Humans Apart*, em tradução livre, significa “teste público de Turing completamente automatizado para distinguir computadores e pessoas”. O Google, em 2009, descobriu um modo de utilizar esses micro testes realizados no cotidiano de navegação de internautas para treinar IAs. Assim, possivelmente, ao responder um CAPTCHA, é possível que você esteja criando mais camadas de processamento em redes neurais⁴⁸, como pode ser visualizado no esquema abaixo:

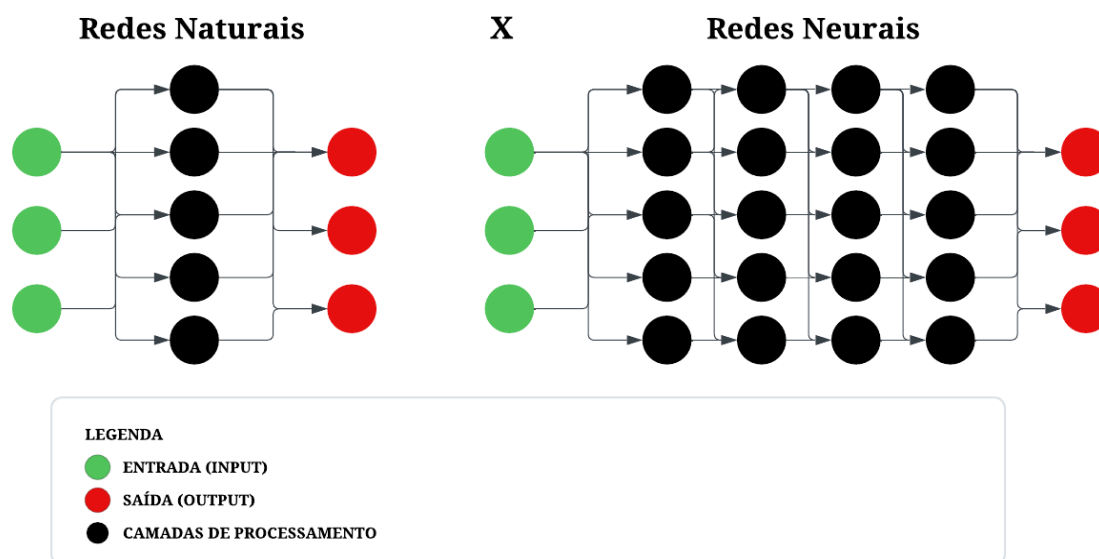
Figura 2 - Diferença visual simplificada de redes naturais e redes neurais

⁴⁵ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, p. 72.

⁴⁶ ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 5.

⁴⁷ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023. p. 75.

⁴⁸ SUPER ABRIL. **Como surgiu o captcha: conheça a história e como funciona o sistema de verificação**. Disponível em: <https://super.abril.com.br/tecnologia/como-surgiu-o-captcha-conheca-a-historia-e-como-funciona-o-sistema-de-verificacao>



Fonte: esquema elaborado pela autora.

Nessa senda, as IAs simbólicas não desapareceram no cenário atual, contudo, há razões para que tenham sido preteridas pelas máquinas de aprendizado profundo, especificamente as redes neurais, destacando-se: (a) a capacidade de processamento amplificada pelo avanço tecnológico; e (b) a quantidade massiva de dados disponíveis na sociedade da informação. Tais fatores permitiram que o *deep learning* sobressaísse e se tornasse fundamentalmente um ativo econômico de *big techs* e governos com conjunto de dados (*datasets*), haja vista poucos agentes econômicos serem capazes de arcar com os custos de supercomputadores e processadores para uma quantidade massiva de dados, e muito menos ter acesso a bancos de informações tão grandes e caros.⁴⁹

A definição de IA está distante de ser um conceito absoluto e consensual em escalas regionais e globais e nos setores acadêmicos e empresariais. Determinados normativos locais e recomendações internacionais podem adotar uma conceituação que convém de modo mais adequado à sua realidade e às suas necessidades. Nesse contexto, a recente regulação aprovada pela União Europeia denominada como AI Act utiliza a definição de sistemas de inteligência artificial estabelecida pela Recomendação do Conselho de Inteligência Artificial elaborado pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE):⁵⁰

⁴⁹ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, p.72.

⁵⁰ OCDE., **Recommendation of the Council on Artificial Intelligence**, 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em 5 de mar de 2024.

Um sistema de inteligência artificial é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, dos dados que recebe, como gerar resultados como previsões, conteúdos, recomendações e decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam nos seus níveis de autonomia e adaptabilidade após a implantação.

No Brasil, em um momento regulatório distinto, devido ao debate legislativo hodierno intensificado em busca de uma “lei de IA”, ainda não estabeleceu o conceito de IA em uma regulação geral como ocorreu no cenário europeu. Contudo, as propostas de legislação apresentam definições com nuances interessantes de serem comparadas e analisadas do ponto de vista técnico. O PL 21/2020, pioneiro na iniciativa de regular a IA no contexto brasileiro, apresenta, no seu art. 2º, uma conceituação que traz delimitações bastante específicas:

Art. 2º Para os fins desta Lei, considera-se sistema de inteligência artificial o sistema baseado em processo computacional que, a partir de um conjunto de objetivos definidos por humanos, pode, por meio do processamento de dados e de informações, aprender a perceber e a interpretar o ambiente externo, bem como a interagir com ele, fazendo previsões, recomendações, classificações ou decisões, e que utiliza, sem a elas se limitar, técnicas como:

I – sistemas de aprendizagem de máquina (machine learning), incluída aprendizagem supervisionada, não supervisionada e por reforço;

II – sistemas baseados em conhecimento ou em lógica;

III – abordagens estatísticas, inferência bayesiana, métodos de pesquisa e de otimização.
Parágrafo único. Esta Lei não se aplica aos processos de automação exclusivamente orientados por parâmetros predefinidos de programação que não incluam a capacidade do sistema de aprender a perceber e a interpretar o ambiente externo, bem como a interagir com ele, a partir das ações e das informações recebidas.⁵¹

Na definição apresentada, a proposta preocupou-se em utilizar termos técnicos específicos do campo de inteligência artificial, com uma relevante ressalva em seu parágrafo único quanto à não aplicação para sistemas que não utilizam o aprendizado de máquina, sendo este, portanto, um requisito necessário para estar sob o escopo da regulação. Tal previsão excluiria, por exemplo, os sistemas simbólicos baseados em regras lógicas da primeira onda de IA.

Em contrapartida, o PL 2338/2023, de relatoria do Senador Rodrigo Pacheco, traz uma abordagem conceitual bastante similar à adotada pelo texto do AI Act (e OCDE) justamente por se encontrar em um contexto de elaboração paralelo ao cenário europeu, como pode ser visualizado:

⁵¹ BRASIL, Congresso Nacional. **Projeto de Lei nº 21 de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/151547>

Art. 4º Para as finalidades desta Lei, adotam-se as seguintes definições:

I – sistema de inteligência artificial: sistema computacional, com graus diferentes de autonomia, desenhado para inferir como atingir um dado conjunto de objetivos, utilizando abordagens baseadas em aprendizagem de máquina e/ou lógica e representação do conhecimento, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões que possam influenciar o ambiente virtual ou real.⁵²

A definição é consideravelmente mais ampla do que a proposta legislativa anterior, principalmente no que se refere à delimitação de “aprendizagem de máquina e/ou lógica e representação do conhecimento”. Infere-se que não há uma limitação da aplicação somente às IAs aprendizes que utilizam do método do *machine learning*. Já o PL 210/2024 de autoria do Senador Marcos do Val adota o conceito mais generalista dos projetos de lei, inclusive, faz uso de poucos termos técnicos específicos:

Art. 2º Para os efeitos desta Lei, são considerados tecnologias de inteligência artificial os sistemas, programas de computador e processos que utilizam recursos de informática, no todo ou em parte, para determinar resultados, tomar ou recomendar decisões, coletar dados ou interagir com indivíduos, comunidades ou máquinas.

Parágrafo único. As tecnologias de inteligência artificial incluem os métodos estatísticos e de processamento de dados para aprendizagem de máquina⁵³

Considerar inteligência artificial como “programas de computador e processos que utilizam recursos de informática” segue uma linha de tentar englobar até mesmo o que não é tecnicamente uma IA, o que, em última instância, gera tremenda insegurança jurídica ao impor uma regulação incapaz de delimitar apropriadamente o objeto regulatório.

Nesse contexto, a presente monografia utilizará “sistemas de inteligência artificial”, “modelos de inteligência artificial” e “máquinas” como sinônimos com fins puramente de simplificação da linguagem de escrita da pesquisa, embora sejam termos não necessariamente referentes aos mesmos objetos em termos de nuances tecnológicas e técnicas do campo da computação. Com as devidas delimitações conceituais feitas a partir do recorte de propostas de lei do ordenamento jurídico brasileiro e europeu, o propósito do presente capítulo busca investigar as nuances técnicas do significado de “decidir”, do conceito de decisão automatizada, do funcionamento das redes neurais de aprendizado profundo, os limites da

⁵² BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁵³ BRASIL, Congresso Nacional. **Projeto de Lei nº 210 de 2024**. Dispõe sobre os princípios para uso da tecnologia de inteligência artificial no Brasil. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/161980>

explicabilidade em decisões realizadas por inteligência artificial e os obstáculos éticos presentes nesse processo decisório, como vieses, discriminação e ruídos. Assim, buscará determinar se o direito à explicação, sob a ótica tecnológica, é possível de ocorrer ou não.

1.2. DECISÕES AUTOMATIZADAS: O CONCEITO DE DECIDIR NO CONTEXTO DA AUTOMAÇÃO

O algoritmo de pagamentos da Uber na cidade de Londres, nos primórdios do sucesso da plataforma, costumava apresentar o preço cobrado dos passageiros, o montante repassado aos motoristas e o percentual retido de cada corrida pela empresa (25%), sendo considerado um exemplo de transparência algorítmica, pois era publicamente sabido como se determinava os preços conforme oferta e demanda em espaços e momentos distintos. Contudo, em 2024, motoristas londrinos se reúnem em uma onda de protestos contra a forma opaca de precificação de corridas sob o argumento de que se tornou impossível compreender o modelo de preços do aplicativo. Ocorrem até mesmo acusações de que a Uber oferece pagamentos personalizados para cada motorista de acordo com as informações pessoais coletadas de cada um, sugerindo um novo algoritmo opaco e desigual, o que é veementemente negado pela plataforma. Motoristas relatam continuar com a mesma quantidade de corridas e carga horário de trabalho, todavia, cada vez mais, recebem menos. Em defesa, representantes da empresa afirmam que utilizam dados agregados, e não dados pessoais de passageiros e motoristas para definir quem recebe uma determinada corrida e o preço que ela custará.⁵⁴

Esse caso traz à tona relevantes discussões sobre opacidade algorítmica e explicabilidade decisória que serão posteriormente abordadas, mas também traz consigo o fato de grandes plataformas de tecnologia, como a Uber, utilizarem massivamente — e serem a base do modelo de negócios de *big techs* e *startups* — sistemas de aprendizado de máquina, primordialmente àqueles de arquiteturas de redes neurais para predições, recomendações e decisões em tempo real, como a política de precificação e sugestão de motoristas e corridas para os passageiros⁵⁵. Assim, não obstante outros modelos de IA eventualmente possam ser aplicados para processos

⁵⁴ MEAKER, Morgan. Drivers Are Rising Up Against Uber's 'Opaque' Pay System. **Wired**, London, 23 de Mai. de 2024. Disponível em: <https://www.wired.com/story/drivers-are-rising-up-against-ubers-opaque-pay-system/>

⁵⁵ UBER. **Engineering More Reliable Transportation with Machine Learning and AI at Uber**. 10 de Nov. de 2017. Disponível em: <https://www.uber.com/en-BR/blog/machine-learning/>

decisórios, devido à constatação de que as DNNs são as modalidades de inteligência artificial mais efetivas e aplicada para essa finalidade⁵⁶, o foco de estudo da presente monografia, em especial deste capítulo, será em relação às propriedades técnicas das redes neurais ao produzirem decisões automatizadas e respectivo dilema da explicabilidade gerado pela natureza dessa implementação tecnológica emergente.

É nesse cenário que emergem polêmicos, mas pertinentes, questionamentos: existem decisões que um modelo de IA nunca deve tomar? Em verdade, uma IA é dotada de capacidade decisória ou é uma atribuição antropomórfica às máquinas? Nessa senda, a definição do substantivo “decisão” pelo Dicionário Online de Português é transcrita da seguinte maneira:

Resolução que se toma após deliberação; sentença ou juízo. [Por Extensão] Tudo aquilo que se resolveu acerca de alguma coisa; deliberação. Habilidade para chegar a uma conclusão ou resolver algum problema sem indecisão.⁵⁷

Nesse contexto, similarmente, a conceituação do verbo “decidir”, com base no dicionário de português do Google, fornecido pela Oxford Languages, é expressa como “emitir (alguém com autoridade ou poder para julgar) juízo final sobre (questão, causa etc.). tomar resolução ou resoluções sobre; deliberar, resolver”⁵⁸. Agora, cabe compreender os aspectos que envolvem a completude do processo decisório como um fenômeno estruturado em etapas complexas.

Com efeito, a fim de responder devidamente às perguntas deste capítulo, James. H. Moor, em seu célebre artigo *Are There Decisions Computers Should Never Make?*, definiu dois significados para tomada de decisões (“*decision making*”). O sentido estrito não engloba o processo de ponderação de variáveis, isto é, restringe-se somente a escolher uma ação, a exemplo de escolher uma carta em um baralho durante um truque de mágica. Não se reflete acerca das motivações por trás da decisão. Em contrapartida, o sentido amplo considera os aspectos conscientes e inconscientes que contornam uma tomada de decisão, como no caso de um jogo de damas, em que cada jogada necessita de uma avaliação de pontos positivos e negativos, de outras possibilidades de movimentos que podem ser feitos, e, finalmente, decidir uma ação com base em toda essa análise.⁵⁹ Em geral, afirma-se que computadores, no sentido estrito, podem decidir ou

⁵⁶ ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 2.

⁵⁷ DECISÃO. In: DICIO, Dicionário Online de Português. Disponível em: <https://www.dicio.com.br/decisao/>

⁵⁸ DECIDIR. In: Google, Dicionário Oxford Languages. Disponível em: [decidir - Pesquisa Google](#)

⁵⁹ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 217.

serem ferramentas auxiliares de processos decisórios, mas não são capazes de tomar decisões no que tange ao sentido mais amplo.⁶⁰

Para o autor, essa constatação é errônea por ser demasiadamente simplista ao enxergar programas computacionais “(...) *simply as a calculator of arithmetic operations. However, computer activity can be understood in many other ways.*”⁶¹ Tal objeção em relação ao senso comum decorre da evolução das IAs simbólicas para o aprendizado de máquina, haja vista “a capacidade de que algoritmos produzam novos códigos e que algoritmos produzam novos algoritmos”⁶², ilustrada pelas máquinas jogadoras de xadrez. No início, todas as regras, possibilidades e probabilidades de uma partida do jogo deveriam ser inseridas pelo programador em um código desenvolvido previamente, no entanto, essas aplicações falharam em vencer os mestres enxadristas humanos. A IA ainda não estava sequer perto de sua missão ao nascer de recriar a inteligência humana. Para tal, os modelos de *machine learning* não são condicionados a seguir parâmetros rígidos estabelecidos no momento do seu desenvolvimento, são, na verdade, treinados para aprenderem com experiências passadas.⁶³

Moor argumenta que o fato do computador utilizar polinômios matemáticos para decidir os seus movimentos, distinguindo-se essencialmente dos parâmetros humanos de tomada de decisão, não significa que isso não o torna um tomador de decisões. O parâmetro de comparação não é decidir como um humano, mas sim se a tecnologia pode decidir como o sentido amplo do termo.⁶⁴ Ademais, ele refuta dois dos principais argumentos comumente adotados por aqueles que não consideram máquinas dotadas de capacidade decisória. O primeiro se refere à incapacidade de um computador decidir de modo consciente como humanos decidem, haja vista os dispositivos não replicarem a consciência, e estarem bem distantes disso. Para contradizer tal argumento, Moor traz situações cotidianas de um indivíduo como dirigir pelo trânsito em que micro e macro

⁶⁰ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, pp. 217-218.

⁶¹ “(...) simplesmente como uma calculadora de operações aritméticas. No entanto, a atividade computacional pode ser entendida de várias outras formas” (em tradução livre para o português). MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 218.

⁶² MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 32.

⁶³ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 32.

⁶⁴ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 219.

decisões são tomadas em menos de segundos, sendo a maior parte delas inconscientemente. A consciência humana varia em distintos momentos decisórios, e, considerando que nem em todas as escolhas os humanos agem conscientemente, para ele, não deve ser um requisito que máquinas o façam.⁶⁵

A segunda alegação é que um computador não decide, pois é somente uma ferramenta de assistência para que um indivíduo tome a decisão. Em refutação, o autor traz a diferenciação entre duas expressões que se confundem entre si: o poder de decidir e a habilidade de decidir. À medida que o poder “*involves being in the appropriate situation and having the authority to make decisions*”⁶⁶ como ser o presidente de uma nação e ter o poder de escolher entre uma política econômica de austeridade e uma de gastos, embora outras pessoas possam ter o conhecimento e experiência exigidos para uma opção tão significativa, ele é o único com a autoridade para tal. Dessa forma, terceirizar esse tipo de decisão impactante significa, em outros vocábulos, conceder o controle, inclusive, no que tange a computadores.⁶⁷

Por fim, com o intuito de delinear e limitar as decisões que uma máquina nunca deve fazer, Moor aponta três premissas indissociáveis e codependentes. A primeira é: as IAs não podem tomar uma decisão sobre assuntos os quais humanos querem decidir sobre. Trata-se de um pressuposto fraco se analisado isoladamente, pois, embora o fato de uma pessoa pretender consideravelmente ser o agente responsável por um processo decisório, outros aspectos relevantes devem ser considerados no *tradeoff* humano vs. máquina. A segunda premissa complementa a anterior, pois determina que uma inteligência artificial nunca deve decidir acerca de temáticas mais bem-sucedidas se decididas por indivíduos. Esse axioma resolve parte do problema do primeiro, no entanto, também traz à tona questões como tarefas perigosas, enfadonhas e demoradas demais se realizadas por pessoas.

Em casos assim, há de se ponderar o custo benefício de máquinas executarem tais tarefas como substitutas, mesmo sendo menos eficientes. A última premissa é a de que IAs não podem tomar decisões que não podem ser ultrapassadas por humanos em um cenário hipotético de máquinas assumirem o controle e nos tornarmos meros figurantes em uma disrupção dos robôs

⁶⁵ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 219.

⁶⁶ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 219

⁶⁷ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 219

no domínio total. Trata-se de um pressuposto que remonta à ficção científica e, assim, deve receber certas precauções, afinal, o intuito do avanço tecnológico é progredir, inclusive no que não somos capazes de fazer.⁶⁸

Para Wimmer e Doneda, decisões tomadas com base em sistemas parcial ou totalmente automatizados podem apresentar falhas. Os autores visualizam duas primordiais. A primeira se refere que é da natureza das máquinas estarem propensas a falhas devido a métodos e técnicas impreciso e não neutros que geram resultados inaceitáveis, em que o exemplo citado por Wimmer e Doneda de uma máquina incapaz de distinguir uma cabeça calva de uma bola de futebol recai no Paradoxo de Moravec: a capacidade cognitiva necessária para exercícios mentais considerados complexos é mais fácil de ser replicada em sistemas de IA do que tarefas simples. A segunda é o uso de decisões de IA para decisões humanas, ensejando questões éticas e de justiça. Assim, dicotomias e dilemas subjetivos de “certo” e “errado” ou ambiguidades são irreprodutíveis no atual estado da arte dessa tecnologia. Com o intuito de mitigá-las, sugerem critérios para equacionar o entrave do “déficit de humanidade” como: (a) transparência e auditabilidade, (b) explicabilidade decisória; e (c) grau de participação humana.⁶⁹

Nesse contexto, desdobram-se dois argumentos: o “*Awful Human Argument*”⁷⁰ e “*Better Together Argument*”⁷¹. O primeiro argumento se funda na premissa de que máquinas decidem melhor que humanos e o segundo segue a trajetória de que máquinas e humanos se complementam para um melhor processo de tomada de decisão.⁷² Para os autores, o aumento da automação decisória leva a uma mudança paradigmática do peso dado a determinados critérios em detrimento de outros ao se decidir, e não necessariamente é considerada uma virada positiva, haja vista que algoritmos alteram a natureza das decisões. Informações quantificáveis prevalecem como elementos determinantes, escanteando dados qualitativos, em um *tradeoff* por mais

⁶⁸ MOOR, James H.. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985, p. 226.

⁶⁹ WIMMER, Miriam; DONEDA, Danilo. “Falhas de IA” e a Intervenção Humana em Decisões Automatizadas: Parâmetros para a Legitimação pela Humanização. **Revista Direito Público**, Brasília, v.18, n. 100, pp. 374-406, out./dez. 2021.

⁷⁰ “O argumento do humano horrível”, em tradução livre.

⁷¹ “O argumento de melhor juntos”, em tradução livre.

⁷² SOLOVE, Daniel J. MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. **Fordham Law Review**, vol. 92, 6 de nov.. de 2023, p. 1

eficiência pelo custo de não incluir emoção, moralidade e juízos de valor, elementos relevantes para se decidir sobre o bem-estar coletivo e individual de pessoas.⁷³

O *Awful Human Argument* baseia-se nas falhas humanas que interferem e prejudicam uma tomada de decisão imparcial e justa, como os preconceitos, a lentidão e ineficiência, a irracionalidade, as emoções e os impulsos típicos da natureza humana. O argumento, então, consiste em considerar máquinas superiores quando se compara a esses erros tão comuns quando as pessoas decidem.⁷⁴ Em contrapartida, o *Better Together Argument* parte do pressuposto que humanos e máquinas, em simbiose, decidem melhor do que humanos sozinhos, pois os modelos algorítmicos fornecem visões, estatísticas e informações adicionais para complementar o que humanos já possuem, em uma perspectiva de quanto mais informação, melhor.⁷⁵

Conclui-se que é um tema em aberto. Não é consenso sobre o que as inteligências artificiais podem decidir, nem se elas propriamente decidem. Para os fins desta monografia, consideram-se os argumentos de Moor em defesa da capacidade e poder das máquinas decidirem, inclusive, a conceituação de tomada de decisão em sentido amplo e que entre os dois principais argumentos quanto às decisões automatizadas, aquele que entende a interação humano-máquina como a melhor maneira de se lidar com elas. Assim, torna-se fundamental atentar-se a outro questionamento do objeto de estudo em análise: o que significa a caracterização “automatizada” sempre associada às decisões realizadas sob o contexto de uma tecnologia auxiliar ou protagonista? O que é, portanto, uma “decisão automatizada”?

O *Article 29 Working Party* da União Europeia elaborou o documento “*Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679*” em que define as decisões automatizadas como:

As decisões exclusivamente automatizadas correspondem à capacidade de tomar decisões através de meios tecnológicos e sem intervenção humana. As decisões automatizadas podem basear-se em qualquer tipo de dados.⁷⁶

⁷³ SOLOVE, Daniel J. MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. *Fordham Law Review*, vol. 92, 6 de nov. de 2023, pp. 2-3.

⁷⁴ SOLOVE, Daniel J. MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. *Fordham Law Review*, vol. 92, 6 de nov. de 2023, pp. 3-4.

⁷⁵ SOLOVE, Daniel J. MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. *Fordham Law Review*, vol. 92, 6 de nov. de 2023, pp. 4-5.

⁷⁶ EUROPEAN COMMISSION. **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679**. Disponível em: <https://ec.europa.eu/newsroom/article29/items/612053>

Analogamente, a automatizada decorre de automação que, conforme a International Society of Automation significa “*the creation and application of technology to monitor and control the production and delivery of products and services*”⁷⁷. De acordo com Renato Leite Monteiro, trata-se de um sentido mais restrito da definição abordada em dicionários, que comumente associam a ações tomadas independente de um operador humano ou, metaforicamente, refere-se à ausência de autonomia de vontade, ocorrido de forma involuntária. Além disso, o autor identifica a inexistência de um “aspecto volitivo”⁷⁸ tanto na definição formal quanto no sentido figurado da expressão.

Para isso, traz o exemplo de aplicativos de geolocalização como Google Maps e Waze, em que os aplicativos, automaticamente, determinam os melhores caminhos para se chegar a um destino. Não há nenhuma participação humana no processo de indicação de rotas, somente no momento de estabelecimento de parâmetros no desenvolvimento da aplicação, o que, em última instância, implica na vontade ser um elemento externo à automatização. Similarmente, o Dicionário Online de Português define automação da seguinte forma:

Funcionamento de uma máquina ou grupo de máquinas que, sob o controle de um programa único, permite efetuar, sem intervenção humana, uma série de operações contábeis, estatísticas ou industriais.⁷⁹

Já o dicionário de português do Google, fornecido pela Oxford Languages, conceitua desta maneira:

Sistema em que os processos operacionais em fábricas, estabelecimentos comerciais, hospitais, telecomunicações etc. são controlados e executados por meio de dispositivos mecânicos ou eletrônicos, substituindo o trabalho humano; automatização.⁸⁰

Por fim, a automação possui classificações como: básica, de processos, de integração e inteligente. A automatização classificada como básica se refere àquela responsável por atividades mais simplórias, em que são empregados instrumentos digitais para centralizar informações e substituir processos fragmentados. Na automação de processos, há uma preocupação com fluxos

⁷⁷ “a criação e a aplicação de tecnologia para monitorar e controlar a produção e entrega de produtos e serviços.” ISA. What is Automation? Disponível em: <https://www.isa.org/about-isa/what-is-automation>. Acesso em: 14 mar. 2024.

⁷⁸ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 58.

⁷⁹ AUTOMAÇÃO. In: DICIO, Dicionário Online de Português. Disponível em: <https://www.dicio.com.br/automacao/>

⁸⁰ AUTOMAÇÃO. In: Google, Dicionário Oxford Languages. Disponível em: [automação definição - Pesquisa Google](#)

de trabalho mais eficientes e transparentes. As outras duas categorizações são as mais relevantes sob o ponto de vista do objeto de estudo em análise, pois se referem àquelas com inteligência artificial com ferramentas auxiliares e substituidoras. A de integração, a exemplo, utiliza máquinas replicadoras de habilidade e comportamentos anteriormente humanos para executar determinadas tarefas, à medida que a automação inteligente é justamente quando a IA toma decisões no lugar de pessoas.⁸¹

A partir das perspectivas doutrinárias e de senso comum apresentadas com intuito de delimitação do objeto de pesquisa, convém explorar as definições estabelecidas nas legislações no ordenamento jurídico brasileiro e europeu. No Regulamento Geral sobre a Proteção de Dados, a definição de automatização decisória influenciou frontalmente a aceção estabelecida posteriormente no Brasil com a promulgação da LGPD, com redações legislativas semelhantes e teologicamente idênticas, como pode ser visualizado no art. 22 da GDPR:

O titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar.⁸²

De forma análoga, no cenário brasileiro, a Lei Geral de Proteção de Dados Pessoais, no seu art. 20, traz tal concepção de automação em situações de processos com tratamento de dados pessoais, a ver:

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.⁸³

De acordo com Monteiro, a lei brasileira conceitua decisão automatizada como aquele interligado a um processo de deliberação simplório de escolha de uma ação, sem levar em consideração outros fatores envolvidos subconscientemente e, assim, adota o significado estrito desenvolvido por Moor. Por conseguinte, na redação legislativa do dispositivo, não ocorre delimitação de aspectos subjetivos dos parâmetros da automação, e toda a discussão ora emergida de quem decide, o homem ou a máquina, torna-se irrelevante sob as lentes regulatórias, pois, na

⁸¹ IBM. O que é automação? Disponível em: <https://www.ibm.com/br-pt/topics/automation>

⁸² PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

⁸³ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

lei, importa tão somente ângulos objetivos. Não há nenhuma menção direta ou indireta acerca dos momentos anteriores à decisão, nem diferenciação se as regras do processo decisório foram determinadas por um programador ou por um algoritmo de aprendizado de máquina, em uma acepção claramente pragmática.⁸⁴

Para o autor, duas propriedades se destacam nas definições adotadas tanto na GDPR quanto na LGPD. A primeira decorre do tratamento de dados necessariamente ter de permitir mais de uma possibilidade de output, haja vista que o direito à revisão explicitado não possuiria lógica de ser assegurado se somente um caminho poderia ter sido alcançado pelo processamento da máquina. Não há o que ser revisto diante de um único resultado possível⁸⁵. Ademais, a segunda característica identificada por Monteiro diz respeito aos efeitos da automação decisória, como explicita:

A redação do artigo, então, utiliza uma perspectiva consequencialista e classifica a sua existência a partir de determinados efeitos. São decisões, para os efeitos do artigo, aquelas que afetam os interesses do titular. Neste sentido, o caminho que os bits, contendo informações pessoais, percorreram, embora automatizado e contingente, não possui, a princípio, efeitos para o titular dos dados. O artigo ainda inclui “[...] as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade”, como forma de apontar que tal efeito não precisa ser necessariamente material, podendo ser meramente potencial, como uma classificação sobre aspectos da personalidade.⁸⁶

Desse modo, considera-se, para os fins da presente pesquisa, que máquinas possuem não somente a habilidade de tomar decisões, como também o poder de decidir na posição de autoridades nas mais variadas aplicações de inteligência artificial. A partir de tal pressuposto, discute-se como entender esse processo decisório. Além disso, devido à presença incontestável do fator ausência humana em todas as conceituações (doutrinária, senso comum, doutrinária e legislativa) de decisão automatizada, enseja a reflexão acerca da correlação entre um direito à explicação e um direito à intervenção humana na tomada de decisões.

⁸⁴ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 67.

⁸⁵ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 59.

⁸⁶ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 68.

1.3. EXPLICABILIDADE DECISÓRIA: UTOPIA TÉCNICA OU PRESSUPOSTO DE TRANSPARÊNCIA E *ACCOUNTABILITY*?

Aviões possuem uma caixa preta que somente pode ser aberta em caso de suspeitas ou comprovados acidentes, pois é um dispositivo de monitoramento de dados em tempo real, e, quando incidentes ocorrem, a caixa revela mistérios que podem ajudar a resolver as causas e eventuais culpados pelo desastre⁸⁷. No entanto, o termo “caixa preta” (*black box*) pode se referir a uma metáfora de um sistema que opera de forma nebulosa, isto é, sem ser possível inferir como as estruturas internas operam, como os resultados são obtidos, e, por consequência, não ser possível identificar cada passo dado até aquele resultado.⁸⁸ Tal expressão é diretamente associada ao funcionamento das redes neurais de aprendizado profundo que possuem o fenômeno *black box* como o maior entrave a ser superado em termos de transparência e *accountability*, mormente no que tange ao uso dessa tecnologia em processos de tomada de decisão que impactam significativamente na esfera individual de pessoas ou na expressão coletiva de grupos minoritários.⁸⁹

A lógica de sistemas de decisões automatizadas se funda no funcionamento fundamental de rede neural: estruturas capazes de lidar com a complexidade de informações presentes no plano mundano, conectá-las e inferir conclusões a partir delas, assim como as redes de neurônios do cérebro humano são evolutivamente preparadas para racionalizarmos a sofisticação do mundo real.⁹⁰ Para assegurar a acurácia de um modelo decisório, exige-se não somente uma vastidão de dados de entrada para que a não linearidade da realidade material seja transcrita em padrões estatísticos, como também uma arquitetura neural, diante do fato de que o fenômeno *black box* decorre diretamente da ausência de informações sobre como uma inteligência artificial treinada com aprendizado de máquina, especialmente a DNN, alcança uma determinada decisão.⁹¹

⁸⁷ VINHOLES, Thiago. Caixa-preta de avião é, na verdade, laranja; entenda por que e como ela funciona. **CNN Brasil**. São Paulo, 09 de out. de 2024. <https://www.cnnbrasil.com.br/economia/caixa-preta-de-aviao-e-na-verdade-laranja-entenda-por-que-e-como-ela-funciona/>. Acesso em 9 de mar. de 2024.

⁸⁸ PASQUALE, Frank. **The Black Box Society: The Secret Algorithms That Control Money and Information**. Harvard University Press, 2015, p. 3.

⁸⁹ ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 2.

⁹⁰ SAMEK, Wojciech; WIEGAND; MULLER, Klaus-Robert. Explainable Artificial Intelligence: understanding, visualizing and interpreting deep learning models. **Cornell University Press**, 28 de Aug. de 2017, p. 1

⁹¹ SAMEK, Wojciech; WIEGAND; MULLER, Klaus-Robert. Explainable Artificial Intelligence: understanding, visualizing and interpreting deep learning models. **Cornell University Press**, 28 de Aug. de 2017, p. 1

Desse modo, parte-se da premissa de que a opacidade algorítmica não era uma preocupação gritante quando os modelos de inteligência artificial eram, mormente, simbólicos ou especialistas. A abordagem tradicional de IA era composta por uma estrutura conhecida e previsível de regras lógicas visíveis durante o ciclo de vida da execução de uma atividade. A ausência de explicabilidade torna-se, portanto, uma problemática com o surgimento do *machine learning*, em especial das redes neurais por serem sistemas cognitivos de aprendizado indutivo. Foi o preço a ser pago por maior acurácia e eficiência em comparação ao paradigma adotado anteriormente.⁹²

Nesse contexto, existem determinados parâmetros para calcular o quanto uma IA é capaz de traduzir o mundo real em estatística. Dentre eles, destaca-se a definição probabilística de acurácia (a frequência de acertos de uma IA em comparação aos erros) e de precisão (a frequência de acertos de resultados positivos) e a revocação (a frequência de falsos resultados negativos). A partir desses parâmetros, é possível avaliar o desempenho da aplicação de uma DNN em diversos setores.⁹³

As DNNs operam, como explicitado previamente, por meio de camadas de processamento, poucas são visíveis e a grande maioria responsável pelo fenômeno “caixa preta” não são observáveis. O funcionamento e integração entre esses *layers* depende de estruturas matemáticas interconectadas denominadas como nó (*nodes*). A semelhança com a arquitetura cerebral decorre justamente da existência desses nós, criados para serem, primordialmente, como neurônios humanos: cada um deles processa um ínfimo fragmento de dados (*features*). Por conseguinte, os nós adquirem um determinado valor ou peso, conectando-se a outros nós em pontos de dados. Isso permite a identificação de padrões de uma forma que IAs simbólicas não eram capazes, e, assim, o aprendizado de máquina permite generalizar em alta escala por meio de estatística.⁹⁴ O funcionamento do CAPTCHA já citado é útil nesse caso, pois cada *feature* processa um pixel de contorno, coloração ou contraste da imagem a ser identificada pelo teste.⁹⁵

No campo das redes neurais, os programadores desenvolveram o procedimento de retroalimentação e autocorreção chamado "*backpropagation*", que ajusta os pesos dos nós e

⁹² MACLURE, Jocelyn. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. **Mind and Machines**, v.31, 23 de ago. 2021, p. 421–438

⁹³ JUNQUILHO, Tainá. **Inteligência Artificial no Direito: limites éticos**. Salvador, Juspodivm, 2022, p. 47.

⁹⁴ MACLURE, Jocelyn. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. **Mind and Machines**, v.31, 23 de ago. 2021, p. 421–438

⁹⁵ MACLURE, Jocelyn. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. **Mind and Machines**, v.31, 23 de ago. 2021, p. 421–438

interconexões após a propagação dos dados pelas camadas, visando aprimorar gradualmente os resultados do algoritmo. Durante o desenvolvimento do algoritmo, programadores precisam ajustar hiperparâmetros e monitorar seu impacto no desempenho. Embora teorias e equações ofereçam alguma orientação, a intuição e um conhecimento não totalmente formalizados são cruciais na fase de design.⁹⁶

Com efeito, o desenvolvimento de grandes *datasets* de disponíveis e da tecnologia do *Big Data* desenvolvida, subverte-se o processo decisório, tradicionalmente influenciado pelo fenômeno da causalidade.⁹⁷ Em geral, os humanos decidem sob o horizonte dos efeitos da decisão, em um pensamento lógico de: se eu decido A, então B ocorre, porém, se eu decido C, conseqüentemente, D acaba por acontecer, e assim sucessivamente, em uma infinidade de possibilidades. O aprendizado profundo escanteia a predominância dessa forma de raciocínio, pois máquinas de redes neurais possuem intrinsecamente uma dificuldade em identificar e compreender causalidades, mormente no que tange a uma ampla gama de variáveis envolvidas, e, por conseguinte, essa nova maneira de aprender traz as correlações como as protagonistas das tomadas de decisões,⁹⁸ diante da percepção de que encontrar padrões, o grande poder das DNNs, é, em síntese, correlacionar em grande escala computacional.⁹⁹

Todavia, há uma problemática inerente à essa disrupção, diante da constatação de que o ato de correlacionar não é neutro. Correlações são tendenciosas, e, muitas vezes, podem ser discriminatórias.¹⁰⁰ A exemplo disso, examina-se as inteligências artificiais utilizadas no processo de avaliação de concessão de crédito. Em muitas delas, empresas e *fintechs* usam endereços postais como parâmetros de correlação, o que permite que um usuário seja prejudicado na sua solicitação devido aos perfilamento do seu bairro apresentar uma alta taxa de inadimplência, gerando, constantemente, distorções de realidade em prol da eficiência estatística, gerando uma

⁹⁶ MACLURE, Jocelyn. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. **Mind and Machines**, v.31, 23 de ago. 2021, p. 421–438

⁹⁷ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial, São Paulo**: Almedina, 2023, p. 75.

⁹⁸ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial, São Paulo**: Almedina, 2023, p. 75.

⁹⁹ MACLURE, Jocelyn. AI, Explainability and Public Reason: The Argument from the Limitations of the Human Mind. **Mind and Machines**, v.31, 23 de ago. 2021, p. 421–438

¹⁰⁰ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial, São Paulo**: Almedina, 2023, p. 76.

“incompatibilidade entre a otimização matemática de alta dimensionalidade, e o raciocínio e a interpretação semântica do ser humano.”¹⁰¹

Torna-se, portanto, uma necessidade regulatória da contemporaneidade questionar se é possível abrir essas caixas pretas de IAs para fins de transparência e explicabilidade de decisões automatizadas. Em verdade, para além de ver o que tem nos sistemas obscuros, é relevante perguntar: é suficiente observar as informações que o interior delas concede? Somos capazes de interpretar e entender? O que significa receber uma explicação de uma decisão? É possível, em termos técnicos, estabelecer um princípio de explicabilidade em aplicações de inteligência artificial? Qual a correlação dessa propriedade com os institutos principiológicos da transparência e *accountability* de um modelo algorítmico? São sinônimos ou englobam aspectos distintos? Responder a alguns desses questionamento será o foco do presente capítulo, a partir da premissa que boa parte das respostas não são consensuais, sendo, na realidade, verdadeiros embates acadêmicos.

Primordialmente, cabe trazer o conceito de explicabilidade. Na Recomendação sobre a Ética da Inteligência Artificial elaborada pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), é apresentado da seguinte maneira no item 40:

A explicabilidade significa tornar inteligível e fornecer informações sobre o resultado dos sistemas de IA. A explicabilidade desses sistemas também se refere à compreensibilidade sobre a entrada, a saída e o funcionamento de cada bloco de construção dos algoritmos e como ele contribui para o resultado dos sistemas. Assim, a explicabilidade está estreitamente relacionada com a transparência, pois os resultados e os subprocessos que conduzem a resultados devem almejar serem compreensíveis e rastreáveis, conforme o contexto.¹⁰²

No documento, ressalta-se a importância de ponderar, de modo equilibrado, a explicabilidade e a transparência haja vista serem institutos passíveis de serem nivelados por outros, como a privacidade, pois não são absolutos e irrestritos. Ademais, a Recomendação prevê a necessidade de usuários serem “plenamente informados quando uma decisão é fundamentada ou tomada com base em algoritmos de IA”¹⁰³ e, nesse contexto, é um direito receberem explicações do responsável pela aplicação de máquina tomadora da decisão, e, por consequência,

¹⁰¹ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial**, São Paulo: Almedina, 2023, p. 78.

¹⁰² UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. 2022. p. 23. Disponível em: unesdoc.unesco.org/ark:/48223/pf0000381137_por

¹⁰³ UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. 2022. p. 23. Disponível em: unesdoc.unesco.org/ark:/48223/pf0000381137_por

possuem a possibilidade de solicitar acesso e revisão desse processo decisório. Na definição da UNESCO, enquadra-se a explicabilidade como um princípio indissociável da transparência, com a ressalva de que ambos são distintos entre si, sendo a transparência definida como:

A transparência tem como objetivo o fornecimento de informações adequadas aos respectivos destinatários, informações essas que permitam sua compreensão e fomentem a confiança. No caso específico dos sistemas de IA, a transparência pode permitir que as pessoas entendam como cada estágio de um sistema é colocado em prática, conforme o contexto e a sensibilidade do sistema de IA. Ela também pode incluir informações sobre os fatores que afetam uma previsão ou decisão específica.¹⁰⁴

Analogamente, a Recomendação do Conselho da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) sobre Inteligência Artificial¹⁰⁵, no tópico 1.3 denominado como *Transparency and explainability*, há a orientação de que atores de IA devem se comprometer em fornecer informações significantes, apropriadas e consistentes, sendo este um princípio para o uso responsável e confiável dessa tecnologia. Para a OCDE, isso deve ocorrer a fim de:

- i. promover uma compreensão geral dos sistemas de IA,
- ii. conscientizar as partes interessadas sobre suas interações com os sistemas de IA, inclusive no local de trabalho,
- iii. para permitir que as pessoas afetadas por um sistema de IA compreendam o resultado e,
- iv. permitir que as pessoas afetadas negativamente por um sistema de IA contestem o seu resultado com base em informações simples e fáceis de compreender sobre os fatores e na lógica que serviu de base para a previsão, recomendação ou decisão.¹⁰⁶

Nessa senda, em ambas as orientações internacionais, explicabilidade e transparência são abordadas de forma conjunta, em uma relação de simbiose, o que também pode ser visualizado na seção de “considerandos” do texto final do AI Act no ordenamento jurídico europeu:

A transparência significa que os sistemas de IA são desenvolvidos e utilizados de forma a permitir uma rastreabilidade e explicabilidade adequadas, sensibilizando ao mesmo tempo os seres humanos para o facto de estarem a comunicar ou a interagir com um

¹⁰⁴ UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. 2022. p. 23. Disponível em: unesdoc.unesco.org/ark:/48223/pf0000381137_por

¹⁰⁵ OCDE. **Recommendation of the Council on Artificial Intelligence**, 2019, p. 8. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em 4 de mar de 2024.

¹⁰⁶ “i. to foster a general understanding of AI systems, i. to make stakeholders aware of their interactions with AI systems, including in the workplace, iii. to enable those affected by an AI system to understand the outcome, and, iv. to enable those adversely affected by an AI system to challenge its outcome based on plain and easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision.” (em tradução livre para o português). OCDE. **Recommendation of the Council on Artificial Intelligence**, 2019, p. 9. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em 4 de mar de 2024.

sistema de IA, informando devidamente os responsáveis pela implantação das capacidades e limitações desse sistema de IA e informando as pessoas afetadas dos direitos que lhes assistem.¹⁰⁷

No ordenamento jurídico brasileiro, o PL 21/2020, pioneiro na tentativa de regulamentar a matéria de inteligência artificial, não traz nenhuma menção aos termos “explicação” ou “explicabilidade” ou “explicar” em todo o texto legislativo e aborda o conceito de transparência, em contramão das recomendações de organizações referenciadas internacionais, da seguinte maneira:

Art. 5º São princípios para o desenvolvimento e a aplicação da inteligência artificial no Brasil

V – transparência: direito das pessoas de serem informadas de maneira clara, acessível e precisa sobre a utilização das soluções de inteligência artificial, salvo disposição legal em sentido contrário e observados os segredos comercial e industrial [...]¹⁰⁸

Em contrapartida, a primeira versão do PL 2338/2023, aborda, analogamente, ambas as definições unidas, mas sem explicitá-la como devem ser definidas e nem de que forma são conceitualmente correlacionadas, com adição de outras propriedades principiológicas para além delas, como a auditabilidade e inteligibilidade, como pode ser visualizado:

Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios:

VI – transparência, explicabilidade, inteligibilidade e auditabilidade;¹⁰⁹

A versão de complementação de voto projeto de lei 2338/23, publicada no dia 4 de julho de 2024, separa a redação do dispositivo princípios em dois incisos distintos, inserindo a “auditabilidade” como outro instituto principiológico, mas, ainda sob a semelhante ótica do panorama global, manteve explicabilidade e transparência associados, com uma nova abordagem de limitações (em caso de segredo comercial e industrial):

Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios:

VI - transparência e explicabilidade, observado o **segredo comercial e industrial**

¹⁰⁷ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

¹⁰⁸ BRASIL, Congresso Nacional. **Projeto de Lei nº 21 de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/151547>

¹⁰⁹ BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline.

VII - devida diligência e auditabilidade ao longo de todo o ciclo de vida do sistema de inteligência artificial, de acordo com o risco envolvido.¹¹⁰ (grifo nosso)

Outro conceito inerente às noções explicitadas é o de *accountability*. Em tradução livre, trata-se de um termo originado da língua inglesa referente à “responsabilização”. Na doutrina de proteção de dados pessoais, surgiu diante de uma pulsante preocupação em estabelecer salvaguardas técnicas, administrativas e legais em um contexto de emergente regulação da disciplina, haja vista uma “ausência de responsabilidade”¹¹¹ de organizações no processamento de dados. Para Bruno Bioni, trata-se do “envelope da mensagem regulatória encaminhada pelas práticas informacionais justas”¹¹², e, assim, era necessário para entidades reguladoras atribuir aos agentes responsáveis pelos processos de tratamento de dados determinados deveres legais.

A concepção de *accountability* que prevaleceu diante do debate acerca dos limites da definição e como ocorreria a especificação de uma ideia tão ampla é a de uma definição prescritiva, entendida como a falta de elementos definidores do que exatamente seria *accountable*, em termos legais, e quais seriam as consequências práticas da responsabilização e como seriam as suas formas de externalização (como costumavam descrever nas leis descrições de códigos de boa conduta, boas práticas organizacionais e políticas de comportamento)¹¹³. Houve uma disputa, entretanto, venceu a acepção “enxuta e camaleão”¹¹⁴ que é, de certo modo, genérica, mas dotada de capacidade de se remodelar. Tal raciocínio se enquadra com o atual contexto regulatório da tecnologia de inteligência artificial, inclusive assumindo a relação simbiótica dos modelos de IA com o processamento de dados.

Em célebre artigo intitulado *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, Ananny e Crawford destringem críticas a inserir a transparência como o ponto focal de modelos de *accountability*. Conforme os autores, supervaloriza-se a noção de que, no momento que for possível visualizar um sistema de IA, abrem-se portas para controlá-lo e governá-lo, como se fosse um estalar de mágica. Nessa senda,

¹¹⁰ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

¹¹¹ BIONI, Bruno Ricardo. **Regulação e proteção de dados pessoais: o princípio da accountability**. Rio de Janeiro: Forense, 2022, p. 17.

¹¹² BIONI, Bruno Ricardo. **Regulação e proteção de dados pessoais: o princípio da accountability**. Rio de Janeiro: Forense, 2022, p. 18.

¹¹³ BIONI, Bruno Ricardo. **Regulação e proteção de dados pessoais: o princípio da accountability**. Rio de Janeiro: Forense, 2022, pp. 25-26.

¹¹⁴ BIONI, Bruno Ricardo. **Regulação e proteção de dados pessoais: o princípio da accountability**. Rio de Janeiro: Forense, 2022, p. 27.

é como se a busca pela transparência fosse, em fato, a busca por uma verdade filosófica da aplicação de inteligência artificial, a partir da premissa de que observar, necessariamente, permite caminhos de mudança e aprimoramentos. Por conseguinte, demonstram como é preciso contornar esse pensamento com novas abordagens regulatórias.¹¹⁵

A virada de chave, portanto, consistiria em interromper a insistência de ver as máquinas sob uma ótica interna e passar a enxergá-las como “sistemas sociotécnicos que não contém complexidade, mas promovem complexidade”¹¹⁶ e, desse modo, romper-se-ia com essa visão tradicional, influenciada pela perspectiva do método científico de correlacionar ter conhecimento sobre determinado objeto de estudo necessariamente como uma jornada de descobrir uma verdade não visível com facilidade, exigindo investigação pormenorizada. Essa lógica ao ser aplicada no sentido da transparência tradicional, caso fosse implementada, levaria a um verdadeiro *accountability*.¹¹⁷ É nesse contexto que Crawford e Ananny explicitam que a noção de transparência não é tão simplória como definem como permitir que tudo seja observável, nítido e claro, mas sim uma promessa de governança.¹¹⁸

No entanto, para o autor e a autora, a transparência, nos moldes como fora construída pela comunidade científica, pelo neoliberalismo e pelo senso comum (no contexto de uma sociedade combativa com corrupção), não passa de um ideal não somente inalcançável, mas estipulado de tal maneira que pode, em muitas situações, ser prejudicial caso seja buscado com tanto afincamento, sem o devido exercício de proporcionalidade e ponderação com outros importantes princípios. Assim, identificam certas limitações a essa acepção.¹¹⁹

Dentre elas, destacam-se as relevantes para a construção do raciocínio do presente capítulo: (a) a transparência pode ser prejudicial, haja vista a possibilidade de ser implementada sem ponderar quais partes de um modelo devem ser reveladas, expondo grupos e indivíduos minoritários já vulneráveis e violando os seus direitos à privacidade e proteção de dados pessoais;

¹¹⁵ ANANNY, M.; CRAWFORD, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, v. 20, n. 3, p. 973–989, 2016.

¹¹⁶ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 62

¹¹⁷ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 62

¹¹⁸ ANANNY, M.; CRAWFORD, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, v. 20, n. 3, p. 973–989, 2016.

¹¹⁹ ANANNY, M.; CRAWFORD, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, v. 20, n. 3, p. 973–989, 2016.

(b) a transparência pode intencionalmente causar opacidade quando organizações utilizam de subterfúgios como conceder, propositalmente, uma quantidade massiva de informações ao ponto de ser impossível identificar quais são relevantes e ilícitas ou não, a fim de intencionalmente estorvar a análise de agentes reguladores, criando uma “transparência resistente”; (c) a transparência pode evocar modelos organizacionais neoliberais ao depositar de informações complexas, em linguagem inacessível, nas mãos do público confiando na capacidade de usuários discutirem e desdobrarem meios de equacionar possíveis irregularidades e desigualdades identificadas, em uma ficção de exacerbado individualismo e, por fim; (d) a transparência pode privilegiar “ver” ao invés de “entender”, pois os verbos significam, evidentemente, ações distintas, e, assim, aprender sobre um sistema de IA não implica necessariamente vê-lo, mas sim interagir com ele e analisar os seus eventuais resultados.¹²⁰

Com efeito, a expressão tão evocada como uma solução milagrosa de abrir a “caixa preta” é materialmente insuficiente, pois não engloba a exigência de entendimento das operações de um sistema de inteligência artificial e não permite, na prática, o real controle da modelação. Assim, para Renato Leite Monteiro, trata-se de “um processo dinâmico de configuração e reconfiguração de plataformas, algoritmos e sistemas de *machine learning* que gerenciam visibilidade”¹²¹, em uma visão de que a questão não envolve visualizar cada uma das camadas de processamento invisíveis de uma rede de aprendizado profundo e engloba, na realidade, a compreensão ampla e analítica dos elementos componentes da arquitetura neural e sua interação com as externalidades da máquina.

Desse modo, a partir das críticas apresentadas ao modelo de *accountability* comumente difundido e adotado em níveis privados e públicos, possibilitou-se uma diferenciação fundacional entre obrigações de transparência e um direito à explicação. Explicar e conceder explicações transborda a mera concepção de que acessar informações garante a clara compreensão de etapas decisórias. Concluiu-se, em síntese, que enxergar não pode ser considerado um sinônimo incontestável de entender, são verbos que delineiam ações frequentemente associadas, mas nunca

¹²⁰ ANANNY, M.; CRAWFORD, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, v. 20, n. 3, p. 973–989, 2016.

¹²¹ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 65.

iguais. Um direito à explicação, exige, portanto, que o indivíduo afetado por uma decisão automatizada receba um esclarecimento cognoscível a seu nível de conhecimento.¹²²

No contexto regulatório de inteligência artificial, *accountability*, portanto, refere-se à capacidade de determinar se a decisão tomada pelo modelo foi produzida conforme padrões de governança, e, caso não tenha sido, haver a possibilidade de responsabilizar alguém pela inconformidade¹²³. Assim, dentre as variadas formas de concretizar no mundo fático o dito camaleão princípio da *accountability* em sistemas de IA, o foco da presente monografia e capítulo será a explicabilidade.

Nesse cenário, convém destrinchar, com maior detalhamento, a explicação tecnicamente possível de ser extraída e exigida legalmente de sua arquitetura opaca e como o debate na literatura recentemente abordou esse apontamento. Para tal, esboça-se um retrato de estudos acerca de uma “*xAI*” (*Explainable Artificial Intelligence*), traduzida, livremente, como a “inteligência artificial explicável”. Os pesquisadores dessa corrente buscam tornar modelos caixa preta de *deep learning* mais compreensíveis sem haver a necessidade de realizar um *tradeoff* entre eficiência e acurácia algorítmica em prol de transparência, com especial enfoque para aplicações de IA em setores sensíveis, como é o caso de modelos utilizados nas áreas militares, bancárias, judiciárias e de saúde.¹²⁴

É de suma relevância entender que não se trata de uma mera minoria da pesquisa no campo da IA em busca de um suspiro de adequação da tecnologia em acelerada inovação com mecanismos de transparência efetivos, mas, em verdade, trata-se de um interesse de grandes instituições governamentais a nível global financiar e desenvolver programas de desenvolvimento de IAs explicáveis, a exemplo de iniciativas de países na linha de frente da corrida tecnológica, a exemplo dos Estados Unidos, o Reino Unido e a França.¹²⁵ As motivações de se construir sistemas explicáveis não se restringem a mero atendimento de requisitos regulatórios e englobam, em fato, maior interação humano-máquina com consertos de *bugs*, pois a opacidade algorítmica

¹²² MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 66.

¹²³ VELEZ, Finale; et. al. Accountability of AI Under the Law: The Role of Explanation. **Berkman Klein Center Working Group on AI Interpretability**, v. 3, 2019, p. 2.

¹²⁴ ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 7.

¹²⁵ VELEZ, Finale; et. al. Accountability of AI Under the Law: The Role of Explanation. **Berkman Klein Center Working Group on AI Interpretability**, v. 3, 2019, p. 3.

também leva a erros operacionais. O design de algoritmos que impactam diretamente em vidas humanas é uma preocupação global.¹²⁶

No senso comum, uma explicação pode ser definida como qualquer informação esclarecedora.¹²⁷ Usamos o termo cotidianamente sem refletir conscientemente os significados inerentes, pois a natureza humana, desde os primórdios, busca explicar, em distintos moldes, o mundo inteligível ao nosso redor. É natural ao homem o desejo por entender. Quando pessoas que conhecemos agem fora do esperado em situações do dia-a-dia, como serem grosseiras repentinamente, estarem tristes em um dia considerado bom, desmarcar um compromisso ou chegar atrasado em outro, exigimos explicações acerca do ocorrido. Não gostamos de ficar no escuro.¹²⁸

Nesse sentido, há um gap de expectativa do que uma explicação significa no campo do aprendizado de máquina e nas áreas de ciências sociais aplicadas (denominadas como “ciências da explicação”). Na seara da filosofia e ciência cognitiva, as explicações requerem certas propriedades que assegurem interatividade social por meio de métodos contrastantes e seletivos e orientadas por “*why-questions*”, levando, em última instância, a uma abordagem *post-hoc* da interpretabilidade do sistema. Para esse campo, o diálogo entre usuários, desenvolvedores, modelos algorítmicos e reguladores é o ponto-chave de uma explicação adequada. Em contrapartida, o termo, para o campo da xAI, as explicações são aproximações simplificadas de processos complexos à luz de modelos científicos, distanciando essencialmente do esperado de uma explicação cognoscível.¹²⁹ Reconhece-se, portanto, que há um considerável esforço para explicações em decisões automatizadas, mas ainda não há consenso dos critérios mais adequados para torná-las acessíveis a um público leigo em linguagem de programação.¹³⁰

Explicações, em geral, significam a comunicação de informações sobre um fenômeno, a exemplo do objeto de estudo desta monografia: a funcionalidade de um modelo ou os critérios e

¹²⁶ BOHLENDER, Dimitri; KÖHL, Maximilian A. **Towards a Characterization of Explainable Systems**. ArXiv [cs], [S.l.], 2019, p. 1.

¹²⁷ VELEZ, Finale; et. al. Accountability of AI Under the Law: The Role of Explanation. **Berkman Klein Center Working Group on AI Interpretability**, v. 3, 2019, p. 3.

¹²⁸ BOHLENDER, Dimitri; KÖHL, Maximilian A. **Towards a Characterization of Explainable Systems**. ArXiv [cs], [S.l.], 2019, p. 1.

¹²⁹ MITTELSTADT, Brent; RUSSELL, Chris; WATCHER, Sandra. Explaining Explanations in AI. **FAT: Proceedings of the Conference on Fairness, Accountability, and Transparency**, 2019, p. 1.

¹³⁰ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 67.

justificativas para uma decisão.¹³¹ Por conseguinte, é compreendida como “a representação de uma informação, que explica um aspecto de um sistema — *explanandum* — para um indivíduo ou um grupo”.¹³² Nas ciências de cognição, pode ser definida como o grau em que se pode elucidar toda a cadeia causal e a necessidade de um evento. Comumente, isso se manifesta na distinção entre explicações científicas e cotidianas. As explicações cotidianas focalizam-se em esclarecer as razões de determinado evento ocorrerem, ao invés de fazer correlações típicas do método científico. Em contrapartida, as explicações interpretáveis após a análise dos dados experimentais de específicas decisões não objetivam revelar o funcionamento intrínseco de um modelo, mas sim ilustrar como ele se comportou e as razões subjacentes a esse comportamento.¹³³

Em geral, a literatura traz critérios para avaliar a efetividade de uma explicação, como a (i) qualidade, que tange à ausência de elementos falsos e aleatórios; (ii) a relevância, para que explicar não seja repleto de um vasto número de informações irrelevantes para a decisão automatizada; (iii) a forma com que a informação chega ao receptor, com linguagem clara, acessível e sem ambiguidade, afastando prolixidade; e, por fim, (iv) a adequação de uma explicação orientada por um contexto, assim, se ela for direcionada a um desenvolvedor experiente em programação, não deverá ter a mesma linguagem e elementos caso seja orientada a um usuário leigo.¹³⁴

Desse modo, as explicações podem assumir dinâmicas e propostas distintas, sem haver uma única forma de ser expressada¹³⁵ como no caso das: (a) explicações locais, as quais buscam elucidar como um modelo fixo chega a uma previsão específica, seja ajustando um modelo local mais simples ao redor de uma decisão específica, seja perturbando variáveis para medir como a previsão se altera; (b) método de explicações baseadas em casos, em que se estabelecem parâmetros para ajustar a distância de similaridade entre dados de treinamento do aprendizado de máquina e os *outputs* gerados em uma determinada situação; (c) explicações de linguagem

¹³¹ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Explaining Explanations in AI. **FAT: Proceedings of the Conference on Fairness, Accountability, and Transparency**, 2019, pp. 2-3.

¹³² MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 68.

¹³³ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Explaining Explanations in AI. **FAT: Proceedings of the Conference on Fairness, Accountability, and Transparency**, 2019, pp. 2-3.

¹³⁴ ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 8.

¹³⁵ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 73.

natural, que analisam as propriedades de *inputs* e correlações com resultados gerados; e (d) técnicas de visualização, as quais o usuário e desenvolver podem participar ativamente no método explicativo por meio de interfaces e recursos visuais.¹³⁶ Além disso, é de suma importância delimitar que são limitadas, pois, em verdade, se nem tudo no mundo inteligível é entendível pelo ser humano, não há que se esperar que em um sistema de inteligência artificial, tudo seja explicável.¹³⁷

Edwards e Vale, no artigo, propõem, com base no texto do Regulamento sobre a Proteção de Dados do ordenamento jurídico europeu e no funcionamento do aprendizado de máquina, duas distinções de possíveis explicações no contexto de decisões automatizadas. A primeira é centrada no modelo, abrangendo lógicas aplicáveis a diferentes sujeitos de decisão, motivações, contexto, variáveis e desempenho do modelo; a segunda trata de previsões específicas de interesse. Mesmo em modelos complexos, algumas informações sobre o “porquê” de uma previsão específica podem ser fornecidas, embora com limitações. As explicações centradas no modelo (“*Model Centric Explanations - MCEs*”) fornecem uma visão abrangente do modelo de aprendizado de máquina, incluindo intenções de modelagem, família do modelo, parâmetros de treinamento, estatísticas dos dados de entrada e saída, métricas de desempenho e lógicas globais estimadas, como escores de importância de variáveis e análise de sensibilidade.¹³⁸

Em contrapartida, explicações centradas no indivíduo (*Subject-Centric Explanations - SCEs*) são formuladas a partir de um registro de entrada específico e são fornecidas em resposta a uma consulta particular, seja ela real, fictícia ou exploratória. Tais explicações podem ser disponibilizadas teoricamente antes ou depois de uma “decisão”, conforme discutido no âmbito da proteção de dados, desde que haja acesso ao modelo. Na área de ciência da computação, esse tipo de explicação é chamado de “local”, pois está restrita à área ao redor de um conjunto específico de dados. Assim, ao se concentrar em partes específicas e relevantes, é possível obter explicações mais úteis. Os autores afirmam existir quatro tipos principais de SCEs: (a) explicações baseadas em sensibilidade, que examinam quais mudanças nos dados de entrada poderiam alterar a decisão; (b) explicações baseadas em casos, que buscam identificar registros

¹³⁶ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Explaining Explanations in AI. **FAT: Proceedings of the Conference on Fairness, Accountability, and Transparency**, 2019, p. 7.

¹³⁷ BOHLENDER, Dimitri; KÖHL, Maximilian A. **Towards a Characterization of Explainable Systems**. ArXiv [cs], [S.l.], 2019, pp 6.-8.

¹³⁸ EDWARDS, L.; VEALE, M. Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. **Duke Law & Technology Review**, v. 16, p. 18-84, 2017, pp. 38-41.

de dados de treinamento mais semelhantes ao registro em questão; (c) explicações baseadas em demografia, que comparam as características de indivíduos que receberam tratamento similar; e (d) explicações baseadas em desempenho, que avaliam a confiança no resultado e a frequência de classificações errôneas em casos semelhantes.¹³⁹

Diante do exposto, defende-se a propriedade de explicabilidade e interpretabilidade de sistemas de IA para que explicações, em suas diversas manifestações, sejam possíveis de serem exigidas e concedidas pelos agentes responsáveis pela aplicação.

A interpretabilidade permite que os desenvolvedores se aprofundem no processo de tomada de decisão do modelo, aumentando sua confiança na compreensão de onde o modelo obtém seus resultados. Em vez de uma simples previsão, a técnica de interpretação fornece uma interface que fornece informações ou explicações adicionais que são essenciais para interpretar o funcionamento subjacente de um sistema de IA.¹⁴⁰

Nessa senda, embora explicações diretas acerca de uma decisão automatizada individual, em muitos cenários, sejam consideradas inexequíveis, excessivamente custosas ou utopias técnicas, não há de se descartar que os modelos, como um ecossistema interativo, não sejam explicáveis nem interpretáveis. Portanto, entende-se a explicabilidade como:

A palavra explicabilidade, em termos de conceitos de IA, significa conhecimento funcional do modelo, cujo objetivo é tentar descrever o comportamento da caixa preta do modelo. É frequentemente usado de forma intercambiável com o termo interpretabilidade na literatura. A explicabilidade expressa o que está ocorrendo no modelo, fornecendo uma explicação legível da decisão do modelo.¹⁴¹

Em síntese, não se pretende exaurir exhaustivamente a listagem de todos os métodos de gerar explicações em inteligência artificial, nem realizar juízos de valor acerca das melhores abordagens, pois não cabe aos objetivos da presente monografia. Pretendeu-se, em verdade, evidenciar que a explicabilidade é possível — ainda que com determinadas limitações técnicas inerentes às arquiteturas de redes neurais — e que as explicações são sociais, embora oriundas de

¹³⁹ EDWARDS, L.; VEALE, M. Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. **Duke Law & Technology Review**, v. 16, p. 18-84, 2017, pp. 38-41.

¹⁴⁰ “Interpretability enables developers to delve into the model’s decision-making process, boosting their confidence in understanding where the model gets its results. Instead of a simple prediction, the interpretation technique provides an interface that gives additional information or explanations that are essential for interpreting an AI system’s underlying functioning” (em tradução livre). ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 6.

¹⁴¹ “The word explainability, in terms of AI concepts, means functional knowledge of the model, the purpose of which is to attempt to describe the model’s black-box behavior. It is often used interchangeably with the term interpretability in the literature. Explainability expresses what is occurring in the model by providing a human-readable explanation of the model’s decision. However, it is difficult to come up with a precise description of what qualifies as an explanation.” (em tradução livre). ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023, p. 7.

máquinas, portanto, a interação entre usuários, desenvolvedores, e aplicações deve ser o ponto-chave¹⁴² ao definir um direito à explicação no contexto de decisões automatizadas, com a participação humana como elemento indispensável para que elas ocorram apropriadamente.

1.4. O SEGREDO DE EMPRESA COMO ELEMENTO LIMITANTE DA EXPLICABILIDADE DE SISTEMAS ALGORÍTMICOS AUTOMATIZADOS

Diante de um cenário de possibilidades variadas de explicações, não é ordinário se questionar: existe um fator limitante para exigí-las? Um titular, no exercício de seus direitos, pode, sempre, querer que uma empresa se explique sobre decisões tomadas com base em seus dados pessoais? Na seara do paradigma de proteção de dados, o *tradeoff* entre transparência e segredo de empresa é uma das mais complexas problemáticas de serem equilibradas, a qual se intensifica com o uso de sistemas de inteligência artificial frequentemente considerados inexplicáveis. Em um caso concreto, qual deve prevalecer? O princípio da transparência ou o segredo de negócios? É possível que ambos sejam respeitados e cumpridos em uma mesma situação, mormente àquelas que se referem à exigência de explicabilidade de decisões automatizadas? A resposta não é trivial e muito menos simplória. A intenção é demonstrar, em perspectiva comparada, como a União Europeia e Brasil lidam com a questão.

No ordenamento jurídico europeu, os termos “segredo comercial”, “segredo de negócios” ou “segredo de empresas” não aparece no textos legislativo do RGPD com tão somente uma menção direta na parte do preâmbulo no Considerando 63¹⁴³ da lei de proteção de dados. No entanto, de forma interpretativa, o posicionamento geral baliza o consenso de que o segredo de empresa não pode ser excessivamente amplo nem ferramenta de escape de imposições legais de conceder informações aos titulares e entidades reguladoras de forma temerária ao fato de que as empresas podem driblar as exigências de transparência do Regulamento sobre a Proteção de Dados, alegando o sigilo corporativo. Contudo, as diretrizes elucidam que, embora exista “certa proteção” contra revelar segredos comerciais, as empresas não podem usar essa proteção como

¹⁴² MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Explaining Explanations in AI. **FAT: Proceedings of the Conference on Fairness, Accountability, and Transparency**, 2019, p. 6.

¹⁴³ (63) (...) Quando possível, o responsável pelo tratamento deverá poder facultar o acesso a um sistema seguro por via eletrónica que possibilite ao titular aceder diretamente aos seus dados pessoais. Esse direito não deverá prejudicar os direitos ou as liberdades de terceiros, incluindo o **segredo comercial** ou a propriedade intelectual e, particularmente, o direito de autor que protege o software. (grifo nosso).

pretexto para negar acesso ou recusar-se a fornecer informações. Embora isto não elimine a exceção aos segredos comerciais do Considerando 63 do RGPD, incentiva as autoridades de proteção de dados a monitorarem quando empresas supostamente alegam a não possibilidade de exercer obrigações de *accountability* com base em um amplo e genérico conceito de segredo de empresa. Dessa forma, as orientações do *Working Party* 29 estabelecem explicitamente que o segredo comercial não pode se sobrepôr à transparência algorítmica.¹⁴⁴

O AI Act, em contrapartida, aborda os termos não somente no preâmbulo, como também nos dispositivos legais, mas com foco nas obrigações entre fornecedores de IA de uso geral e aquelas aplicações de IA que irão utilizá-la. O texto traz no seu Considerando 107 a transparência como elemento a ser seguido nos modelos de IA de propósito geral, apesar do segredo comercial, com a seguinte disposição:

Para aumentar a transparência dos dados utilizados na pré- formação e treinamento de modelos de IA de uso geral, incluindo textos e dados protegidos por direitos autorais lei, é adequado que os fornecedores de tais modelos elaborem e disponibilizem publicamente um resumo suficientemente detalhado do conteúdo usado para treinar o programa de uso geral modelo. Embora tendo em devida conta a necessidade de **proteger segredos comerciais e informações confidenciais informações comerciais**, este resumo deve ser geralmente abrangente em seu escopo em vez de tecnicamente detalhado para facilitar as partes com interesses legítimos, incluindo titulares de direitos de autor, para exercerem e fazerem valer os seus direitos ao abrigo do direito da União, por exemplo, através listando as principais coleções ou conjuntos de dados usados no treinamento do modelo, como grandes bancos de dados ou arquivos de dados privados ou públicos, e fornecendo uma explicação narrativa sobre outras fontes de dados usadas (...) (grifo nosso)¹⁴⁵

Esse considerando se relaciona diretamente com o art. 53 do AI Act que prevê as obrigações para os provedores de modelos de IA de propósito geral, como a obrigatoriedade de disponibilizar informações e documentação acerca da modelagem da inteligência artificial, como pode ser visto:

1. Os fornecedores de modelos de IA de uso geral devem: (...) (b) Elaborar, manter atualizados e disponibilizar informações e documentação aos fornecedores de sistemas de IA que pretendam integrar o modelo de IA de uso geral em seus sistemas de IA. Sem prejuízo da necessidade de **respeitar e proteger os direitos de propriedade intelectual e as informações comerciais confidenciais ou os segredos comerciais**, em conformidade com o direito da União e o direito nacional, as informações e a documentação devem: (i) permitir que os fornecedores de sistemas de IA tenham uma

¹⁴⁴ KAMINSKI, M. E. The Right to Explanation, Explained. *Berkeley Technology Law Journal*, v. 34, n. 189, 2019, p.15.

¹⁴⁵ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

boa compreensão das capacidades e limitações do modelo de IA de uso geral e cumpram as suas obrigações nos termos do presente regulamento;¹⁴⁶ (grifo nosso).

Além disso, destaca-se o art. 74 do AI Act, que prevê disposições legais acerca da fiscalização dos sistemas de IA, trazendo as principais atribuições e poderes das autoridades reguladoras competentes. Em específico, examina-se as previsões que tratam do *tradeoff* entre transparência e segredo de negócios:

(12) Sem prejuízo dos poderes previstos no Regulamento (UE) 2019/1020, e sempre que pertinente e limitado ao necessário para o desempenho das suas funções, os prestadores devem conceder às autoridades de fiscalização do mercado total acesso à documentação, bem como aos conjuntos de dados de treino, validação e teste utilizados para o desenvolvimento dos sistemas de IA de risco elevado, inclusive, se for caso disso e sob reserva de salvaguardas de segurança, através de interfaces de programação de aplicações ou outros meios e ferramentas técnicas pertinentes que possibilitem o acesso remoto

(13) Deve ser concedido às autoridades de fiscalização do mercado **o acesso ao código-fonte dos sistemas de IA de risco elevado** mediante pedido fundamentado e apenas se estiverem preenchidas ambas as condições que se seguem: a) **O acesso ao código-fonte é necessário para avaliar a conformidade de um sistema de IA de risco elevado** com os requisitos estabelecidos no capítulo III, secção 2; e, b) Os procedimentos de testagem ou auditoria e as verificações com base nos dados e na documentação apresentados pelo prestador foram esgotados ou revelaram-se insuficientes.¹⁴⁷ (grifo nosso)

Desse modo, de forma análoga, o AI Act segue a linha de não considerar o segredo de negócios um instituto jurídico capaz de transpassar as obrigações de transparência, assim como ocorre na interpretação sistemática do RGPD: o fornecimento de informações e documentos não necessariamente devem conter conteúdo estratégico confidencial e sigiloso de uma empresa, no entanto, em casos de aplicações de alto risco, prevê-se a possibilidade de acesso a códigos fontes em prol da mitigação dos riscos a direitos fundamentais em casos específicos e limitados. Trata-se, portanto, de um desafio das autoridades de controle europeias em como efetivamente realizar esse alinhamento.

Em contrapartida, no ordenamento jurídico brasileiro, ainda não há entendimentos consensuais interpretativos de como lidar com esse *tradeoff* — com a LGPD citando “segredo comercial e industrial” mais de uma dezena de vezes¹⁴⁸ a serem protegidos e respeitados pela

¹⁴⁶ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

¹⁴⁷ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

¹⁴⁸ FRAZÃO, Ana. Transparência de algoritmos x segredo de empresa. **Jota**, 9 de jun. de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/transparencia-de-algoritmos-x-segredo-de-empresa-09062021>

regulamentação, sem explicitar como e deixando a cargo da ANPD dispor detalhadamente sobre¹⁴⁹ — como podem ser visualizados no texto legal do princípio da transparência¹⁵⁰, no direito à informação prévia acerca do tratamento de dados¹⁵¹, restrições ao relatório de impacto à proteção de dados pessoais no caso da base legal do legítimo interesse¹⁵² e o direito à informação no contexto de decisões automatizadas.¹⁵³

Como consequência lógica, a ausência de orientação da entidade reguladora permite a discricionariedade de juízes e tribunais construírem compreensões a partir das situações fáticas sob as suas jurisdições. Conforme o relatório “Painel LGPD nos Tribunais”, o principal artigo relacionado aos direitos dos titulares mais citados em decisões judiciais brasileiras é o de revisão de decisões automatizadas com o art. 20 da LGPD atingindo o maior número de menções. A conclusão do relatório é de que os pedidos, em sua maioria, são indeferidos e ocorrem no contexto aplicativos de transporte com situações de (i) de reconhecimento de vínculo empregatício ou (ii) revisão de banimentos e restrições de usuários, principalmente motoristas. Os outros casos se referem a exclusões e restrições de acesso a contas em redes sociais, em uma lógica similar.¹⁵⁴

Esses dados estatísticos podem ser exemplificados com os constantes casos de recusa de perícias algorítmicas no contexto de conflitos jurisdicionais entre motoristas de aplicativo e plataformas como Uber e 99 devido à proteção expressa de informações intrínsecas ao segredo de negócios diante dos pedidos de acesso a elementos como “distribuição de corridas, a definição de valores a serem cobrados e repassados, as restrições ou preferências no acesso a chamados, além

¹⁴⁹ Art. 55-J. Compete à ANPD: II - **zelar pela observância dos segredos comercial e industrial**, observada a proteção de dados pessoais e do sigilo das informações quando protegido por lei ou quando a quebra do sigilo violar os fundamentos do art. 2º desta Lei e; X - dispor sobre as formas de publicidade das operações de tratamento de dados pessoais, **respeitados os segredos comercial e industrial**; (grifo nosso).

¹⁵⁰ 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** (grifo nosso).

¹⁵¹ Art. 9º O titular tem direito ao acesso facilitado às informações sobre o tratamento de seus dados, que deverão ser disponibilizadas de forma clara, adequada e ostensiva acerca de, entre outras características previstas em regulamentação para o atendimento do princípio do livre acesso: II - forma e duração do tratamento, **observados os segredos comercial e industrial**; (grifo nosso)

¹⁵² Art. 10 (...) § 3º A autoridade nacional poderá solicitar ao controlador relatório de impacto à proteção de dados pessoais, quando o tratamento tiver como fundamento seu interesse legítimo, **observados os segredos comercial e industrial**. (grifo nosso)

¹⁵³ Art. 20 (...) § 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.

¹⁵⁴ MENDES, Laura Schertel; FUJIMOTO, Mônica (org.). **Painel LGPD nos Tribunais**. Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, 2024, pp. 79-80.

das trocas de mensagens”.¹⁵⁵ O entendimento do Tribunal Superior do Trabalho (TST) é de que não é possível permitir que se acesse esse tipo de informação das empresas.¹⁵⁶

Nesse contexto, os projetos de regulamentação de IA no Brasil também trazem à tona novas camadas à discussão. O texto inicial do PL 2338/23 cita somente duas vezes o termo “segredos industrial e comercial” no contexto de conclusões de avaliações de impacto algorítmico e bases de dados públicas, à medida que o PL 21/20, ao tratar da transparência, estabelece que esse instituto principiológico deve observar o sigilo de negócios, sendo este um elemento limitante¹⁵⁷ e, por fim, o PL 210/24 não cita o termo nenhuma vez.

Analogamente, o parecer relativo à Complementação de Voto PL 2338/2023 e outros do dia 4 de julho de 2024 (a última versão de texto legal do PL 2338/23 ao momento da escrita do presente trabalho) acatou as ideias das emendas 75 e 96 para estabelecer como fundamentos no uso da IA no Brasil o segredo comercial e industrial¹⁵⁸, os princípios da transparência e explicabilidade¹⁵⁹ também estão sujeitas aos limites desse instituto jurídico e o direito à explicação em decisões automatizadas em sistemas de alto risco também é limitado pelo segredo de negócios¹⁶⁰, além de outras menções expressas a esse mecanismo de proteção em dispositivos legais referentes à auditoria, à governança e às avaliações de impacto.

Ademais, na complementação do PL 2338/23, destaca-se um poder regulamentar concedido de acessar informações acerca dos sistemas no âmbito da Administração Pública,

¹⁵⁵ PAIVA, Letícia. Perícia em algoritmo da 99 também é suspensa pelo TST. **Jota**, 1 de jun. de 2021. Disponível em: <https://www.jota.info/tributos-e-empresas/trabalho/pericia-algoritmo-aplicativo-99-suspensa-tst-01062021>

¹⁵⁶ PAIVA, Letícia. Perícia em algoritmo da 99 também é suspensa pelo TST. **Jota**, 1 de jun. de 2021. Disponível em: <https://www.jota.info/tributos-e-empresas/trabalho/pericia-algoritmo-aplicativo-99-suspensa-tst-01062021>

¹⁵⁷ (...) V – transparência: direito das pessoas de serem informadas de maneira clara, acessível e precisa sobre a utilização das soluções de inteligência artificial, salvo disposição legal em sentido contrário e observados os **segredos comercial e industrial**, nas seguintes hipóteses (...) c) sobre critérios gerais que orientam o funcionamento do sistema de inteligência artificial, assegurados **os segredos comercial e industrial**, quando houver potencial de risco relevante para os direitos fundamentais; (grifo nosso).

¹⁵⁸ Art. 2º O desenvolvimento, a implementação e o uso de sistema de inteligência artificial no Brasil têm como fundamentos: XVII - proteção de direitos de propriedade intelectual e ao segredo comercial e industrial; In: BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline

¹⁵⁹ Art. 3º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial observarão a boa-fé e os seguintes princípios: VI - transparência e explicabilidade, observado o **segredo comercial e industrial**; (grifo nosso). BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline

¹⁶⁰ Art. 6º (...) Parágrafo único. A explicação solicitada no âmbito do inciso I, respeitando o **segredo comercial e industrial**, incluirá informações suficientes, adequadas e inteligíveis, nos termos do Regulamento (grifo nosso). BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível

dentre as atribuições da autoridade competente em fiscalizar o uso de IA no escopo do projeto de lei:

Art. 44. Cabe à autoridade competente:

IV – solicitar, a qualquer momento, às entidades do poder público que desenvolvam ou utilizem sistemas de inteligência artificial, informe específico sobre o âmbito, a natureza dos dados e os demais detalhes do tratamento realizado, com a possibilidade de emitir parecer técnico complementar para garantir o cumprimento desta Lei;¹⁶¹

Além disso, nesse dispositivo legal, também contém as atribuições de realização de auditorias e validação desses procedimentos, em respeito aos segredos comercial e industrial, em mais uma forma aberta e ampla de ponderação entre a transparência e o sigilo de negócios:

Art. 44. Cabe à autoridade competente:

VII – realizar auditorias internas de sistemas de inteligência artificial quando necessária para a aferição de conformidade com esta Lei, garantido o tratamento confidencial das informações em atenção aos segredos comercial e industrial (...)

IX – credenciar instituições, mediante critérios estabelecidos em regulamento sujeito a consulta pública, para acesso a dados para fins de auditorias, garantido a confidencialidade da informação em atenção aos segredos comercial e industrial;¹⁶²

Uma possível solução razoável e proporcional para o *tradeoff*, ou um meio-termo, seria o conceito de transparência qualificada de Frank Pasquale. O autor também se alinha ao pensamento de transparência “completa” ou excessiva não é satisfatória, inclusive, considera que seria uma grande problemática para a sociedade em termos de violações de privacidade e propriedade intelectual. Para ele, é importante que um seleto grupo de especialistas seja responsável pelas investigações algorítmicas, limitando as informações acessadas para respeitar o interesse de todos os envolvidos.¹⁶³ Entende-se, portanto, que essa formatação de transparência permitiria “que determinados experts possam atestar a qualidade, a validade e a confiabilidade dos sistemas algorítmicos, mantendo o segredo de negócios”¹⁶⁴

Depreende-se, portanto, a partir do exposto, ser exacerbada a concepção de que exigir explicações de decisões automatizadas necessariamente implica em violações do segredo

¹⁶¹ BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline

¹⁶² BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline

¹⁶³ PASQUALE, Frank. **The Black Box Society: The Secret Algorithms That Control Money and Information**. Harvard University Press, 2015, p. 142.

¹⁶⁴ FRAZÃO, Ana. Transparência de algoritmos x segredo de empresa. **Jota**, 9 de jun. de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/transparencia-de-algoritmos-x-segredo-de-empresa-09062021>

comercial e industrial, haja vista ser um artifício comumente utilizado por empresas para contornarem obrigações de transparência e *accountability*. Em verdade, as explicações e explicabilidade não exigem códigos-fonte nem conteúdos estratégicos sigilosos e confidenciais, mas sim informações significativas interativas, cognoscíveis e capazes de dialogar com o receptor acerca dos parâmetros que ensejaram o resultado da tomada de decisão.¹⁶⁵ Outros procedimentos de prestação de contas podem vir a exigir registros e documentações que possam diretamente entrar em conflito com o sigilo de negócios. Há de se refletir acerca dos riscos individuais e coletivos de empresas utilizarem o instituto jurídico do segredo de negócios como verdadeiros “escudos”, mormente em disputas judiciais.¹⁶⁶

2. A DIMENSÃO EXISTENCIAL DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS

Em vista da possibilidade técnica de estabelecer parâmetros e medidas capazes de gerarem não somente explicações de decisões automatizadas individuais, como também explicabilidade de modelos de inteligência artificial, ainda mais em uma emergente corrente de *explainable AI*, urge a questão: um direito à explicação deveria ser assegurado e debatido no Brasil com base em que fundamentos jurídicos? Para atender a tal questionamento, explorar-se-á a evolução de institutos como a privacidade e a proteção de dados pessoais¹⁶⁷ como direitos subjetivos fundamentais de todo e qualquer cidadão brasileiro, mormente no contexto de decisões automatizadas tomadas com o processamento de dados que identificam ou possam identificar um indivíduo em modelos de inteligência artificial. A trajetória do direito à proteção de dados pessoais será relevante para identificar como direitos podem surgir e se concretizar em dado ordenamento jurídico, e aplicar a lógica semelhante para a defesa de um direito à explicação decorrente da materialidade do escopo de proteção dos dados pessoais e de institutos

¹⁶⁵ FRAZÃO, Ana. Transparência de algoritmos x segredo de empresa. **Jota**, 9 de jun. de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/transparencia-de-algoritmos-x-segredo-de-empresa-09062021>

¹⁶⁶ FRAZÃO, Ana. Transparência de algoritmos x segredo de empresa. **Jota**, 9 de jun. de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/transparencia-de-algoritmos-x-segredo-de-empresa-09062021>

¹⁶⁷ Art. 5º Para os fins desta Lei, considera-se: I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável. A ver: BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

princípios que também regem a estrutura infraconstitucional como a autodeterminação informativa e o livre desenvolvimento da personalidade.

Em 2020, em um cenário de disrupção da pandemia de COVID-19, com fragilidades atenuadas nas esferas políticas, econômicas e sociais, a temática de proteção de dados emergiu no plenário do Supremo Tribunal Federal (STF) com a edição da Medida Provisória n.º 954 (MP 954) pelo governo Bolsonaro para permitir o compartilhamento de dados por empresas de telecomunicações durante a emergência de saúde pública com o Instituto Brasileiro de Geografia e Estatística (IBGE). Desse modo, as operadoras de telefonia eram, por lei, obrigadas a repassar informações pessoais como telefone celular e endereço, sob a justificativa se dava diante da impossibilidade dos recenseadores realizarem entrevistas com a população de modo presencial diante do distanciamento social estabelecido para fins de segurança pandêmica.¹⁶⁸

Em decisão paradigmática, o STF, com ampla maioria de 10 votos seguindo a relatoria da Ministra Rosa Weber, reconheceu o direito fundamental autônomo à proteção de dados pessoais em julgamento das Ações Diretas de Inconstitucionalidade (ADIs) ajuizadas por diversos partidos políticos contrários à MP. A Ministra centrou os argumentos contrários à medida governamental ao redor da relevância dos dados pessoais serem protegidos sob a égide da Constituição em um Estado Democrático de Direito, haja vista a situação excepcional de crise sanitária não ser suficiente para justificar medidas de vigilância em massa. Para Rosa Weber, uma vez coletados esses dados, torna-se uma árdua tarefa garantir que não serão utilizados para finalidades distintas dos fins de recenseamento durante o período da pandemia. Assim, destacam-se dois elementos fundantes do voto da relatora: a ruptura da falácia do dado neutro e o reconhecimento como um direito autônomo dissociado de outras garantias fundamentais.¹⁶⁹

A ministra rompe com a falsa aceção — inclusive estabelecida previamente em outros julgamentos do STF — de que existe uma categoria específica de dados a serem protegidos constitucionalmente, isto é, aqueles considerados excessivamente íntimos ou compreendidos

¹⁶⁸ MENDES, Laura Schertel. Decisão histórica do STF reconhece direito fundamental à proteção de dados pessoais. **Jota**, 10 de mai. de 2020. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/decisao-historica-do-stf-reconhece-direito-fundamental-a-proteco-de-dados-pessoais-10052020>.

¹⁶⁹ MENDES, Laura Schertel. Decisão histórica do STF reconhece direito fundamental à proteção de dados pessoais. **Jota**, 10 de mai. de 2020. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/decisao-historica-do-stf-reconhece-direito-fundamental-a-proteco-de-dados-pessoais-10052020>.

como sensíveis.¹⁷⁰ Em uma economia movida pela tecnologia do *Big Data* e com grandes bancos de dados em operação na iniciativa pública e privada para construção de perfis de cidadãos brasileiros, não há possibilidade de dados passíveis de identificar uma pessoa sejam considerados irrelevantes ou neutros, haja vista serem ativos econômicos em uma sociedade informacional. Ademais, nesse contexto de processamento em massa de dados, o direito à proteção dos dados pessoais deve se desvincular do paradigma da privacidade e intimidade, pois não mais o direito de ser deixado só é capaz de lidar com os entraves da sociedade da informação, e, assim, passa a ser um direito que vai para além de ser equiparado ao sigilo e assume contornos singulares.¹⁷¹

No ano de 2022, em consonância ao entendimento adotado pelo órgão máximo do Poder Judiciário, o Legislativo promulgou uma emenda constitucional (EC 115) para alterar o texto constitucional a fim de a proteção de dados pessoais constar no rol de garantias fundamentais do art. 5º, com a atribuição da competência exclusiva da União para legislar sobre o tema. Para o relator na Câmara dos Deputados, a nova emenda, conjuntamente a LGPD e o Marco Civil da Internet fixam as bases de uma arquitetura normativa para um Brasil cada vez mais digitalizado, com a seguinte redação legislativa presente na CRFB/88:¹⁷²

Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes:
LXXIX - é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais.¹⁷³

Por conseguinte, reconhece-se a dimensão subjetiva e objetiva desse novo direito fundamental, entendidas, respectivamente, como a capacidade de ensejar uma pretensão de determinado comportamento ou na produção de efeitos em relações jurídicas e de instituir

¹⁷⁰ Art. 5º Para os fins desta Lei, considera-se: II - **dado pessoal sensível**: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural; (grifo nosso).

¹⁷¹ MENDES, Laura Schertel. Decisão histórica do STF reconhece direito fundamental à proteção de dados pessoais. **Jota**, 10 de mai. de 2020. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/decisao-historica-do-stf-reconhece-direito-fundamental-a-protecao-de-dados-pessoais-10052020>.

¹⁷² OLIVEIRA, José Carlos. Promulgada PEC que inclui proteção de dados pessoais entre direitos fundamentais do cidadão. **Agência Câmara de Notícias**, 10 de fev. de 2022. Disponível em: <https://www.camara.leg.br/noticias/850028-PROMULGADA-PEC-QUE-INCLUI-A-PROTECAO-DE-DADOS-PESSOAIS-ENTRE-DIREITOS-FUNDAMENTAIS-DO-CIDADA0>. Acesso em 17 de mai. de 2024.

¹⁷³ BRASIL. Constituição (1988). Constituição da República Federativa do Brasil. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

influência principiológica e norteando todo o Poder Público.¹⁷⁴ Assim, a proteção de dados pessoais configura uma nova ordem valorativa, expandindo-se para além do escopo constitucional, como um ideal a ser perseguido em todos os âmbitos da sociedade e induzir o reconhecimento de novos direitos não necessariamente constitucionais. É nessa lógica que se defenderá a existência de um direito à explicação em decisões automatizadas.

Ordenamentos jurídicos possuem determinadas características, e duas das mais relevantes que convêm de serem exploradas para justificar a necessidade de se assegurar um direito à explicação em processos decisórios para preencher um espaço gerado pela disrupção tecnológica, é a de “completude”, e por consequência lógica, a sua falha e ausência, denominada como “lacuna”. Para Norberto Bobbio, podem ser definidas como:

Por “completude” se entende a propriedade pela qual um ordenamento jurídico tem uma norma para regular qualquer caso. Posto que a ausência de uma norma se chama, com frequência, de “lacuna (em um dos sentidos do termo “lacuna”), “completude” significa “ausência de lacunas” (...) podemos dizer que um ordenamento é completo quando nunca se verifica uma situação na qual não se possa demonstrar a pertinência nem de uma determinada norma nem de sua norma contraditória.¹⁷⁵

Por conseguinte, se nenhuma norma existe, tanto em caráter positivo, permitindo determinada ação ou omissão, quanto em caráter negativo, proibindo ou vedando uma conduta, compreende-se um ordenamento incompleto.¹⁷⁶ Para ocupar esse vácuo legislativo, Bobbio prevê duas metodologias instrumentais. A primeira — em uma perspectiva de resgate de fundamentos externos — se refere à heterointegração, a qual busca mecanismos em fontes distintas da lei e em outros ordenamentos. O segundo designa a autointegração — na perspectiva interna, com elementos já presentes e inerentes ao próprio ordenamento jurídico — por meio da analogia e dos princípios gerais do direito.¹⁷⁷

Para os fins do presente capítulo, utilizar-se-á a heterointegração no que tange à análise comparada com o ordenamento jurídico da União Europeia e a influência direta em institutos jurídicos basilares para fundamentar um direito à explicação. Quanto à autointegração, esta será essencial como método no que tange a dissertar como garantir explicações e explicabilidade de

¹⁷⁴ MENDES, Gilmar; BRANCO, Paulo. **Curso de Direito Constitucional**. São Paulo: Saraiva Educação (Série IDP), 2021, p. 171.

¹⁷⁵ BOBBIO, Norberto. **Teoria do Ordenamento Jurídico**. São Paulo: Edipro, 2 ed., 2014, p. 113.

¹⁷⁶ BOBBIO, Norberto. **Teoria do Ordenamento Jurídico**. São Paulo: Edipro, 2 ed., 2014, p. 113.

¹⁷⁷ BOBBIO, Norberto. **Teoria do Ordenamento Jurídico**. São Paulo: Edipro, 2 ed., 2014, pp. 138-141.

sistemas autônomos de decisões efetiva os princípios da autodeterminação informativa, o livre desenvolvimento da personalidade e a dignidade da pessoa humana.

Nessa senda, confronta-se, em lógica semelhante ao ocorrido com o paradigma de proteção de dados, legitimado em um dispositivo legal a partir da promulgação da LGPD, como um direito à explicação pode surgir e ser assegurado na esfera infraconstitucional do ordenamento jurídico brasileiro diante de uma evidente lacuna na automação decisória. Para isso, é necessário compreender como determinados direitos subjetivos podem nascer. O caminho mais óbvio e intuitivo é o direito decorrer de uma previsão estabelecida por lei, mas também é possível construir a defesa da existência de um determinado direito por meio de interpretações teleológicas e sistemáticas de arquiteturas normativas que englobam direitos mais abrangentes, e, por fim, direitos também podem surgir a partir de derivações morais de uma filosofia ética, como ocorreu com a corrente dos direitos naturais.¹⁷⁸

Para os fins desta monografia, especialmente o presente capítulo, considerar-se-ão as duas primeiras formas para destrinchar a dimensão existencial de um direito à explicação em decisões automatizadas. Além disso, citar-se-ão situações fáticas e estatísticas que colocam em xeque princípios de igualdade entre indivíduos, haja vista a discriminação algorítmica e os vieses serem uma realidade no processamento de informações pessoais automatizados. Assim, defender-se-á a explicação como um meio de combater práticas discriminatórias.

2.1. O DIREITO À EXPLICAÇÃO COMO DECORRÊNCIA DO DEVIDO PROCESSO LEGAL E DEVIDO PROCESSO INFORMACIONAL

Em um Estado Democrático de Direito, a garantia do devido processo legal se funda a partir da concepção de que os processos decisórios estatais acerca da vida dos cidadãos somente são legítimos se houver sólida fundamentação e, em outras palavras, se as decisões são pertinentemente motivadas.¹⁷⁹ No ordenamento jurídico brasileiro, trata-se de um direito

¹⁷⁸ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 74.

¹⁷⁹ KAUFMAN, Dora; JUNQUILHO, Tainá; REIS, Priscila. Externalidades negativas da inteligência artificial: conflitos entre limites da técnica e dos direitos humanos. **Revista Direitos e Garantias Fundamentais**, Vitória, v. 24, n. 3, p. 43-71, set./dez. de 2023.

fundamental previsto constitucionalmente na CRFB/88 no art. 5º, inciso LIV.¹⁸⁰ Por conseguinte, a análise do presente capítulo parte do pressuposto que um direito à explicação em decisões automatizadas decorre diretamente desse (e outros) preceitos constitucionais e infraconstitucionais.

Tal entendimento é ensejado no cenário digital brasileiro pela fundamentação utilizada no célebre caso da Ação Direta de Inconstitucionalidade 6.389 pelo Supremo Tribunal Federal — conhecido popularmente como “Caso IBGE” — com a “releitura da cláusula do devido processo em meio a um cenário crescente de automatização de processos de tomada de decisão que afetam as liberdades dos indivíduos”¹⁸¹, o que, no voto do Ministro Gilmar Mendes, deriva-se em um “devido processo informacional”, como pode ser visualizado:

É possível identificar como corolário da dimensão subjetiva do direito à proteção de dados pessoais, a preservação de verdadeiro “devido processo informacional” (*informational due process privacy right*), voltado a conferir ao indivíduo o direito de evitar exposições de seus dados sem possibilidades mínimas de controle, sobretudo em relação a práticas de tratamento de dados capazes de sujeitar o indivíduo a julgamentos preditivos e peremptórios.¹⁸²

O devido processo legal, portanto, transfigura-se em uma nova figura em uma sociedade algorítmica. As decisões automatizadas que interferem na esfera jurídica individual e coletiva de cidadãos não escapam da jurisdição constitucional que funda o *rule of law*. Se não é possível garantir um devido processo informacional na seara digital, abre-se margem para uma nova ordem incapaz de exigir a prestação de contas necessária: todos têm, em um Estado Democrático, direito a entender os fundamentos das decisões que interferem no seu viver.¹⁸³ A distinção entre o preceito tão bem estabelecido doutrinária e jurisdicionalmente do processo legal é que é, desde a sua acepção, é diretamente associado a uma prestação estatal, contudo, o devido processo informacional exige que as regras de motivação e explicação se estendam aos atores privados.

¹⁸⁰ BRASIL. **Constituição (1988)**. Constituição da República Federativa do Brasil. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

¹⁸¹ BIONI, Bruno. MARTINS, Pedro. Devido processo informacional: um salto teórico-dogmático necessário? **Jota**, 15 de jul. de 2015. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/devido-processo-informacional-um-salto-teorico-dogmatico-necessario-15072020>

¹⁸²
¹⁸³ BIONI, Bruno. MARTINS, Pedro. Devido processo informacional: um salto teórico-dogmático necessário? **Jota**, 15 de jul. de 2015. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/devido-processo-informacional-um-salto-teorico-dogmatico-necessario-15072020>

Esse fenômeno à medida que os julgamentos se digitalizam, plataformas consolidam julgamentos dos usuários em níveis nunca antes vistos na história humana, seja por classificações, ranqueamentos, sistemas de avaliação ou por sanções sociais como verdadeiras praças públicas. Em geral, essas práticas são regidas por regras, o que, por consequência, permitem lógicas mercadológicas que perfilam e classificam os cidadãos de forma irrestrita. Com algoritmos secretos dificultando a identificação de ilicitudes, a vulnerabilidade dos cidadãos aumenta. Historicamente, o devido processo legal protegia contra o Estado, mas agora deve ser repensado para enfrentar o poder crescente dos agentes privados nas relações digitais.¹⁸⁴

Em verdade, para Bruno Bioni o paradigma da privacidade é intrinsecamente relacionado aos preceitos constitucionais, mormente, o do devido processo e relembra relevante pronunciamento dos “princípios de prática de informação justa” do Departamento de Estado dos EUA:

Existe uma certa combinação entre temas de privacidade e os demais temas constitucionais. Minha visão pessoal é que provavelmente um dos maiores bastiões constitucionais da privacidade ainda não explorado pelas cortes ou pelos defensores ativistas é o conceito de devido processo legal, a noção de que governos não podem privá-lo de sua vida, liberdade, propriedade, sem o devido processo legal, uma restrição que recai tanto sobre o governo nacional quanto, claro, sobre os estados e governos locais. Me parece que está por ser escrito o capítulo sobre devido processo informacional. E certos tipos de levantamentos, usos e disseminação de informação governamental podem ser desafiados quando violarem o devido processo (...)¹⁸⁵

Desse modo, entende-se o devido processo legal de modo abrangente ao não se restringir à visão restrita processualista, mas sim de compreendê-lo como um instrumento de razoabilidade, proporcionalidade e simetria de poder entre partes. O devido processo informacional surge justamente para controlar o aumento desenfreado de assimetria informacional entre empresas e consumidores de tecnologia.¹⁸⁶

Convém, a partir do exposto, apresentar os fenômenos que ensejam o fortalecimento do devido processo legal no mundo digital na figura do devido processo informacional.

¹⁸⁴ FRAZÃO, Ana. Devido processo digital. **Jota**, 20 de out. de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/devido-processo-digital-20102021>

¹⁸⁵ MILLER, Arthur. Transcription of the 1st Meeting Part I of the Secretary's Advisory Committee on Automated Personal Data Systems of the U.S. Department of Health, Education and Welfare, p. 267.

¹⁸⁶ BIONI, Bruno. MARTINS, Pedro. Devido processo informacional: um salto teórico-dogmático necessário? **Jota**, 15 de jul. de 2015. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/devido-processo-informacional-um-salto-teorico-dogmatico-necessario-15072020>

2.2. O INDIVÍDUO NO CAPITALISMO DE VIGILÂNCIA E NA SOCIEDADE DO ALGORITMO

Com a complexificação de procedimentos tecnológicos e sob o manto da literatura catastrófica da vigilância em massa encontrada no *Grande Irmão* de George Orwell e do *Panopticon* de Jeremy Bentham¹⁸⁷, a privacidade, embora seja uma concepção que remonta aos primórdios das discussões filosóficas da natureza humana, somente se tornou uma temática concreta em tribunais e plenários no século XX.¹⁸⁸

Nessa senda, a tutela da privacidade é uma inquietação da modernidade — como um elemento diretamente associado ao livre desenvolvimento da personalidade — ao ponto de não ser um instituto jurídico suficiente para enfrentar os desafios do avanço tecnológico e, assim, acaba por se transfigurar para dar origem ao novo direito de proteção de dados pessoais.¹⁸⁹ O nascimento do paradigma dos dados surge da compreensão de que não mais bastava o *right to be alone* (direito de ser deixado só) em uma acepção individualista e de isolamento¹⁹⁰. Trata-se do reconhecimento dos ordenamentos jurídicos de que a contemporaneidade se deparou com uma nova era com a informação como fenômeno central, com variadas denominações para esse novo momento histórico que remontam à mesma realidade: “sociedade da informação, ou sociedade do conhecimento, economia da informação ou sociedade em rede”.¹⁹¹

Com efeito, é inegável a propriedade disruptiva das tecnologias como pontos de virada na ordem social, econômica, jurídica e ideológica da sociedade, e, por conseguinte, a ascensão de uma sociedade algorítmica desvia o Poder Público como a única força institucional preocupada com o respeito a um Estado Democrático de Direito que assegure uma rede de proteção às garantias e direitos fundamentais.¹⁹² Agora, quem produz, desenvolve e controla os algoritmos

¹⁸⁷ RODOTÀ, Stefano. **A vida na sociedade da vigilância: a privacidade hoje**. Rio de Janeiro: Renovar, 2008, p. 47.

¹⁸⁸ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, p. 27.

¹⁸⁹ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, pp. 27-31.

¹⁹⁰ WARREN, Samuel; BRANDEIS, Louis. The right to privacy. **Harvard Law Review**, v. IV, n. 5, p. 195, 1890.

¹⁹¹ MENDES, Laura Schertel. **Privacidade, proteção de dados e defesa do consumidor: linhas gerais de um novo direito fundamental**. 1ª Edição. São Paulo: Editora Saraiva, 2014, p. 19.

¹⁹² MICKLITZ, Hans. W et al. **Constitutional Challenges in The Algorithmic Society**. Cambridge, Cambridge University Press, 2022, pp. 1-4.

não somente possui a responsabilidade para com uma sociedade democrática, como ainda deve ser exigido legal e moralmente de atores privados a tutela da ordem constitucional.¹⁹³

Tal ruptura ocorreu devido à descoberta de que dados pessoais são valiosa matéria-prima de um sofisticado processo mercadológico da economia da era do *Big Data*¹⁹⁴, em razão de suas propriedades essenciais para as principais operações das grandes empresas de tecnologia: utilidade, volume e baixo custo. Nossas informações pessoais podem servir para incontáveis e inimagináveis possibilidades, como serem vendidas para fins de publicidade e anúncios, serem cedidas para prever o futuro e comportamentos, e, em casos extremos, serem utilizadas para nos controlar e manipular. Para tal, grandes quantidades são necessárias, porque mais dados significam mais poder em variadas manifestações, como o poder de computação, o poder do conhecimento e o poder do dinheiro. Por fim, à primeira vista, aparentam ter um baixo valor para quem os detém, afinal, não há como comparar as vantagens de receber um desconto ou um emprego em troca de algo supostamente banal como quem você é. Em certas situações, são irrisórios a ponto de não custar supostamente nada, pois são minerados sem a ciência do titular, como subprodutos da relação pessoa-tecnologia.¹⁹⁵

Convém, portanto, realizar uma distinção primordial entre dado e informação, embora sejam utilizados, convenientemente, como sinônimos para o fins desta monografia, assim como as legislações e doutrinas adotam os conceitos de forma igual, embora não sejam propriamente a mesma coisa. Para Danilo Doneda, ambos designam a representação de um aspecto fático da realidade, e, por essa razão, confundem-se, contudo, o dado é a forma mais rudimentar, anterior a um processamento interpretativo e a informação seria a transfiguração do dado em um determinado contexto, processado pela cognição.¹⁹⁶

Entretanto, é um engano supor que nossas informações íntimas não possuem valor. Na era do capitalismo de vigilância, cada ínfimo aspecto de quem somos é utilizado por *Big Techs* a fim de alimentar inteligências de máquina para não somente predizerem comportamentos humanos, mas também para moldá-los em larga escala. Conforme Shoshana Zuboff, “a meta é

¹⁹³ MICKLITZ, Hans. W et al. **Constitutional Challenges in The Algorithmic Society**. Cambridge, Cambridge University Press, 2022, p. 15.

¹⁹⁴ VÉLIZ, Carissa. **Privacidade é Poder: porque e como você deveria retomar o controle de seus dados**. São Paulo: Editora Contracorrente, 2021, p. 21.

¹⁹⁵ VÉLIZ, Carissa. **Privacidade é Poder: porque e como você deveria retomar o controle de seus dados**. São Paulo: Editora Contracorrente, 2021, p.129.

¹⁹⁶ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, p. 139.

nos automatizar”.¹⁹⁷ Para a autora fundadora do termo *capitalismo de vigilância*, ao realizarmos esse *tradeoff* sobre nossa individualidade em prol do prazer de estar conectado e atuantes em plataformas sociais causa um “entorpecimento psíquico”¹⁹⁸ de uma falsa sensação de decisão altamente normalizada, pois o indivíduo passa a ser um prisioneiro, sem liberdade de agir e ser conforme às suas vontades, mas um preso supostamente feliz. E o grande problema está nessas grandes empresas serem entidades oniscientes sobre nós, mas, de forma diametralmente oposta, não sabemos nada sobre como operam os seus algoritmos.¹⁹⁹

Não se trata de um produto, serviço ou tecnologia, mas sim uma nova forma racionalização de ações a serem tomadas em um empreendimento movido por dados. A questão do capitalismo de vigilância ultrapassa as fronteiras da materialidade de *softwares*, algoritmos e máquinas, Essa nova realidade foi fundada pelo Google por meio da vanguarda em estabelecer propaganda direcionada automatizada em mecanismo de busca com os resquícios de dados na interação humano-máquina, e Shoshana compara o pioneirismo da empresa com as inovações de Ford na revolução do industrialismo.²⁰⁰ Esses rastros de informações são escavados por meio de uma técnica denominada *data-mining* (mineração de dados), definida como:

Ela consiste na busca de correlações, recorrências, formas, tendências e padrões significativos a partir de quantidades muito grandes de dados, com o auxílio de instrumentos estatísticos e matemáticos. Assim, a partir de uma grande quantidade de informação em estado bruto e não classificada, torna-se possível identificar informações de potencial interesse.²⁰¹

Os engenheiros do Google já estavam cientes da existência de dados residuais quando usuários realizavam qualquer busca na plataforma, como a forma que escrevemos, o tempo que ficamos em determinadas páginas, os cliques que damos, de onde estamos navegando. Esses pacotes de detalhamento de experiência do indivíduo permitiram a meta fundacional de Larry Page — um dos cofundadores do Google — em torná-lo em um buscador com algoritmos de aprendizado autossuficientes. Tornou-se uma relação de codependência de aprendizado entre plataforma e usuários. Quem acessava o Google, procurava por um conhecimento sobre algo, e o buscador aprendia com as pegadas digitais dos usuários.²⁰²

¹⁹⁷ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Intrínseca, 2021, p. 31.

¹⁹⁸ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Intrínseca, 2021, p. 38.

¹⁹⁹ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Intrínseca, 2021, p. 39.

²⁰⁰ ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Intrínseca, 2021, p. 80.

²⁰¹ DONEDA, Danilo. *Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados*. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, p. 150.

²⁰² ZUBOFF, Shoshana. *A Era do Capitalismo de Vigilância*. Rio de Janeiro: Intrínseca, 2021, pp. 81-85.

Nessa senda, o capitalismo de vigilância ultrapassa os limites da razoabilidade. O recente caso da atriz Scarlett Johansson escancara a teoria de Zuboff sobre o roubo de identidades de indivíduos por parte de grandes empresas com o uso de tecnologias de automação.²⁰³ Ocorre que a estrela de Hollywood contestou o uso de uma voz no ChatGPT curiosamente similar à sua, especialmente devido ao filme "Her" (2013), onde ela dubla um sistema operacional de IA. A atriz afirmou que, após recusar a oferta de Sam Altman, CEO da OpenAI, para usar sua voz, a empresa lançou a nova voz "Sky", que muitos conhecidos da atriz e as redes sociais compararam à dela. A OpenAI respondeu pausando o uso da voz "Sky" e afirmando que ela foi baseada em outra atriz, não em Johansson, pedindo desculpas pela má comunicação e garantindo que não houve intenção de imitar a voz da atriz.²⁰⁴

O poder na era digital, portanto, emana do controle dos dados pessoais, porque, em última instância, isso significa o poder de influenciar em ampla escala. Influenciar no sentido de modular — na acepção deleuzeana como base de uma sociedade de controle — de uma entidade na posição de autoridade ao “introjetar comportamento dentro de cada pessoa”²⁰⁵. Desse modo, a modulação algorítmica — uma forma bastante específica de modulação deleuzeana — se funda no capitalismo de vigilância por exercer o poder de “cristalizar uma determinada subjetividade na memória, no cérebro das pessoas”²⁰⁶ em uma sociedade de controle algorítmica, sendo sua característica primordial a sutileza de forjar vontades à distância.

A modulação do comportamento humano torna-se um ativo econômico de base para Big Techs que possuem o poder dos modelos de aprendizado de máquina com a maior acurácia. Com a recomendação de anúncios personalizados e de perfis para serem seguidos e a sugestão de

²⁰³ FRAZÃO, Ana. O ChatGPT e voz de Scarlett Johansson. **Jota**, 30 de mai. de 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/o-chatgpt-e-a-voz-de-scarlett-johansson-29052024?non-beta=1>. Acesso em: 30 de mai. de 2024.

²⁰⁴ Bloomerang. Scarlett Johansson contrata advogados para contestar voz 'estranhamente semelhante' à dela no ChatGPT. **O Globo**, 20 de mai. de 2024. Disponível em: <https://oglobo.globo.com/economia/tecnologia/noticia/2024/05/20/scarlett-johansson-contrata-advogados-para-conte-star-voz-estranhamente-semelhante-a-dela-no-chatgpt.ghtml>

²⁰⁵ CASSINO, João Francisco. Modulação deleuzeana, modulação algorítmica e manipulação midiática. In: AVELINO, Rodolfo; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu da. **A Sociedade do Controle: manipulação e modulação nas redes digitais**. São Paulo: Hedra, 2018, pp. 14-15.

²⁰⁶ CASSINO, João Francisco. Modulação deleuzeana, modulação algorítmica e manipulação midiática. In: AVELINO, Rodolfo; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu da. **A Sociedade do Controle: manipulação e modulação nas redes digitais**. São Paulo: Hedra, 2018, p. 15.

conteúdos em *timelines* (linhas de tempo de redes sociais), cada uma destas microdecisões tomadas por algoritmos de inteligência artificial exerce o poder modular sob os usuários.²⁰⁷

A situação se exacerba nos riscos quando se trata de grandes bancos de dados em execução da técnica de *profiling* — manipulação de dados pessoais por meio de métodos estatísticos e técnicas de inteligência artificial almeja a obtenção de uma “metainformação”²⁰⁸ — como nossos padrões de consumo, navegação, curtidas, pesquisas e tempo gasto em determinadas páginas nas plataformas digitais. Tal prerrogativa pode ser aplicada a sujeitos individuais, bem como estendida a coletividades. O resultado pode ser utilizado para delinear um quadro das tendências de decisões futuras, comportamentos e destino de uma pessoa ou grupo. Assim, as técnicas de previsão de padrões de comportamento podem ocasionar uma diminuição da esfera de liberdade do indivíduo, uma vez que o perfil eletrônico se torna a única faceta da personalidade visível a terceiros.²⁰⁹

Sob o prisma da modulação algorítmica e a exemplo da técnica apresentada, cita-se o histórico e marcante caso da *Cambridge Analytic*, em que a democracia vista sob a égide digital, comportou-se de modo inesperado e paradigmático, sobretudo. Em 2018, fora revelado o escândalo nas eleições presidenciais estadunidenses, ocorridas em 2016, com a exposição da enorme utilização de dados pessoais para manipulação eleitoral nesse processo e na escolha popular da permanência ou não do Reino Unido na União Europeia, popularmente denominado como *Brexit*.²¹⁰

No caso em tela, um pesquisador da empresa Cambridge publicou na plataforma do Facebook questionários voltados para a pesquisa pessoal de personalidade, com suposto intuito acadêmico, oferecendo a quem o respondesse à quantia de cinco dólares. Com o nome de “*This Is Your Digital Life*”, o aplicativo coletou dados não apenas dos usuários que o instalaram, mas também dos amigos de rede desses usuários, resultando em uma ampla coleta de informações.

²⁰⁷ OLIVEIRA, Carla. Aprendizado de máquina e modulação do comportamento humano. In: AVELINO, Rodolfo; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu da. **A Sociedade do Controle: manipulação e modulação nas redes digitais**. São Paulo: Hedra, 2018, p. 79-89.

²⁰⁸ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, p. 149.

²⁰⁹ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, p. 149.

²¹⁰ Cf. CONFESSORE, Nicholas. **Cambridge Analytica and Facebook: The Scandal and the Fallout So Far**. New York Times. New York, 4 abr. 2018. Disponível em: <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>_Acesso em: 25 de mar. de 2024.

Com o grande número de respostas, a pesquisa e seu processo massivo de coleta de dados teve resultado crucial nas votações mencionadas. Isso explica-se com que a coleta de dados que possibilitou o direcionamento (*micro-targeting*) e a formulação de perfis (*profilings*), além do afunilamento e entrega persuasiva de tais conteúdos para usuários considerados mais vulneráveis e suscetíveis ao convencimento por intermédio digital.²¹¹

Por conseguinte, a empresa teve acesso a dados de 87 milhões de usuários do Facebook, indevidamente, sem que muitos deles soubessem e sem que a própria rede social tivesse lhes dado permissão. O escândalo veio a público em março de 2018, quando o denunciante Christopher Wylie, ex-funcionário da *Cambridge Analytica*, revelou detalhes sobre a coleta e uso indevido de dados. Suas revelações foram publicadas em reportagens do jornal britânico *The Guardian* e do jornal americano *The New York Times*.²¹²

De modo similar, o cenário brasileiro não escapou do fenômeno da modulação deleuzeana no contexto do capitalismo de vigilância. A Lei das Eleições – Lei nº 9.504/97 – prevê a possibilidade dos partidos políticos realizarem propagandas eleitorais para promover suas campanhas e seus candidatos²¹³. Historicamente, é reconhecida a influência da mídia e das TIC's (Tecnologias de Comunicação) para a visão que a própria população tem sobre a sociedade, contexto que se acentua no processo eleitoral. A televisão, por exemplo, era considerada um vetor fundamental para o cenário político das eleições brasileiras, principalmente por ser o meio de comunicação que transmite o horário de propagando gratuita eleitoral²¹⁴.

Entretanto, a eleição presidencial de 2018 trouxe um cenário disruptivo. O processo eleitoral foi marcado pelo forte engajamento da população através das plataformas digitais. Não somente diversos perfis foram criados para circular informações sobre os candidatos políticos (com as suspeitas e posteriores confirmações de que muitos eram contas “robô”, conhecidos

²¹¹ WYLIE, Christopher. **Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.** The Guardian, London, 17 mar. 2018. Disponível em: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Acesso em: 25 de mar. de 2024.

²¹² WYLIE, Christopher. **Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach.** The Guardian, London, 17 mar. 2018. Disponível em: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Acesso em: 25 de mar. de 2024.

²¹³ BRASIL. **Lei nº 9.504, de 30 de setembro de 1997.** Estabelece normas para as eleições. [S. l.], 30 set. 1997. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/19504.htm. Acesso em: 25 de mar. de 2024.

²¹⁴ SILVA, Elisa de Macedo da. **Redes Sociais, Eleição Presidencial de 2018 e Questões de Credibilidade da Mídia.** 2019. Monografia (Pós Graduação em Comunicação) - Mestrado Acadêmico, [S. l.], 2019. Disponível em: <https://repositorio.ufjf.br/jspui/bitstream/ufjf/12040/1/elisademacedodasilva.pdf>. Acesso em: 25 de mar. de 2024.

como *bots*), como a possibilidade de se comunicar diretamente com o próprio candidato pelas redes sociais foi um fator antes não tão profundamente explorado. A ruptura da tradição eleitoral é contribuída por outro fator, que é a modificação do comportamento dos usuários através do "poder instrumentário", conforme definido por Zuboff.²¹⁵ Em uma célebre dissertação de mestrado pela Universidade Federal de Minas Gerais (UFMG), Ana Luiza Marques disserta:

Nesse ecossistema, estão inseridos “fazendas de trolls” (troll farms), bots (robôs ou contas automatizadas que buscam mimetizar o comportamento humano), ciborgues (mesclam automatização com a curadoria humana), e contas falsas humanas, utilizados com o objetivo de “amplificar narrativas ou produzir discordância”.²¹⁶

A presença das redes sociais para o processo eleitoral de 2018 foi tão expressiva que, embora não tenha aparecido nos principais canais midiáticos para promover sua campanha durante o primeiro e segundo turno, o candidato Jair Bolsonaro venceu as eleições. Sem debater suas propostas e ideias com os seus adversários em rede televisiva nacional, como costumava ser o método mais eficiente de promoção política até a eleição anterior. Nesse contexto, projetos de pesquisa de comunicação questionaram se essa votação marcou a ascensão das mídias sociais como o meio midiático primordial para as campanhas políticas, em contrapartida, com o fim da influência da televisão no processo eleitoral como o mecanismo de divulgação com maior disputa.²¹⁷

Todavia, a utilização das redes sociais como principal meio de propaganda político-eleitoral durante esse momento envolveu a ocorrência de diversos tratamentos inadequados com informações pessoais. Notadamente, foi reconhecido que o número de celular de usuários do Facebook foi utilizado para o envio de mensagens políticas em massa no WhatsApp – o principal aplicativo de mensagem utilizado por brasileiros²¹⁸. Além disso, vale ressaltar a participação das corretoras de dados (*data brokers*), uma indústria destinada a coletar dados pessoais e compartilhá-los entre diversas empresas. Durante a Comissão Parlamentar Mista de Inquéritos da Fake News, o ex-funcionário da empresa de marketing digital Yacows Hans

²¹⁵ ZUBOFF, Shoshana. **A Era do Capitalismo de Vigilância**. Rio de Janeiro: Intrínseca, p. 352, 2021.

²¹⁶ MARQUES, Ana Luiza Pinto Coelho. **Erosão constitucional no Brasil e tecnologias disruptivas: análise a partir do processo eleitoral de 2018**. Dissertação (Pós Graduação em Direito) - Mestrado Acadêmico, Belo Horizonte, 2022, p. 86.

²¹⁷ MARQUES, Ana Luiza Pinto Coelho. **Erosão constitucional no Brasil e tecnologias disruptivas: análise a partir do processo eleitoral de 2018**. Dissertação (Pós Graduação em Direito) - Mestrado Acadêmico, Belo Horizonte, 2022, p. 86.

²¹⁸ MAGENTA, Matheus et al. Como telefones de usuários do Facebook foram usados por campanhas em 'disparos em massa' no WhatsApp. **BBC News Brasil**, 20 out. 2018. Disponível em: <https://www.bbc.com/portuguese/brasil-45910249>. Acesso em: 26 jun. 2023.

River afirmou que a empresa realizou comunicações eleitorais com o uso de dados pessoais não autorizados.²¹⁹

Nesse cenário, um direito à explicação emerge como uma forma de se opor ao regime do capitalismo de vigilância. Receber explicações de decisões, previsões e recomendações automatizadas e ter acesso à explicabilidade de sistemas de inteligência artificial permite que o indivíduo se emancipe diante de uma sociedade feita para controlá-lo e moldá-lo. Saber os motivos que ensejaram. Em suma, é função do Direito na sua acepção de produção de leis e normativos e construção doutrinária e dogmática acompanhar as lacunas cavadas pela constante erosão de garantias fundamentais pela tecnologia em desenvolvimento irrestrito.²²⁰

2.3. O DIREITO À EXPLICAÇÃO COMO EFETIVAÇÃO DO LIVRE DESENVOLVIMENTO DA PERSONALIDADE E DA AUTODETERMINAÇÃO INFORMATIVA

A Segunda Grande Guerra foi um terrível marco na história humana em termos de humanidade. A fim de iniciar uma nova era, organizações internacionais e ordenamentos jurídicos nacionais se uniram em prol de uma nova ordem jurídica em que a pessoa humana, em todas as suas expressões, seja o centro da esfera protetiva do Estado e de entes privados por meio de declarações inegociáveis de direitos humanos. Para tal, urgiu o reconhecimento de que o ser humano é dotado de personalidade, e, por conseguinte, consagra-se o princípio da dignidade da pessoa humana.²²¹ No entanto, ambos os institutos não se sustentam sem que o outro esteja assegurado, pois a dignidade de um indivíduo somente é completa, ampla e concreta quando se garante que a sua personalidade se desenvolva livremente sem a intervenção de um agente externo a manipular, modelar e influir sobre quem uma determinada pessoa deve ser.²²²

Tal acepção se contrapõe diretamente com a nova realidade da contemporaneidade no capitalismo de vigilância e na sociedade algorítmica, tornando necessário garantir, portanto,

²¹⁹ **CPMI das Fake News: depoente diz que houve uso não autorizado de dados pessoais nas eleições de 2018.** [S. L.], 11 fev. 2020. Disponível em: <https://www.camara.leg.br/noticias/636315-cpmi-das-fake-news-depoente-diz-que-houve-uso-nao-autorizado-de-dados-pessoais-nas-eleicoes-de-2018/>. Acesso em: 26 jun. 2023.

²²⁰ MENDES, Laura Schertel. **Privacidade, proteção de dados e defesa do consumidor: linhas gerais de um novo direito fundamental.** 1ª Edição. São Paulo: Editora Saraiva, 2014, p. 19.

²²¹ MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013, pp. 1176-11211.

²²² MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013, pp. 1176-11211.

mecanismos de oposição às formas de modulação e intervenção nas vontades inerentes à personalidade. Em um mundo onde quem somos é ditado por grandes empresas de tecnologia e as suas poderosas máquinas, entender como somos influenciados e manipulados ocupa uma lacuna gerada em um ordenamento jurídico, conforme pensamento de Bobbio. O direito à explicação é uma forma dentre variadas de ocupar esse espaço.

Entender, portanto, o livre desenvolvimento da personalidade, é compreender o que são os direitos de personalidade e como são assegurados no ordenamento jurídico brasileiro e europeu.²²³ Nesse contexto, a personalidade é tutelada pelo direito civil sob a ótica da existência. A condição de aplicação de proteção é ser um humano, em uma ambivalência de sujeito e objeto de direitos: a pessoa humana, englobada tanto por seus elementos corpóreos quanto imateriais.²²⁴ Desse modo, permite exigência ativa por parte do titular de defender todos os bens associados à sua personalidade, de natureza espiritual e física²²⁵. Decorrem, portanto, do princípio da dignidade da pessoa humana como elemento constitucional a partir de uma visão antropocêntrica, pois o ser humano é um fim em si mesmo²²⁶, assim, são, em última instância, “essenciais, inatos e permanentes”²²⁷ e estão diretamente associados a uma noção de liberdade inerente à pessoa natural:

A própria noção de personalidade já enseja, necessariamente, uma conotação de liberdade. Portanto, a criação de um caráter, ou de um indivíduo, com ideologias, jeito, modo de pensar e de agir, essência, referência, identidade, e outras características inerentes à pessoa, realizadas de forma artificial e sem desenvolvimento livre, não poderia gerar uma noção de personalidade.²²⁸

A história desses direitos personalíssimos não segue uma linha reta, com origens nebulosas e longos períodos de intervalos de interrupção na construção de uma lógica jurídica com a personalidade como característica inata da pessoa, a exemplo do período de abstração no positivismo e outras escolas preocupadas demasiadamente com aspectos patrimoniais.²²⁹ Bruno Bioni articula a problemática de que, embora os direitos da personalidade emanem de um

²²³ MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013, pp. 1176-11211.

²²⁴ OLIVEIRA, Carlos Elias; NETO, João Costa. **Direito Civil**. Rio de Janeiro: Forense, 2022, p. 162.

²²⁵ SOUSA, Rabindranath de. **O Direito Geral de Personalidade**. Coimbra Editora, 1995, p. 117.

²²⁶ AMARAL, Francisco. **Direito Civil: Introdução**. São Paulo: Saraiva Educação, 2018, pp. 353-357.

²²⁷ AMARAL, Francisco. **Direito Civil: Introdução**. São Paulo: Saraiva Educação, 2018, p. 355.

²²⁸ MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013, pp. 1176-11211.

²²⁹ BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, pp. 93-95.

parâmetro constitucional, no âmbito infraconstitucional, o Código Civil não estabeleceu um rol taxativo para essa gama de garantias. Em virtude disso, sua noção jurídica vincula-se a duas concepções doutrinárias. A primeira se refere ao direito de personalidade como um prolongamento do indivíduo. Em consonância, a segunda diz respeito “à percepção de que o ser humano é, por excelência, um ser social, devendo-se, portanto, assegurar sua esfera relacional”.²³⁰

No Brasil, o direito ao livre desenvolvimento da personalidade é um princípio fundamental que, embora não explicitamente mencionado na Constituição Federal de 1988, está implicitamente incorporado em nosso sistema jurídico. Este princípio é essencial para a proteção da intimidade e para o reconhecimento e preservação da autonomia privada dos indivíduos,²³¹ como pode ser visualizado:

Ao tutelar um desenvolvimento da personalidade, consagra-se um direito de liberdade individual em relação à constituição da personalidade, integrando um “direito à diferença”, dizendo-se que o problema, no fundo, é permitir a cada um que eleja o seu modo de vida, desde que não cause prejuízo a terceiros. Assim se garante a autonomia de constituir uma personalidade livre, sem qualquer imposição de outrem, preconizando um direito à individualidade.²³²

O desenvolvimento da personalidade emerge, portanto, como um direito à individualidade de se distinguir essencialmente do próximo, desde que não o afete de nenhuma maneira, entretanto, não se restringe a essa mera concepção de liberdade individual autônoma. Trata-se, em verdade, de um direito cujas manifestações exigem uma dimensão negativa (omissão) de não interferência estatal e uma dimensão positiva de prestação jurisdicional do Poder Público (ação), assegurando-lhe formas de exercer livremente seus aspectos personalíssimos primordiais,²³³ como é o caso de prever na esfera constitucional e infraconstitucional determinados direitos derivados.

Com efeito, à medida que o Regulamento Geral sobre a Proteção de Dados da União Europeia não cita expressamente o desenvolvimento da personalidade como um de seus fundamentos no texto legislativo nem no corpo dos conhecidos “considerandos”, a Lei Geral de

²³⁰ BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, p. 98.

²³¹ LUDWIG, Marcos de Campos. O direito ao livre desenvolvimento da personalidade na Alemanha e possibilidades de sua aplicação no Direito privado brasileiro. **Revista da Faculdade de Direito da UFRGS**, v. 19, mar. de 2001, p. 261.

²³² MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013, pp. 1176-11211.

²³³ MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013, pp. 1176-11211.

Proteção de Dados Pessoais buscou, logo no início de sua redação legislativa, trazer o livre desenvolvimento como base norteadora para todo e qualquer tratamento de dados pessoais:

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.²³⁴

Além disso, no rol de fundamentos da lei, também é, novamente, abordado tal instituto em consonância com a dignidade da pessoa humana e os direitos humanos, como uma tríade de escopo de proteção do indivíduo:

Art. 2º A disciplina da proteção de dados pessoais tem como fundamentos:
VII - os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais.²³⁵

Sob a ótica das propostas de legislar e regular sistemas de inteligência artificial, também é possível encontrar o livre desenvolvimento da personalidade, juntamente com outros princípios associados no caso do PL 2338/23, embora não se possa realizar tal afirmação em relação às outras propostas brasileiras como o PL 21/2020 e PL 210/2024 nem no AI Act da União Europeia:

Art. 2º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial no Brasil têm como fundamentos:
I – a centralidade da pessoa humana;
II – o respeito aos direitos humanos e aos valores democráticos;
III – o livre desenvolvimento da personalidade.²³⁶

Assim, os direitos da personalidade e o livre desenvolvimento da personalidade, em virtude de sua natureza subjetiva, resguardam a vida humana em sua evolução, bem como na progressão tecnológica e científica. Ao se proceder à análise do direito à proteção de dados, entre outros projetos de lei que visam regulamentar a inteligência artificial, almeja-se garantir a eficácia e a salvaguarda dos direitos personalíssimos. É imperativo não perder de vista que a vida humana constitui o epicentro do direito, sendo a tecnologia e a ciência instrumentos para o seu desenvolvimento.

²³⁴ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

²³⁵ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

²³⁶ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

Nesse contexto, convém a análise do surgimento de outro instituto elementar para a compreensão de uma doutrina de proteção de dados. A autodeterminação informativa se consagra como um direito na paradigmática decisão do Tribunal Constitucional Alemão em 1983, mas os caminhos fundacionais na dogmática da Alemanha percorrem outros institutos jurídicos – como o direito geral de personalidade, a liberdade de ação e o direito ao respeito à esfera privada –²³⁷, sendo definida por um dos doutrinadores pioneiros da disciplina de proteção de dados no Brasil, Danilo Doneda:

Concebido como um direito fundamental, na esteira do direito geral de personalidade, o direito à autodeterminação informativa proporciona ao indivíduo o controle sobre as suas informações. Na tradição democrática alemã, este direito fundamental é entendido como uma afirmação do personalíssimo (...)²³⁸

O direito de autodeterminação decorre, no ordenamento jurídico alemão, de uma decisão que previa um “direito à autodeterminação do cidadão”, em um contexto de crescente e descontrolado processamento em massa de informações, e, assim, as discussões do caso do recenseamento no Tribunal Constitucional Alemão se centram no objetivo de não se perder de vista os riscos dos efeitos da automação de processamento de dados.²³⁹ A decisão se referia à Lei do Censo Alemão (Volkzählungsgesetz) de 1983. A referida norma estipulou que os cidadãos deveriam fornecer uma gama de dados pessoais para a mensuração estatística da distribuição espacial e geográfica da população e, como grande problemática, “previa a possibilidade de que os dados coletados fossem cruzados com outros registros públicos para a finalidade genérica de execução de “atividades administrativas”²⁴⁰

A autodeterminação informativa, por conseguinte, enseja um domínio do cidadão sobre as informações pessoais processadas por entes públicos e privados. Em outros termos, a faculdade do indivíduo de exercer autodeterminação sobre seus dados pessoais constitui elemento essencial de seu direito ao livre desenvolvimento de sua personalidade.²⁴¹

²³⁷ MENDES, Laura Schertel Ferreira. Autodeterminação informativa: a história de um conceito. **Pensar**, Fortaleza, v. 25, n.4, p. 1-18, out./dez. 2020.

²³⁸ DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020, pp. 160-161.

²³⁹ MENDES, Laura Schertel Ferreira. Autodeterminação informativa: a história de um conceito. **Pensar**, Fortaleza, v. 25, n.4, p. 1-18, out./dez. 2020.

²⁴⁰ BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, p. 128.

²⁴¹ BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, p. 129.

Em decisão histórica, a Corte Interamericana de Direitos Humanos, no Caso *Membros da Corporação Coletivo de Advogados “José Alvear Restrepo” vs. Colômbia*, reconheceu que a nação havia violado uma série de direitos, incluído, no extenso rol, o direito à autodeterminação informativa:

A Corte destacou que o conteúdo da Convenção Americana sobre Direitos Humanos deriva o direito autônomo à autodeterminação informativa, que inclui o direito de acessar e controlar os dados pessoais que constem em arquivos públicos. Nesse sentido, conforme indicado na Sentença, a atuação estatal configurou a violação a esse direito.²⁴²

Em lógica similar ao livre desenvolvimento da personalidade, o RGPD não estabeleceu uma menção direta ao instituto da autodeterminação informativa na estrutura legislativa como fundamento ou princípio a orientar o tratamento de dados pessoais, contudo, de forma diametralmente oposta, a LGPD prevê expressamente:

Art. 2º A disciplina da proteção de dados pessoais tem como fundamentos:
II - a autodeterminação informativa;²⁴³

No corpo legislativo dos PLs de regulamentação de IA no Brasil, não se encontra a previsão da autodeterminação informacional no PL 21/2020 nem no PL 210/2024, entretanto, prevê-se o instituto no PL 2338/23:

Art. 2º O desenvolvimento, a implementação e o uso de sistemas de inteligência artificial no Brasil têm como fundamentos:
VIII – a privacidade, a proteção de dados e a autodeterminação informativa;²⁴⁴

Em síntese, a relevância das explicações em decisões automatizadas se interliga indissociavelmente ao princípio da autodeterminação informativa e ao livre desenvolvimento da personalidade como meios do indivíduo controlar as suas informações pessoais e ter a liberdade individual de ser quem é, sem a imposição de vontades externas de *Big Techs*, startups e a Administração Pública por meio de modelos de inteligência artificial.

²⁴² CORTE INTERAMERICANA DE REECHOS HUMANOS. **Comunicado: A Colômbia é responsável internacionalmente por haver realizado atividades arbitrárias de inteligência contra defensores de direitos humanos, que também foram vítimas de atos de violência e estigmatização por parte das autoridades estatais.** 2024. Disponível em https://www.corteidh.or.cr/docs/comunicados/cp_16_2024_port.pdf

²⁴³ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

²⁴⁴ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023.** Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

2.4. O DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS COMO MECANISMO DE COMBATE À DISCRIMINAÇÃO ALGORÍTIMICA

Diante do exposto, convém questionar para os objetivos aqui propostos: todos os indivíduos e coletividades são alvos de erros em decisões automatizadas da mesma maneira? A automação de processos decisórios interfere igualmente em todos? O impacto sobre garantias fundamentais ocorre igualmente para todos os titulares de dados? O objetivo a seguir, por conseguinte, será explicitar que a resposta para essas perguntas é um enfático “não” e, por conseguinte, o direito às explicações se torna ainda mais essencial de ser assegurado quando se trata de públicos minoritários e historicamente desfavorecidos. Portanto, a análise deste tópico será orientada pela premissa máxima da igualdade humana como “como uma busca pela equiparação das oportunidades, da ocupação dos espaços políticos e sociais de forma justa (...)”²⁴⁵

Algoritmos não são neutros. A utopia da neutralidade pode ter prevalecido por um período de tempo, contudo, a premissa cai por terra quando se analisam os modelos estatísticos — mormente aqueles denominados como *algoritmos de destruição em massa (ADMs)*.²⁴⁶ Cathy O’Neil, em célebre obra, desmistificou tal pressuposto utilizado como escudo para preconceitos por matemáticos que escreviam fórmulas que ditavam massivamente a vida das pessoas. Conforme a autora, camuflam-se lógicas tendenciosas e prejudiciais a determinados grupos de indivíduos em algoritmos e perpetuam assimetrias de poder, privilegiando mais ainda os privilegiados e excluindo os mais pobres.²⁴⁷

Nessa senda, avalia-se o caso do INSS no Brasil. Às 9h58min, um trabalhador brasileiro de 53 anos solicita, com documentos comprobatórios anexados, o seu direito constitucional à aposentadoria perante o sistema online do Instituto Nacional do Seguro Social (INSS). Às 10h04min, recebe a resposta: o pedido havia sido negado. O tempo transcorrido da solicitação até o momento da negativa fora de exatos 6 (seis) minutos²⁴⁸. A decisão foi tomada pelo Isaac, a

²⁴⁵ LINDOSO, Maria Cristine Branco. **Discriminação de gênero em processos decisórios automatizados**. Dissertação (Mestrado - Mestrado em Direito) - Universidade de Brasília, 2019, p 49.

²⁴⁶ O’NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. São Paulo: Editora Rua do Sabão, 2020.

²⁴⁷ O’NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. São Paulo: Editora Rua do Sabão, 2020, pp. 14-15.

²⁴⁸ GERCINA, Cristiane. INSS aumenta análise de aposentadorias por robôs e nega benefício em seis minutos. **Folha de São Paulo**, São Paulo, 31 de jul. de 2023. Disponível em: <https://www1.folha.uol.com.br/mercado/2023/07/inss-aumenta-analise-de-aposentadorias-por-robos-e-nega-beneficio-em-seis-minutos.shtml>

inteligência artificial (IA) do INSS responsável pelas requisições de aposentadoria, que, embora possua nome de humano, é uma ferramenta algorítmica desenvolvida para agilizar a extensa e crônica fila dos benefícios de seguridade social no Brasil, mas não é a única. O uso da tecnologia de IA no INSS iniciou há anos, e, recentemente, entre 2022 e 2023, a análise automatizada subiu de 17% para 36%, sendo a meta ser pelo menos 50% das decisões serem realizadas automaticamente até o ano de 2026.²⁴⁹

Em resposta, o INSS justificou que o pedido do trabalhador havia sido indeferido devido a um erro no momento do preenchimento da solicitação, e, assim, a IA, ao reconhecer uma resposta incompatível, negou com uma velocidade surpreendente. Em suma, o simples ato de preencher um campo e uma caixa de seleção incorretamente ou clicar no lugar errado acidentalmente basta para que um pedido sequer seja analisado junto aos documentos anexados. No caso em tela, o frentista havia comprovado o direito com os registros exigidos por lei, contudo, por descuido, selecionou a categoria incorreta de tempo de aposentadoria.²⁵⁰ Para conseguir obter o seu direito, são limitadas as opções: recorrer administrativamente da decisão contrária, o que o insere em uma fila de prazo médio de resposta de mais de 1000 (dias) conforme dados do Tribunal de Contas da União (TCU)²⁵¹ ou esperar, no mínimo, 30 (trinta) dias para realizar uma nova requisição que ainda será novamente avaliada. Por fim, pode, em um cenário extremo, ingressar com uma ação judicial. Em todos os casos, trata-se de um impacto considerável na vida do trabalhador que esperará, no mínimo, mais de um mês.

Trata-se de uma decisão automatizada individual, mas não é um caso único nem excepcional, pois uma coletividade de beneficiários com variados pedidos indeferidos sofrem da mesma problemática. Em 2022, houve um acréscimo de 325,7% nos casos de indeferimentos automáticos como aponta o relatório elaborado pela Controladoria Geral da União (CGU)²⁵² e,

²⁴⁹ BRANCO, Ana Paula. Entenda a nova inteligência artificial do INSS para detectar fraude em atestados médicos. **Folha de São Paulo**, São Paulo, 15 de jan. de 2024. Disponível em: <https://www1.folha.uol.com.br/mercado/2024/01/entenda-a-nova-inteligencia-artificial-do-inss-para-detectar-fraude-em-atestados-medicos.shtml?origin>

²⁵⁰ GERCINA, Cristiane. INSS aumenta análise de aposentadorias por robôs e nega benefício em seis minutos. **Folha de São Paulo**, São Paulo, 31 de jul. de 2023. Disponível em: <https://www1.folha.uol.com.br/mercado/2023/07/inss-aumenta-analise-de-aposentadorias-por-robos-e-nega-beneficio-em-seis-minutos.shtml>

²⁵¹ GERCINA, Cristiane. Órgão que julga recursos de aposentadoria levará 16 anos para zerar fila, estima TCU. **Folha de São Paulo**, São Paulo, 30 de mar. de 2023. Disponível em: <https://www1.folha.uol.com.br/mercado/2023/03/orgao-que-julga-recursos-de-aposentadorias-levara-16-anos-para-zerar-fila-estima-tcu.shtml>

²⁵² BRASIL, Controladoria Geral da União. **Relatório de Avaliação do Instituto Nacional do Seguro Social: exercícios 2021 a 2023**. Brasília, 2023, p. 14.

por consequência, a fila de recursos administrativos em análise aumentou em 32%. Isso implica, inevitavelmente, que pessoas com dificuldades com aparatos tecnológicos (como celulares, aplicativos e sites) por motivos de ausência de letramento digital, idade, deficiências cognitivas e físicas (como questões de visão) podem vir a ter direitos constitucionalmente assegurados negados por suas características individuais dificultarem o acesso por meios totalmente automatizados, em um claro processo discriminatório.

Outras plataformas também possuem *ADMs* com lógicas discriminatórias ilícitas. No ano de 2015, o Google Fotos, aplicativo de armazenamento em nuvem de imagens categorizou uma foto de um casal de origem étnica do Senegal como “gorilas”²⁵³. Analogamente, em 2018, a Amazon foi acusada de utilizar um modelo de IA que discriminava candidatas mulheres para vagas de emprego.²⁵⁴ Já em 2020, o Tiktok escondeu, propositadamente, vídeos produzidos por minorias sociais com estereótipos de pobreza, como casas mal decoradas, reboco aparente, além de listar conteúdos que remetem a miséria e “favela” para serem menos recomendados aos usuários.²⁵⁵ Por fim, em 2021, o aprendizado de máquina de recomendações do Facebook classificou um vídeo com homens negros — sem nenhuma aparição ou forma semelhante a um animal presente nas imagens — como “conteúdo de primatas”²⁵⁶.

De maneira mais objetiva, grupos minoritários, no uso de tecnologias de reconhecimento facial treinadas por aprendizado de máquina são grandes alvos de discriminações algorítmicas. A exemplo, as pessoas transexuais se deparam com grandes barreiras de inclusão no que tange ao momento de realizarem o reconhecimento de suas faces em aplicativos, principalmente de bancos. No ano de 2019, uma investigação da Universidade do Colorado evidenciou que as tecnologias de análise facial podem falhar na determinação do gênero de indivíduos trans em até 40% das ocasiões. Esses experimentos foram conduzidos com 2.450 imagens obtidas no Instagram. A biometria facial, conforme as provedoras da solução, alcançaria uma precisão

²⁵³ KASPERKEVIC, Jana. Google says sorry for racist auto-tag in photo app. **The Guardian**. New York. 1 de julho de 2015. Disponível em: <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>. Acesso em: 21 de mai. de 2024.

²⁵⁴ BBC NEWS. Amazon scrapped “sexist AI” tool. 10 de out. de 2018. **BCC NEWS**, Disponível em: <https://www.bbc.com/news/technology-45809919>

²⁵⁵ BIDDLE, Sam; RIBEIRO, Paulo Vítor; DIAS, Tatiana. CENSURA INVISÍVEL: Tik Tok escondeu ‘feios’ e favelas para atrair novos usuários e censurou posts políticos. **The Intercept**, 16 de mar. de 2020. <https://www.intercept.com.br/2020/03/16/tiktok-censurou-rostos-feios-e-favelas-para-atrair-novos-usuarios/>

²⁵⁶ MAC, Ryan. Facebook apologizes after A.I. puts ‘primates’ label on videos of black men. **The New York Times**, 3 de set. de 2021. Disponível em: <https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html>. Acesso em: 21 de mai. de 2024.

superior a 95%. Todavia, para pessoas transgênero, essa taxa de acurácia reduz-se para menos de 60%. Isso decorre do fato de que o algoritmo de identificação baseia-se em medidas faciais, tais como a distância entre os olhos e entre o nariz e o lábio superior, o que pode resultar em equívocos na determinação do gênero do indivíduo identificado.²⁵⁷

Similarmente, pessoas negras também sofrem com o uso de reconhecimento facial, mormente nas aplicações para fins de segurança pública e investigações policiais. Nesse contexto, entre março e outubro de 2019, a Rede de Observatórios de Segurança acompanhou as experiências iniciais de cinco estados brasileiros com tecnologias de reconhecimento facial aplicadas em câmeras de videomonitoramento em todo o seu território (Bahia, Rio de Janeiro, Santa Catarina, Paraíba e Ceará.) A entidade constatou que 90,5% dos indivíduos detidos por monitoramento facial no Brasil eram negros.²⁵⁸

Esses casos e as estatísticas apresentadas revelam que não se tratam de meros casos isolados. A discriminação algorítmica é uma realidade crescente e nociva na era do capitalismo de vigilância e na sociedade do algoritmo. No entanto, há confusões quanto ao termo. Discriminar é um verbo, no senso comum, diretamente associado a tratar de forma diferenciada, mas ambos são conceitos diferenciados e não entendidos como sinônimos pela academia e por legislações, a exemplo da Lei Geral de Proteção de Dados Pessoais que define a não discriminação como um dos princípios do tratamento de dados pessoais:

Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:
IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;²⁵⁹

Nesse cenário, compreende-se, a partir da interpretação do dispositivo legal, que tratar indivíduos e grupos de forma distintas não enseja, necessariamente, discriminação. Em verdade, trata-se de uma cláusula de não proibição (isto é, possível em contexto não especificados pela

²⁵⁷ TEIXEIRA, Pedro S. Reconhecimento facial erra gênero de pessoa trans. São Paulo, **Folha de São Paulo**, 5 de mai. de 2024. Disponível em: <https://www1.folha.uol.com.br/tec/2024/05/reconhecimento-facial-erra-genero-de-pessoa-trans.shtml>

²⁵⁸ SOUSA, Bruno. **Panóptico: reconhecimento facial renova velhas táticas racistas de encarceramento**. Rede Observatórios de Segurança, 22 de abr. de 2022. Disponível em: <https://observatorioseguranca.com.br/panoptico-reconhecimento-facial-renova-velhas-taticas-racistas-de-encarceramento/>

²⁵⁹ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

legislação), com exceções, pois é requisito necessário que o tratamento diferenciado não aconteça de maneira ilegal e abusiva, especialmente no que tange a grupos minoritários.

Portanto, o conceito de discriminação adotado pela presente pesquisa carrega consigo necessariamente uma implicação negativa a determinados indivíduos ou coletividades que os impede de concorrer em condições justas com grupos majoritários pela igualdade material.²⁶⁰ Assim, segue-se o seguinte entendimento para fins de produção acadêmica:

Ao pensar no conceito de discriminação, geralmente imaginamos um cenário no qual certa pessoa é excluída de um grupo pelo fato de possuir determinada característica. Alguém é considerado menos merecedor de um emprego porque não frequentou uma universidade de primeira linha ou não é chamado para uma festa por ser considerado antissocial. Embora essa seja uma forma comum de se compreender a discriminação, o foco deste trabalho são os resultados discriminatórios que decorrem do fato de alguém pertencer a determinado grupo e ser julgado a partir das características desse grupo; um cenário no qual as características individuais de uma pessoa são desconsideradas, e aquela pessoa é vista somente como um membro de um dado conjunto de pessoas.²⁶¹

Por conseguinte, a partir dos dados e das inúmeras, mas não exaustivas situações narradas, nota-se que entre os grupos mais afetados, encontram-se as pessoas negras em uma lógica de perpetuação do racismo estrutural e do “racismo por denegação”²⁶² Esse é um fenômeno definido como “racismo algorítmico”, conceituado da seguinte maneira.

(...) o modo pelo qual a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados. Tal ordenação pode ser vista como uma camada adicional do racismo estrutural, que, além do mais, molda o futuro e os horizontes de relações de poder, adicionando mais opacidade sobre a exploração e a opressão global que já ocorriam desde o projeto colonial do século XVI.²⁶³

Nesse cenário, para que máquinas discriminem ilicitamente, comumente, encontram-se vieses (*bias*) nas etapas de treinamento.²⁶⁴ Ressalta-se que o termo não é oriundo da computação, embora seja uma das grandes barreiras éticas no que tange ao uso de IA. Nasce do estudo do

²⁶⁰ LINDOSO, Maria Cristine Branco. **Discriminação de gênero em processos decisórios automatizados**. Dissertação (Mestrado - Mestrado em Direito) - Universidade de Brasília, 2019, p. 51.

²⁶¹ MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. *RDU*, v. 16, n. 90, p. 39-64, nov.-dez. 2019.

²⁶² KREMER, Bianca. **Direito e tecnologia em perspectiva amefricana : autonomia, algoritmos e vieses raciais**. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Direito, 2021, p. 21.

²⁶³ SILVA, Tarcízio. **Mapeamento de Danos e Discriminação Algorítmica**. Desvelar, 2023. Disponível em: <https://desvelar.org/casos-de-discriminacao-algoritmica/>. Acesso em 21 de mai. de 2024.

²⁶⁴ BRAGA, Carolina. Discriminação nas decisões por algoritmos: polícia preditiva. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil Conteúdo e Tecnologia Ltda, 2019, p. 701.

pensamento humano, e pode ser definido como “pré-concepções que levam o cérebro a concluir questões de forma irrefletida [...], que atuam de maneira inconsciente”²⁶⁵ e, portanto, entende-se:

(...) vieses (*bias*) como pesos desproporcionais a favor ou contra algo ou alguém. Uma decisão enviesada, ou tendenciosa, ganha contornos de unilateralidade, isto é, é composta pela visão de mundo, pela experiência, pelos valores e até mesmo pela intuição de um sujeito ou de um grupo em relação ao contexto no qual está inserido.²⁶⁶

Tal definição é relevante para o contexto de decisões automatizadas tomadas por inteligência artificial diante dos dados de treinamento utilizados no método de aprendizado de máquina poderem conter padrões discriminatórios de classe social, raça, gênero, ideologia e outros aspectos de minorias sociais, os quais são reproduzidos sucessivamente nas camadas de processamento das *DNN*. Com efeito “falar de vieses algoritmos é assumir que humanos são preconceituosos, tendenciosos e tomam decisões de forma enviesada, o que se reflete na aprendizagem estatística”²⁶⁷, e, portanto, superar o fenômeno *black box* de forma exequível e não utópica se torna ainda mais urgente para evitar ao máximo a discriminação algorítmica.

Em contrapartida, há autores que reconhecem a capacidade de tomada de decisões enviesadas por parte de IAs, contudo, esse potencial é igualmente encontrado humanos também são responsáveis por julgamentos, inclusive, sendo suscetíveis a ruídos, definidos “como a variabilidade de julgamentos de um mesmo caso pela mesma pessoa em distintas ocasiões”²⁶⁸. A controvérsia está justamente no fato de que máquinas são capazes de fazer isso em escalas substancialmente maiores que humanos, gerando, inevitavelmente, erros com maior impacto e frequência, pois a amostragem de treinamento de ferramentas decisórias se referem a dados em massa do passado.²⁶⁹ Os dados inseridos no *input* das máquinas nunca se referem ao *hoje*, e, assim, geram uma assimetria temporal problemática, diante da constatação de que “ (...) a realidade é essencialmente aleatória. Fatos e fenômenos relevantes no passado não necessariamente serão relevantes no futuro.”²⁷⁰

²⁶⁵ JUNQUILHO, Tainá. **Inteligência Artificial no Direito: limites éticos**. Juspodivm, 2022, p. 77.

²⁶⁶ KREMER, Bianca et al. **Racismo Algorítmico**. Rio de Janeiro: O Panóptico, 2023. p. 27.

²⁶⁷ JUNQUILHO, Tainá. **Inteligência Artificial no Direito: limites éticos**. Juspodivm, 2022, p. 77.

²⁶⁸ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, pp. 81-82.

²⁶⁹ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, pp. 76-77.

²⁷⁰ KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, pp. 76-77.

Ressalta-se, similarmente, que um grande obstáculo para a igualdade material é a incompatibilidade entre a linguagem algorítmica com discussões subjetivas dotadas de juízos de valor que ocorrem em processos decisórios. Tal problematização decorre da supervalorização dos métodos puramente estatísticos que ignoram o dinamismo e a complexidade de determinadas situações humanas e partem do pressuposto de que “todos os aspectos da natureza humana podem ser convertidos em fórmulas matemáticas e que seres humanos podem ser julgados a partir delas”.²⁷¹ Ademais, ressalta-se que os modelos probabilísticos geram, inevitavelmente, diante da quantidade massiva de dados e consequente complexidade das redes neurais, riscos de correlações absurdas (“*allucinating patterns*”) e de resultados não intencionais, colocando em ponderação a acurácia tão eficiente das IAs *versus* a possibilidade de decisões injustas,²⁷² mormente no que tange às minorias sociais.

Desse modo, a partir do que fora argumentado, as propriedades de explicabilidade e interpretabilidade de modelos de IA e a explicações das decisões automatizadas que afetam negativamente grupos minoritários são indispensáveis como mecanismos de compreensão da discriminação algorítmica e de resultados não intencionais envidados de máquinas. Assim, a previsão e exigências de explicações interativas e contundentes, inclusive como muitas das plataformas explicaram nos casos apresentados, permite que os sujeitos de dados e os órgãos reguladores examinem os critérios e processos utilizados pelas inteligências artificiais, identificando possíveis violações aos princípios de igualdade e não discriminação.

3. A DIMENSÃO LEGISLATIVA DO DIREITO À EXPLICAÇÃO EM DECISÕES AUTOMATIZADAS

Se um direito à explicação não somente pode existir em termos técnicos (com certas limitações inerentes à tecnologia), como ainda existem fortes argumentos a fim de defender a sua existência como meio de assegurar institutos jurídicos como a autodeterminação informativa e o livre desenvolvimento da personalidade em ordenamentos jurídicos moldados pelo capitalismo de vigilância e pela sociedade algorítmica, convém compreender, portanto, se há previsão legal para

²⁷¹ FRAZÃO, Ana. Obstáculos para a consideração de questões éticas nos julgamentos algorítmicos. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, p. 38.

²⁷² FRAZÃO, Ana. Obstáculos para a consideração de questões éticas nos julgamentos algorítmicos. In: FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023, p. 39.

tal direito nas leis de proteção de dados e nas propostas de regulamentação da inteligência artificial. Em outros vocábulos, o foco do presente capítulo é investigar a dimensão legislativa deste direito: ele existe em legislações atuais e ou pode vir a existir em futuras? Trata-se de uma discussão restrita ao campo doutrinário e acadêmico ou existem mecanismos legais positivados para que se invoque tal direito por titulares de dados pessoais e pessoas afetadas por decisões automatizadas tomadas por modelos de inteligência artificial?

Para tal, serão analisados, em perspectiva comparada, normativos brasileiros e europeu. Optou-se por estabelecê-los como parâmetros devido às similaridades regulatórias do paradigma de proteção de dados, como a recente proximidade entre as abordagens de riscos, centradas na pessoa humana e garantidora de direitos sobre a regulação de inteligência artificial entre ambos os ordenamentos. Assim, analisar-se-á o Regulamento Europeu sobre a Proteção de Dados e o AI Act (recém-aprovado pelo Parlamento Europeu) como lente de análise para a União Europeia, com ressalvas convenientes à outros materiais legislativos como a Diretiva 95/46/CE do Parlamento Europeu, relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados — o normativo que originou o RGPD — e os *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* — frutos do *Article 29 Working Party* da Comissão Europeia. Em consonância, destrinchar-se-á as sementes de um possível direito à explicação em leis setoriais brasileiras e, em sequência temporal, fundamentos e dispositivos legais da Lei Geral de Proteção de Dados Pessoais e possíveis regulamentações de IA em tramitação no Congresso Nacional — PL 21/20, PL 2338/23 (versão inicial e relatório final) e PL 210/2024.

3.1. O DIREITO À EXPLICAÇÃO NO REGULAMENTO GERAL SOBRE A PROTEÇÃO DE DADOS DA UNIÃO EUROPEIA

O debate em torno da existência ou não um direito à explicação no contexto de decisões automatizadas na União Europeia — com a evidente ausência de alinhamento se existe, os seus limites e escopo de aplicação²⁷³ — orbita em torno de três correntes de autores, primordialmente: (a) os contrários por não haver previsão legal expressa e vinculativa; (b) os defensores por meio

²⁷³ CASEY, B.; FARHANGI, A.; VOGL, R. Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise. *Berkeley Technology Law Journal*, v. 34, 2018, p.145-189.

de uma interpretação sistemática; e (c) não ortodoxos aderentes à concepções de que a discussão deveria ter outro foco para além do texto do RGPD.²⁷⁴ À medida que os académicos das duas primeiras linhas buscam contradizer os argumentos de seus opositores em torno de instrumentos e recursos positivados no Regulamento para embasar ou não um direito à explicação²⁷⁵, a terceira corrente utiliza abordagens mais amplas ao considerar o posicionamentos de autoridades de proteção de dados, mecanismos de *accountability* como *Privacy By Design* e Relatório de Impacto à Proteção de Dados Pessoais (RIPD) e argumentos de ordem moral e filosófica como pontos de relevância a serem considerados.²⁷⁶

O artigo precursor da controvérsia remonta à época da aprovação do texto final do Regulamento no ano de 2016, em reconhecimento da “virada copernicana”²⁷⁷ que a legislação impunha sobre a regulação de proteção de dados, intitulado como *European Union regulations on algorithmic decision-making and a “right to explanation”*. Diferenciam, portanto, o alcance e escopo de aplicação de uma “Diretiva” e um “Regulamento”, pois no paradigma anterior na União Europeia, somente havia regras gerais orientativas para serem implementadas na legislação nacional por cada Estado-membro conforme interpretações e limitações locais, e, com o RGPD, torna-se a lei seguida, imprescindivelmente, por todos.²⁷⁸

Pioneiros, Goodman e Flaxman apresentam uma abordagem com enfoque em aspectos técnicos do funcionamento do aprendizado de máquina no processo de tomada de decisões. Para isso, distinguem o *machine learning* entre o aprendizado supervisionado e o não supervisionado. Os modelos supervisionados utilizam variáveis com valores predefinidos (rótulos) para analisar grandes volumes de dados e gerar resultados. Em contraste, os modelos não supervisionados trabalham com dados sem rótulos e buscam identificar padrões ou estruturas subjacentes que não são previamente conhecidas pelos programadores. Como resultado o fenômeno *black box* gera

²⁷⁴ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, pp. 53-54.

²⁷⁵ KIM, Tae Wan; ROUTLEDGE, Bryan R.; Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. **Business Ethics Quarterly**, v. 32, n°1, Jan. de 2022, pp. 75-102.

²⁷⁶ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, pp. 53-54.

²⁷⁷ GOODMAN, Bryce; FLAXMAN, Bryce. European Union regulations on algorithmic decision-making and a “right to explanation”. **ArXiv**, 31 de aug. de 2016, p.1.

²⁷⁸ GOODMAN, Bryce; FLAXMAN, Bryce. European Union regulations on algorithmic decision-making and a “right to explanation”. **ArXiv**, 31 de aug. de 2016, p.2.

demandas por maior transparência, sem ser explícito como essa exigência se apresenta no design algorítmico.²⁷⁹

Por conseguinte, exploram dois desafios práticos: a garantia de um direito a não discriminação expresso e um possível direito à explicação²⁸⁰. Os autores trazem o conceito de vieses e investigam as nuances técnicas do ato de discriminar no contexto de IA. No entanto, considera-se rasa a análise referente à explicação como uma garantia do titular de dados²⁸¹, haja vista que os autores buscam explorar a temática em somente uma página de todo o artigo acadêmico, retornando ao conceito de transparência, sem aprofundar nas implicações dos artigos do RGPD citados (arts. 13 a 15) e afirmam que a explicação deve ser compreensível a nível humano e esclareça como as entradas se correlacionam com as decisões efetuadas pelo sistema²⁸², e, seja possível inferir respostas para questionamentos como “É mais ou menos provável que o modelo recomende um empréstimo se o requerente for parte de uma minoria social? Quais recursos desempenham um papel mais relevante nesta previsão?”²⁸³ Assim, entende-se que, em verdade, o objetivo principal era estimular a reflexão para que outros trilhassem o caminho ainda a ser pavimentado.

Previamente à análise detalhada das vertentes que mergulham, de fato, no texto do RGPD para defender ou contrariar à existência ou não de um direito à explicação, convém, de forma pertinente para clareza da redação legislativa, principalmente no que tange aos vocábulos específicos utilizados, apresentar os principais dispositivos legais do Regulamento que permeiam o debate para compreender como os autores os correlacionam: os arts. 13, 14, 15 e 22 e o Considerando 71 (*Recital 71*).²⁸⁴

²⁷⁹ GOODMAN, Bryce; FLAXMAN, Bryce. European Union regulations on algorithmic decision-making and a “right to explanation”. *ArXiv*, 31 de aug. de 2016, p. 6.

²⁸⁰ GOODMAN, Bryce; FLAXMAN, Bryce. European Union regulations on algorithmic decision-making and a “right to explanation”. *ArXiv*, 31 de aug. de 2016, pp. 3- 6.

²⁸¹ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 60.

²⁸² GOODMAN, Bryce; FLAXMAN, Bryce. European Union regulations on algorithmic decision-making and a “right to explanation”. *ArXiv*, 31 de aug. de 2016, pp. 6-7.

²⁸³ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 141.

²⁸⁴ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 51.

Ressalta-se, no entanto, o caráter não formalmente vinculante do conhecidos *recitals* do Regulamento — embora apresentem contextos, justificativas e orientações para a aplicação e interpretação de determinados elementos do corpo legislativo — não implica, necessariamente, em um descarte do que está previsto, haja vista ser um elemento de impacto na construção de guias do “comportamento das empresas e entidades, o que corresponde à própria concepção da GDPR como um regime de governança colaborativa.”²⁸⁵ Desse modo, o considerando 71 forma a sustentação de argumentos em defesa de um direito à explicação no ordenamento jurídico europeu justamente por citar, expressamente, tal garantia, e, por consequência, também é elemento basilar da argumentação contrária, por se tratar de dispositivo não vinculativo e assumidamente ser escolha dos membros do Parlamento Europeu e não ser citá-lo no corpo obrigatório do RGPD, como pode ser visualizado:

O titular dos dados deverá ter o direito de não ficar sujeito a uma decisão, que poderá incluir uma medida, que avalie aspetos pessoais que lhe digam respeito, que se baseie exclusivamente no tratamento automatizado e que produza efeitos jurídicos que lhe digam respeito ou o afetem significativamente de modo similar, como a recusa automática de um pedido de crédito por via eletrónica ou práticas de recrutamento eletrónico sem qualquer intervenção humana. Esse tratamento inclui a definição de perfis mediante qualquer forma de tratamento automatizado de dados pessoais para avaliar aspectos pessoais relativos a uma pessoa singular, em especial a análise e previsão de aspetos relacionados com o desempenho profissional, a situação económica, saúde, preferências ou interesses pessoais, fiabilidade ou comportamento, localização ou deslocações do titular dos dados, quando produza efeitos jurídicos que lhe digam respeito ou a afetem significativamente de forma similar. (...) Em qualquer dos casos, tal tratamento deverá ser acompanhado das garantias adequadas, que deverão incluir **a informação específica ao titular dos dados e o direito de obter a intervenção humana, de manifestar o seu ponto de vista, de obter uma explicação sobre a decisão tomada na sequência desta avaliação** e de contestar a decisão (...) (grifo nosso)²⁸⁶

Os arts. 13 a 15 como um todo integram a seção de “*informação e acesso aos dados pessoais*”, à medida que o art. 13 e o art. 14, em conjunto, formam o rol de direitos de notificação e informação de controladores de dados, pois se referem, respectivamente, às “*Informações a facultar quando os dados pessoais são recolhidos junto do titular*” e “*Informações a facultar*

²⁸⁵ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 134.

²⁸⁶ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

quando os dados pessoais não são recolhidos junto do titular”²⁸⁷, compondo um rol de deveres de transparência para o responsável pelo tratamento dos dados pessoais.²⁸⁸ Em consonância, o art. 15 prevê “direito de acesso do titular dos dados”, com especial enfoque ao previsto na alínea “h” por abordar expressamente as decisões automatizadas e um direito à informação prévia de determinados aspectos envolvidos na automação decisória como a lógica do modelo e eventuais impactos para o indivíduo afetado:

1. O titular dos dados tem o direito de obter do responsável pelo tratamento a **confirmação de que os dados pessoais que lhe digam respeito são ou não objeto de tratamento** e, se for esse o caso, o direito de aceder aos seus dados pessoais e às seguintes informações:
 - h) A existência de decisões automatizadas, incluindo a definição de perfis, referida no artigo 22.o, n.os 1 e 4, e, pelo menos nesses casos, **informações úteis relativas à lógica subjacente, bem como a importância e as consequências** previstas de tal tratamento para o titular dos dados.²⁸⁹ (grifo nosso).

Esses três mecanismos integram o conceituado direito à informação significativa (*meaningful information*)²⁹⁰ quando ocorre o processo decisório durante um tratamento de dados pessoais. Por fim, o art. 22 assume papel de extrema relevância na discussão ao tratar, diretamente, do processamento de dados no contexto de decisões automatizadas, prevendo que o titular de dados pessoais — em uma cláusula geral proibitiva com exceções²⁹¹ — não pode sofrer decisões que impactem a sua dimensão jurídica tomadas unicamente por meios de automação:

1. O titular dos dados tem o **direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado**, incluindo a definição de perfis, que produza **efeitos na sua esfera jurídica ou que o afete significativamente** de forma similar.
2. O n.o 1 não se aplica se a decisão:
 - a) For necessária para a celebração ou a execução de um contrato entre o titular dos dados e um responsável pelo tratamento;
 - b) For autorizada pelo direito da União ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito, e na qual estejam igualmente previstas medidas adequadas

²⁸⁷ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

²⁸⁸ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 52.

²⁸⁹ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

²⁹⁰ SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017.

²⁹¹ EUROPEAN DATA PROTECTION BOARD. **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 679/2016**. Bruxelas: European Commission, 2016. Disponível em: https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=612053

para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados;
ou

c) For baseada no consentimento explícito do titular dos dados.

3. Nos casos a que se referem o n.o 2, alíneas a) e c), **o responsável pelo tratamento aplica medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados, designadamente o direito de, pelo menos, obter intervenção humana por parte do responsável, manifestar o seu ponto de vista e contestar a decisão.**

4. As decisões a que se refere o n.o 2 não se baseiam nas categorias especiais de dados pessoais a que se refere o artigo 9.o, n.o 1, a não ser que o n.o 2, alínea a) ou g), do mesmo artigo sejam aplicáveis e sejam aplicadas medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular.²⁹² (grifo nosso).

Assim, o artigo prevê três situações de exceção para processamentos totalmente automatizados, sendo a primeira delas a base legal de execução contratual, a segunda quando o direito local de um Estado-membro permitir e a terceira possibilidade é quando o titular consentir expressamente para que isso ocorra. O mais relevante, contudo, quanto ao art. 22, é a previsão estabelecida no (3) em que o agente de tratamento deve implementar salvaguardas (de transparência e accountability) para proteger o titular, além de prever, expressamente, o direito à intervenção humana e o de contestação e revisão da decisão tomada.

Nesse cenário, aborda-se o primeiro trabalho acadêmico produzido por pesquisadores do Instituto de Internet de Oxford que buscou responder aos apontamentos iniciais de Goodman e Flaxman, tornando-se, inevitavelmente, o cerne gravitacional da corrente contrária à existência de um direito à explicação: *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, por Wachter, Mittelstadt e Floridi. Para isso, estabelecem, inicialmente, premissas-chave de argumentação que permeiam múltiplas razões para questionar tanto a existência jurídica quanto a viabilidade prática de tal direito. Para os autores, em contraste com o suposto direito à explicação de decisões automatizadas específicas, o Regulamento exige, de forma expressa e vinculante, somente que os titulares dos dados recebam informações relevantes, embora devidamente limitadas (os arts. 13 a 15), sobre a lógica do modelo decisório, bem como a significância e as consequências destes, denominando-o como um “direito de ser informado”, mas não necessariamente de receber uma explicação.²⁹³ Ademais, ressaltam a ambiguidade e o escopo restrito de um direito a não ser submetido a decisão

²⁹² PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

²⁹³ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. *Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation*. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 1.

totalmente automatizada previsto no art. 22 e, assim, concluem que o RGPD carece de uma linguagem precisa, bem como de direitos e salvaguardas explícitas e bem delineadas contra a tomada de decisão automatizada.²⁹⁴

É fundamental compreender como os autores definem um direito à explicação, caso ele de fato existisse e quais seriam os seus contornos, e, para isso, diferenciam as explicações entre genéricas ou específicas. As primeiras se referem à operacionalidade do sistema, isto é, à lógica do modelo de IA, a eventuais resultados e à funcionalidade geral, como a especificação de requisitos do sistema, árvores de decisão e critérios e estruturas de classificação. As segundas explicações se relacionam à decisões particulares ligadas a determinado indivíduo ou coletividade, sendo, em outros termos, à justificativa, às razões e às circunstâncias individuais de uma decisão automatizada delimitada, a exemplo de informações sobre grupos de referência ou perfis.²⁹⁵

Outro aspecto relevante é a distinção das explicações quanto ao timing decisório, ou seja, quando elas ocorrem. Nessa senda, uma explicação *ex ante* ocorre antes da tomada de decisão automatizada (e esta somente pode se referir a como opera previamente o algoritmo, já que não é possível prever uma decisão que sequer ocorreu). Já as explicações *ex post* ocorrem após a realização de uma decisão automatizada (em contrapartida, esta pode abordar tanto a funcionalidade do sistema quanto a justificativa de uma decisão específica). Em ilustração, a situação de um sistema automatizado de pontuação de crédito pode ser instrumentalizada como exemplo, pois o desenvolvedor pode conceder informações acerca da lógica geral e consequências da decisão (como quem poderá utilizar a classificação para concessão ou não de crédito ao titular), e, após o processo decisório, aspectos específicos que a ensejaram naquele caso em particular podem ser identificados e compartilhados.²⁹⁶

Wachter, Mittelstadt e Floridi partem do pressuposto que a condição de existência de um direito à explicação no RGPD poderia vir a decorrer de três bases normativas: o art. 22 (3) (e o Considerando 71); os arts. 13 e 14 em representação de obrigações de notificação por parte do

²⁹⁴ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 1.

²⁹⁵ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 3.

²⁹⁶ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 3.

controlador (e os Considerandos 60-62); e o art. 15, na figura do direito de acesso (e seu respectivo Considerando, o 63).²⁹⁷ Desse modo, buscam refutar cada uma dessas possibilidades de argumentação favoráveis separadamente.

Com relação ao art. 22 (3), os autores esclarecem que são citadas expressamente diversas salvaguardas a serem asseguradas como garantias do titular de dados, como a intervenção humana e o direito de oposição, entretanto, em nenhum momento é mencionado diretamente nem com o uso de termos derivados e assumidamente sinônimos, e, como elemento complementar à essa ausência²⁹⁸, o artigo “sugere ter sido intencional, já que o texto do Considerando é praticamente idêntico ao do art. 22, à exceção deste único trecho sobre o direito à explicação.”²⁹⁹, e, assim, fora uma escolha do legislador europeu retirar o direito à explicação como previsão expressa.

Nesse contexto, a força jurídica de um *Recital* é evocada para ressaltar que o Considerando 71 se apresenta como tão somente um mecanismo legal de orientação interpretativa, porém, não é capaz de estabelecer um direito vinculante, nem positivizar institutos jurídicos por conta própria e muito menos gerar expectativas legais. Em complemento a esse argumento, utilizam jurisprudências da Corte Europeia de Justiça que reforçam o entendimento de que *Recitals* desempenham a função de dissolver ambiguidades, afastando a possibilidade desses dispositivos legais, em si, constituírem regras gerais obrigatórias. Por conseguinte, o art. 22 não exige nenhuma necessidade de recorrer ao Considerando 71 para equacionar alguma dificuldade de interpretação, haja vista ser escrito com linguagem clara acerca do que se é exigido. Em síntese, a discussão de uma possível recorrência a um *Recital* pode ocorrer no caso de, a exemplo, delimitar o escopo de aplicação de determinado direito, contudo, nunca se ele deve existir ou não, pois isso não pode haver dúvidas para os agentes de tratamento de dados.³⁰⁰

²⁹⁷ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 4.

²⁹⁸ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 4.

²⁹⁹ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 52.

³⁰⁰ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 5.

Ademais, recorre-se à história legislativa do Regulamento, com minuciosa investigação de suas versões iniciais, inclusive da Diretiva 95/46/CE, a fim de compreender com as devidas certezas se, de fato, ocultar expressamente um direito à explicação fora a intenção dos legisladores europeus. Em ordem cronológica, o texto proposto pelo Conselho Europeu, inicialmente, não continha a previsão legal do direito, e com emenda do Parlamento Europeu posterior, o direito à explicação fora sugerido a ser inserido com a limitação de ocorrer em um momento *ex post* à uma avaliação humana do processo decisório. No entanto, essa mudança não prevaleceu e não foi incluída, deliberadamente, no texto final do Regulamento sobre a Proteção de Dados.³⁰¹

Em segundo plano, quanto aos arts. 13 e 14, notadamente o art. 13 (2) e 14 (2) Wachter, Mittelstadt e Floridi argumentam o RGPD ao prever um de um direito de notificação, não necessariamente estabelece associada uma explicação, haja vista uma barreira de linguagem nesse aspecto, pois, notificar não implicar na ação de explicar, trata-se de um verbo que ocupa um tempo anterior à tomada de decisão.³⁰² Além disso, como esses dispositivos são utilizados como base para um direito à explicação *ex post* de decisões específicas quando correlacionados ao art. 22 (3), os autores ressaltam que ligar esses dispositivos legais uns aos outros, assim como utilizar o Considerando 71, partem de uma premissa de interpretação sistemática, pois não há nenhum indício de que a previsão do 22 (3) está interligada aos direitos de informação e notificação dos arts. 13 e 14³⁰³. Assim, o que se tem é tão somente um possível direito à explicação *ex ante*, haja vista o próprio “vocabulário dos artigos, que sugere uma informação genérica sobre como os sistemas funcionam (“lógica envolvida”) e uma previsão de impacto (“consequências vislumbradas”), antes de qualquer decisão.”³⁰⁴ Nesse sentido, o Considerando 60 emerge para corroborar o posicionamento de que se trata de um dever informacional anterior ao

³⁰¹ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 6.

³⁰² MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 143.

³⁰³ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 7.

³⁰⁴ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 143.

momento da decisão, isto é, como um elemento adicional a como interpretar o que está previsto, de fato, nos arts. 13 e 14:

Os princípios do tratamento equitativo e transparente exigem que **o titular dos dados seja informado da operação de tratamento de dados e das suas finalidades**. O responsável pelo tratamento deverá **fornecer ao titular as informações adicionais necessárias** para assegurar um tratamento equitativo e transparente tendo em conta as circunstâncias e o contexto específicos em que os dados pessoais forem tratados. O **titular dos dados deverá também ser informado da definição de perfis e das consequências que daí advêm**. (...) ³⁰⁵ (grifo nosso).

Em suma, não há como extrair um direito à explicação sobre decisões automatizadas específicas sobre um titular de dados do texto legislativo desses artigos, pois a construção linguística permite tão somente (e possivelmente), um direito de informação prévio acerca do funcionamento geral do algoritmo e de eventuais impactos do processo decisório, o que, por consequência lógica, não engloba nenhuma circunstância referente a um momento *ex post*.³⁰⁶

Por fim, os autores rebatem o terceiro pilar de um possível direito à explicação: derivá-lo do art. 15 que prevê o direito de acesso às informações acerca do tratamento de dados pessoais. Para tal, a lógica referente aos arts. 13 e 14 é novamente invocada, pois as previsões estabelecidas no art. 15 (1), alínea “h” se assemelha ao que os arts. 13 (2) e 14 (2), e, assim, os conceitua como verdadeiras “cartas magnas” dos direitos dos titulares no RGPD, compondo um conjunto unitário e consensual com a mesma estratégia de linguagem. Nesse aspecto, também não é possível derivar um direito *ex post*, diante da obrigatoriedade do controlador de dados apenas prever possíveis consequências que a decisão pode afetar na esfera individual do titular, mas não explicar, posteriormente a ela ser tomada, porque e como, especificamente, ela ocorreu com aquela pessoa e, a fim de corroborar que o direito à explicação *ex post* não pode decorrer do direito de acesso, destrincham a história desse instituto nas versões anteriores do Regulamento, como na Diretiva 95/46/CE.³⁰⁷

Nessa senda, o artigo suscita uma questão de grande relevância para o debate sobre o direito à explicação, especialmente em relação ao art. 22 do GDPR. Para Wachter, Mittelstadt e

³⁰⁵ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

³⁰⁶ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 8.

³⁰⁷ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, pp. 9-13.

Floridi, há divergências latentes na interpretação desse dispositivo como uma proibição ou um direito de oposição, pois as implicações dessa interpretação resultam em diferentes obrigações para agentes de tratamento. Caso seja considerada uma proibição, os controladores estariam impedidos de tomar decisões baseadas exclusivamente em processos automatizados sobre um titular, exceto as três possibilidades de execução de contrato, consentimento e direito local do Estado-membro (e garantisse as salvaguardas adequadas). Caso a interpretação tendesse, a restrição se aplicaria apenas mediante objeção ativa do titular. Em face dessas restrições e ambiguidades, os autores advogam por um “direito a ser informado” sobre a existência de decisões automatizadas e as funcionalidades dos sistemas empregados.³⁰⁸

Adicionalmente, o artigo argumenta que, embora não exista um direito à explicação no Regulamento, se houvesse movimentação de órgãos legislativos, cortes e autoridades de proteção de dados a fim de que de tal direito seja expresso, capaz de ser exigido, isso, por conseguinte, poderia fomentar a transparência e a responsabilização nos processos decisórios automatizados. Não obstante, mesmo que um direito à explicação para decisões específicas fosse estabelecido, outras disposições do texto do RGPD restringiram significativamente o seu escopo de aplicação, como a definição restritiva de processo decisório automatizado no art. 22 (1), que exige “efeitos legais ou outros efeitos significativos”³⁰⁹ de decisões baseadas “exclusivamente” em processos automatizados, ambos termos vagos e sujeitos a interpretações futuras.³¹⁰

Como réplica, Selbst e Powles publicam o trabalho intitulado como *Meaningful information and the right to explanation* que parte da mesma premissa de Wachter, Mittelstadt e Floridi: não existe previsão legal explícita denominada como “direito à explicação” no RGPD. No entanto, isso não necessariamente implica que tal direito não existe ou seja uma ilusão de quem o defende, pois os arts. 13 a 15 asseguram direitos a “informações relevantes sobre a lógica envolvida” (“*meaningful information*”) em decisões automatizadas, e, assim, defendem que esse

³⁰⁸ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 144.

³⁰⁹ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

³¹⁰ WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017, p. 20

conjunto legal constitui um direito à explicação, independentemente da utilização ou não dessa terminologia.³¹¹

Para Selbst e Powles, o direito à explicação deve ser interpretado de forma funcional, flexível e deve, ao menos, permitir que o titular dos dados exerça outros direitos assegurados no Regulamento e em normativos de direitos humanos, diante da aparente violação de que decisões automatizadas sem participação humana violam princípios de autonomia e personalidade, justificando disposições que assegurem informações relevantes aos titulares de dados.³¹²

Nesse contexto, os autores defendem a leitura sistemática e unificada dos arts. 13 (2), alínea “f”, art. 14 (2), alínea “g”, art. 15 (1), alínea “h” e art 22 suportam a existência direito à explicação no RGPD³¹³. Com efeito, os três primeiros dispositivo são construídos em lógica similar e redação legislativa praticamente idêntica, com a distinção de se tratar do contexto (se os dados pessoais são recolhidos em conjunto do titular ou não) preveem:

A existência de decisões automatizadas, incluindo a definição de perfis, referida no artigo 22.o, n.os 1 e 4, e, pelo menos nesses casos, **informações úteis relativas à lógica subjacente, bem como a importância e as consequências** previstas de tal tratamento para o titular dos dados.³¹⁴ (grifo nosso).

Desse modo, são apresentadas quatro observações por Selbst e Powles. A primeira delas é a de que os artigos 13 a 15 do GDPR versam sobre os direitos dos titulares dos dados, estipulando que as informações fornecidas devem ser pertinentes e inteligíveis para eles, em linguagem clara e acessível, nem ser um requisito conhecimento técnico acerca de linguagem de programação utilizada em algoritmos. Nesse sentido, o critério para determinar se a informação é pertinente deve ser funcional, permitindo ações como o direito de contestar uma decisão automatizada, conforme disposto no art. 22(3). Em segundo lugar, existem duas abordagens principais para compreender o valor da explicação: como um valor instrumental, que oferece uma maneira tangível de aferir a pertinência da explicação, ou como um valor intrínseco, associado à autonomia e à personalidade do indivíduo. O enfoque no valor instrumental é mais mensurável,

³¹¹ SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017, p. 1.

³¹² SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017, p. 1.

³¹³ SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017, p. 3.

³¹⁴ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

enquanto o valor intrínseco pode variar significativamente entre diferentes indivíduos e enfrentar conflitos com interesses estabelecidos, como o segredo comercial.³¹⁵

Além disso, deve existir um limiar mínimo de funcionalidade para as informações fornecidas, assegurando que sejam adequadas para que os titulares dos dados possam exercer seus direitos. Por último, a exigência de fornecer explicações deve ser interpretada com flexibilidade para não restringir desnecessariamente a pesquisa e o desenvolvimento. Regras excessivamente rígidas podem obstruir o uso eficaz de sistemas de aprendizado de máquina mais sofisticados, como redes neurais, que podem ser elucidados de maneiras alternativas e úteis.³¹⁶ Inclusive, nesse aspecto, utilizam de um arcabouço argumentativo relativo às traduções do RGPD para todas as línguas da União Europeia, pois encontram-se divergências de termos que não são propriamente sinônimos para se referir à “*meaningful information*”³¹⁷

Ademais, os autores argumentam que o artigo de Wachter, Mittelstadt e Floridi é problemático por duas razões principais. Primeiro, realiza manobras retóricas irresponsáveis, aparentando um argumento robusto contra o direito à explicação, enquanto na verdade o redefine e tenta refutar uma versão restrita propondo um novo direito que deveria ter sido incluído desde o início. De forma contraditória, eles introduzem um “direito a ser informado”, um direito limitado à explicação da funcionalidade dos sistemas de tomada de decisão automatizada, mas, que em verdade, trata-se de um direito à explicação *ex ante*. Segundo, o argumento central do trabalho acadêmico baseia-se em uma estrutura estática sustentada por pressupostos legais e técnicos não declarados, compondo uma estrutura fictícia tratada como verdade. No entanto, para Selbst e Powles essa estrutura é apenas uma abstração.³¹⁸ Assim, “considerados estes pontos, os autores afirmam que todos os argumentos de Wachter et al. não invalidam a existência de um direito à explicação em si, mas apenas da sua própria versão limitada deste direito.”³¹⁹

³¹⁵ SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017, p. 4.

³¹⁶ SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017, p. 4.

³¹⁷ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 150.

³¹⁸ SELBST, A. D.; POWLES, J. Meaningful Information and the Right to Explanation. **International Data Privacy Law**, v. 7, n. 4, p. 233-242, 2017, pp. 6-7.

³¹⁹ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 151.

Esses três artigos destrinchados compõem uma trilogia doutrinária que busca confrontar a existência ou não de um direito à explicação em decisões automatizadas por meio de dispositivos legais presentes no Regulamento sobre a Proteção de Dados. No entanto, a discussão não se restringe meramente à fonte jurídica legislativa, expandindo-se para outras formas de se argumentar a favor de um direito existir, como a jurisprudência de tribunais e o entendimento de entidades reguladoras como as de proteção de dados. Tais acepções serão exploradas a seguir como complemento ao debate.

3.2. O DIREITO À EXPLICAÇÃO PARA ALÉM DOS ARTS. 13, 14, 15 E 22 DO REGULAMENTO SOBRE A PROTEÇÃO DE DADOS

Em uma linha de defesa para a existência de um direito à explicação no ordenamento jurídico europeu, assim como Selbst e Powles, mas a partir de parâmetros não convencionais, os quais buscam explorar argumentos para além da previsão legal expressa ou não, Kim e Routledge fundamentam a favor em seu artigo *Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach* — teorizado com uma abordagem baseada na confiança entre titulares de dados e empresas de tecnologia — sob a égide de elementos filosóficos éticos e morais, intencionalmente transbordando as fronteiras legais. Para isso, partem de duas premissas: (a) o instituto jurídico do consentimento informado em cenários de decisões automatizadas é insuficiente sem uma garantia de confiança; e (b) a iniciativa empresarial de cumprir um possível um direito à explicação *ex post* permite estabelecer uma ponte (possível) de confiança com usuários. Assim, infere-se, por meio de lógica silogística, que um direito à explicação posterior auxilia a permitir o consentimento informado.³²⁰

Nesse contexto, relatam um caso real para ilustrar como a relação entre consumidor e empresa enseja a reflexão de um direito à explicação. Em 2019, a Apple e a Goldman Sach — um dos mais relevantes agentes de banco de investimentos globais — instituíram parceria comercial para a produção de um *Apple Card* para os seus clientes. Um casal aplicou para utilizá-lo, no entanto, deparou-se com a seguinte situação: Jamie Hansson aplicou e recebeu um limite de crédito vinte vezes menor do que seu marido, David Hansson, embora ela possuísse

³²⁰ KIM, Tae Wan; ROUTLEDGE, Bryan R.; Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. **Business Ethics Quarterly**, v. 32, n°1, pp. 75-102. Jan. de 2022, pp. 2-3.

uma pontuação de crédito melhor. O incidente evocou preocupações sobre discriminação de gênero nos algoritmos de avaliação de crédito usados pelo *Apple Card*, decorrendo uma investigação do Departamento de Serviços Financeiros de Nova York para verificar possíveis violações da lei estadual contra discriminação. Quando o serviço de atendimento ao cliente da Apple fora contatado, a atendente culpou o algoritmo black box que decidiu, sozinho e automaticamente, quem receberia mais crédito.³²¹

Outro caso narrado é o conhecido da empresa Target e uma consumidora grávida. O pai de uma adolescente ainda no ensino médio reclamou com a loja, pois a filha recebia anúncios personalizados de cupons de roupas de bebê. No entanto, foi descoberto que a garota estava realmente grávida quando o progenitor buscou informações dos motivos dessa propaganda direcionada em particular. Quem fez a revelação e decidiu indicá-la foi a inteligência artificial da Target.³²² Assim, em situações como a relatada, o direito à explicação surgiria como um importante mecanismo moral para compreender nuances de uma decisão, previsão ou recomendação.³²³

Os autores buscam analisar os tipos de explicação que foram apresentadas por Wachter, Mittelstadt e Floridi. A partir disso, compreendem que podem haver três tipos diferentes de explicação: (a) explicação *ex ante* sobre como opera um sistema (explicação genérica anterior); (b) explicação *ex post* de funcionamento do modelo algorítmico (explicação genérica posterior); e (c) explicação *ex post* de uma decisão referente a um indivíduo ou coletividade (explicação específica posterior). A primeira se refere ao “direito de ser informado” à medida que as explicações posteriores se diferem destas em razão do processamento de dados em máquinas, mormente as de redes neurais, alterar a lógica envolvida antes e após a decisão. Desse modo, se uma empresa, como no caso *Apple Card* e Target, prejudicarem um indivíduo com instrumentos

³²¹ HANSSON, Jamie Heinemeier. I Applied for an Apple Card. What they offered was a sexist insult. **Fast Company**, 11 de nov. de 2019. Disponível em: <https://www.fastcompany.com/90429224/i-applied-for-an-apple-card-what-they-offered-was-a-sexist-insult>.

³²² HILL, Kashmir. How Target Figured Out a Teen Girl Was Pregnant before Her Father Did. **Forbes**, 16 de fev. de 2012. Disponível em: <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-fat-her-did/>

³²³ KIM, Tae Wan; ROUTLEDGE, Bryan R.; Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. **Business Ethics Quarterly**, v. 32, n°1, pp. 75-102. Jan. de 2022, p. 4

automatizados, ele possui o direito de receber as explicações posteriores (genéricas e específicas), em um denominado como “direito à explicação como remédio” (em tradução livre).³²⁴

Figura 4 - Tipos de explicação para cada titular de dados

Table 2: Three Different Types of Data Subjects and Their Rights

Data subjects	Harmed	Influenced	Benefit others	Right	Explanation
Type 1	Yes	Yes	Yes	Both remedial and updating explanations	Specific
Type 2	No	Yes	Yes	Only updating explanation	Moderately specific
Type 3	No	No	Yes	Only updating explanation	Less specific

Fonte: KIM, Tae Wan; ROUTLEDGE, Bryan R. (2022, p. 32)

No entanto, se o direito à explicação decorre de valores filosóficos éticos e morais, não necessariamente o elemento dano deve existir para que uma pessoa ou grupo delas o invoque de forma legítima. Ilustram a situação hipotética de um estudante em busca de um emprego, o qual pesquisa novas oportunidades em ferramentas de busca como o Google (máquinas de aprendizado), no entanto, há uma bolha de filtragem de resultados devido às buscas anteriores do usuário. Outro indivíduo fazendo a mesma pesquisa, com os mesmos termos e palavras-chave, poderia visualizar páginas completamente diferentes. Quais fatores ensejam essa distinção que pode impactar significativamente no acesso à serviços e no exercício de possíveis direitos fundamentais, como o de trabalho? Esse questionamento paira sobre quem é uma minoria social ou grupo vulnerável, que, constantemente, sofrem da discriminação algorítmica. Esse estudante, portanto, deveria poder entender como seus dados de navegação prévia influenciam a procura por emprego atual. Para os autores, nesses casos, é um direito à atualização (“*update*”).³²⁵

Nesse contexto, uma explicação atualizada (“*updated explanation*”) é o elemento que estabelece confiança entre o usuário e o empreendimento com o poder de tomada de decisões automatizadas. Argumentam, por conseguinte, que a garantia de um direito à explicação apesar

³²⁴ KIM, Tae Wan; ROUTLEDGE, Bryan R.; Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. **Business Ethics Quarterly**, v. 32, n°1, pp. 75-102. Jan. de 2022, pp. 6-8.

³²⁵ KIM, Tae Wan; ROUTLEDGE, Bryan R.; Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. **Business Ethics Quarterly**, v. 32, n°1, pp. 75-102. Jan. de 2022, pp. 8-9.

de ocorrer dano ou não é o laço entre empresa-consumidor, diante de uma economia do *Big Data* que interfere e afeta aqueles não objetivamente prejudicados, com contornos mais sutis e nebulosos, como no caso de perfilamentos em massa, direcionamento de conteúdos em redes sociais e propagandas personalizadas.³²⁶ Nessas situações, é razoável, portanto, que a explicação assuma um formato distinto, menos robusto e menos específico, dos contextos de indivíduos serem diretamente prejudicados, mas, que ainda sim, possa ser exigida, como pode ser visualizado na tabela abaixo:

Figura 5 - Contornos de um direito à explicação

Table 1: Contours of a Right to Explanation

	Ethicists' view (philosophical)	Data scientists' view (operational)
Ex ante	The traditional notion of informed consent (or a right to ex ante explanation)	Offering an ex ante generic explanation
Ex post	A right to remedial explanation (or ex post explanation for redress)	Offering both an ex post generic explanation and an ex post specific explanation
	A right to updating explanation (or an ex post explanation for opt-out)	Offering both an ex post generic explanation and an ex post specific explanation

Fonte: KIM, Tae Wan; ROUTLEDGE, Bryan R. (2022, p.)

As próximas iniciativas direcionam o olhar da discussão sob a ótica dos conceitos e mecanismos de *compliance* e regime de *accountability* como uma abordagem sistêmica³²⁷. O segundo já fora previamente explorado ao longo do presente trabalho, convém, então, produzir breves apontamentos acerca da definição de *compliance* e como se apresenta como ferramenta relevante para a explicabilidade de decisões automatizadas tomadas por inteligências artificial. Oriundo originalmente do inglês, e traduzido livremente, significa “conformidade”. Em termos acadêmicos, pode ser entendido, no contexto do processamento de dados, como:

A adesão dos agentes de tratamento de dados pessoais a padrões e normas, oriundas de leis, do mercado e das próprias organizações, aos princípios de boa governança e aos padrões éticos e sociais comumente aceitos.³²⁸

³²⁶ KIM, Tae Wan; ROUTLEDGE, Bryan R.; Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. *Business Ethics Quarterly*, v. 32, n°1, pp. 75-102. Jan. de 2022, pp. 16-18.

³²⁷ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 55.

³²⁸ MULHOLLAND, Caitlin; GOMES, Rodrigo Dias de Pinho. Inteligência Artificial e seus principais desafios para os programas de compliance e as políticas de proteção de dados. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas. **Compliance e Políticas de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2021, p. 163.

Para isso, ferramentas práticas devem ser implementadas, como códigos e políticas de condutas que estabelecem procedimentos padronizados de atuação, regras gerais de governança, produção de relatórios de riscos e adoção de padrões de *privacy by design*.³²⁹ Em sentido amplo, ele também pode ser definido da seguinte maneira:

O compliance pode ser visto como um conjunto de ações a serem adotadas no ambiente corporativo para que se reforce a anuência da empresa à legislação vigente, de modo a prevenir a ocorrência de infrações ou, já tendo ocorrido o ilícito, propiciar o imediato retorno ao contexto de normalidade e legalidade.³³⁰

Assim, conjuntamente à noção de *accountability*, integram o panorama orientativo para não somente empresas, como ainda a Administração Pública deve, em termos práticos, adequarem-se às previsões normativas regulatórias de proteção de dados e inteligência artificial. Sob essa égide, outros autores do ordenamento jurídico europeu buscam trazer novas visões para o direito à explicação, inclusive, compreendendo que talvez o foco não deva ser discutir se ele existe ou não no RGPD. Conforme Kate Kaminski, o Regulamento sobre a Proteção de dados inaugura um paradigma de responsabilização algorítmica mais robusto e exigente do que o oferecido pela Diretiva 95/46/CE e o debate proposto pela tríade de artigos precursores em torno somente dos arts. 13, 14, 15 e 22 do Regulamento é excessivamente restrito, principalmente no que tange à se propor discutir existência ou não de um direito à explicação.³³¹

Para tal, evocam o guia orientativo do *Article 29 Working Party* — atualmente denominado como Conselho Europeu de Proteção de Dados — como elemento essencial de orientar como as empresas deverão se adequar à regulamentação, diante de nascer com o objetivo de padronizar entendimentos, tornando-os mais consensuais para o *enforcement* de entidades reguladoras. Em seguida, a autora se debruça sobre a natureza jurídica dos Considerandos, refutando aqueles que desconsideram por não ser vinculante o conteúdo presente no preâmbulo do RGPD, pois a regulação fora construída para ser “fundamentalmente colaborativa”³³². Assim, é totalmente plausível compreender que “se o texto do Regulamento é repleto de parâmetros

³²⁹ MULHOLLAND, Caitlin; GOMES, Rodrigo Dias de Pinho. Inteligência Artificial e seus principais desafios para os programas de compliance e as políticas de proteção de dados. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas. **Compliance e Políticas de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2021, p. 163.

³³⁰ FRAZÃO, Ana. Propósitos, desafios e parâmetros gerais dos programas de compliance e das políticas de proteção de dados. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas. **Compliance e Políticas de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2021, p. 36.

³³¹ KAMINSKI, M. E. The Right to Explanation, Explained. **Berkeley Technology Law Journal**, v. 34, n. 189, 2019, p. 5.

³³² KAMINSKI, M. E. The Right to Explanation, Explained. **Berkeley Technology Law Journal**, v. 34, n. 189, 2019, p. 7.

abertos, isso decorre de um objetivo maior de que a sua substância seja desenhada ao longo do tempo por meio de um diálogo constante entre reguladores e entidades”.³³³ Dessa maneira, os *Recitals* e os *Guidelines* compõem um sistema dinâmico de *accountability*.

Para ela, não se deu a devida importância, nos trabalhos acadêmicos anteriores, as previsões acerca do art. 22 do RGPD no guia do Conselho de Proteção de Dados Europeu, pois, para escapar da aplicação de tal dispositivo, conforme o documento, é necessário que a participação humana seja significativa, e, assim, controladores de dados que tentarem, obtusamente, burlá-lo inserindo etapas com “humanos” envolvidos sem a interferência mínima necessária para alterar a decisão, continuam sob o escopo do artigo, e, assim deve ser mantido o entendimento por autoridades de proteção de dados e cortes europeias.³³⁴

O artigo 22.º, n.º 1, alude às decisões «**tomadas exclusivamente com base**» no tratamento automatizado. Por outras palavras, não há nenhuma intervenção humana no processo decisório.

O responsável pelo tratamento não pode eximir-se do disposto no artigo 22.º fabricando uma intervenção humana. Por exemplo, se alguém aplicar de forma sistemática perfis gerados automaticamente a pessoas sem ter qualquer influência efetiva no resultado, tratar-se-á, ainda assim, de uma decisão tomada exclusivamente com base no tratamento automatizado. Para que se considere haver uma intervenção humana, **o responsável pelo tratamento tem de garantir que qualquer supervisão da decisão seja relevante, e não um mero gesto simbólico.** Essa supervisão deve ser levada a cabo por alguém com autoridade e competência para alterar a decisão e que, no âmbito da análise, deverá tomar em consideração todos os dados pertinentes.³³⁵ (grifo nosso)

Analogamente, o raciocínio também se aplica à limitação de quais decisões podem ser consideradas capazes de gerar resultados significativos na esfera individual e coletiva de um indivíduo ou grupo previstos no art. 22, ressaltando que, para engatilhar os efeitos jurídicos do dispositivo legal, os impactos devem ser seriamente consideráveis,³³⁶ como pode ser visualizado:

Para haver efeitos jurídicos, é necessário que a decisão, tomada exclusivamente com base no tratamento automatizado, **afete os direitos legais de alguém**, como, por exemplo, a liberdade de associação com outrem, de votar numa eleição ou de intentar ações judiciais. **Um efeito jurídico pode igualmente verificar-se quando o estatuto jurídico de uma pessoa ou os seus direitos no âmbito de um contrato são afetados.** Como exemplos deste tipo de efeito contam-se as decisões automatizadas sobre uma pessoa que resultam: • na rescisão de um contrato; • na atribuição ou recusa de uma

³³³ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 151.

³³⁴ KAMINSKI, M. E. The Right to Explanation, Explained. *Berkeley Technology Law Journal*, v. 34, n. 189, 2019, p. 13.

³³⁵ EUROPEAN COMMISSION. **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679**. Disponível em: <https://ec.europa.eu/newsroom/article29/items/612053>

³³⁶ KAMINSKI, M. E. The Right to Explanation, Explained. *Berkeley Technology Law Journal*, v. 34, n. 189, 2019, p. 14.

prestação social específica prevista na legislação, tais como um abono de família ou um subsídio de habitação; • na recusa de admissão num país ou no indeferimento de um pedido de aquisição de nacionalidade.³³⁷ (grifo nosso)

Dessa forma, pode-se sintetizar que o pensamento de Kaminski decorre da definição de um devido processo algorítmico individual, diante do fato da existência de um direito à explicação ser fundamental para que um titular de dados possa exercer outros direitos, como contestar uma decisão ou expressar sua opinião (direito de oposição, por exemplo), que estão explicitamente previstos no texto da GDPR. Em síntese, um indivíduo deve possuir o direito à explicação de uma decisão automatizada diante de um panorama de interdependência de direitos estabelecido pelo Regulamento. Uma pessoa somente pode impugnar uma decisão específica ou expressar seu ponto de vista se ele realmente compreender “como ela foi tomada e com base em quê”³³⁸.

Assim, o posicionamento da autora é de que o Regulamento e seus documentos acessórios interpretativos exigem essa clareza explicativa, em contraponto à lógica argumentativa de Wachter, Mittelstadt e Floridi. Nesse sentido, é relevante ponderar que, no fim, não se tratam de informações e explicações irrestritas da natureza do algoritmo. Em verdade, se o titular dos dados não possui (e não precisa possuir) acesso a códigos-fonte e outras informações complexas em linguagem de programação, outros atores no sistema podem ter esse acesso, e existem mecanismos apropriados para que essas informações sejam apresentadas de maneira que respeite os segredos das empresas. Por outro lado, as medidas de responsabilidade sistêmica significam que o direito individual à explicação e os direitos de transparência, conforme garantidos pela GDPR, asseguram acesso a toda e qualquer informação relevante sobre um algoritmo.³³⁹

Nesse contexto similar, a partir de uma perspectiva mais contemporânea, Casey, Farhangi e Vogl — em seu trabalho *Rethinking Explainable Machines: the GDPR’s “Right to Explanation” Debate and the Rise of Algorithmic Audits in Enterprise* — atentam-se ao protagonismo que autoridades de proteção de dados possuem na construção de um ambiente regulatório estimulante à medidas de compliance de empresas por meio princípios de *Privacy by*

³³⁷ EUROPEAN COMMISSION. **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679**. Disponível em: <https://ec.europa.eu/newsroom/article29/items/612053>

³³⁸ KAMINSKI, M. E. The Right to Explanation, Explained. **Berkeley Technology Law Journal**, v. 34, n. 189, 2019, p. 16.

³³⁹ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 151.

Design (PbD), transferindo a ótica de um paradigma centralizado no indivíduo (e um direito à explicação) para uma abordagem sistêmica.³⁴⁰ Para isso, direcionam o olhar para os capítulos 6 e 8 do Regulamento, os quais, prevêm, respectivamente, redações legislativas sobre “autoridades de controle independentes” e “vias de recurso, responsabilidade e sanções”. No capítulo 6, são estabelecidas as diretrizes para a criação, função, composição e prerrogativas das autoridades de proteção, à medida que o capítulo 8 prevê faculdades ao titular de dados em ingressar com reclamações e ações judiciais, direito de indenizações, além de prever regras gerais para aplicação de sanções e multas aos controladores de dados.³⁴¹

Por conseguinte, o pensamento de Casey, Farhangi e Vogl se funda em uma acepção ampla de um direito à explicação, pois se trata de um remédio capaz de gerar resultados que transcendem a esfera individual, alterando a lógica de uma explicação *ex post* para necessidade de se estabelecer uma abordagem sistêmica *ex ante* de avaliação de riscos, desonerando a figura do titular dos dados, de forma a partilhar o ônus da explicabilidade com entidades reguladoras, controladores e outros *stakeholders*.³⁴²

Além disso, lembram que quando o uso de novas tecnologias pode resultar em alto risco para os direitos dos titulares, um *Data Protection Impact Assessment (DPIA)* deve ser produzido (o conhecido Relatório de Impacto à Proteção de Dados - RIPD - no contexto brasileiro), conforme a previsão do art. 35 do RGPD, o qual não estabelece em termos exatos como deve ser formatado e composto um *DPIA*, mas o *Article 29 Working Party* concede parâmetros mais delimitados para empresas seguirem.³⁴³ Conforme o art. 35 Regulamento, o documento deve ser, obrigatoriamente elaborado no caso de decisões automatizadas com impactos significativos:

1. Quando um certo tipo de tratamento, em particular que utilize novas tecnologias e tendo em conta a sua natureza, âmbito, contexto e finalidades, for suscetível de implicar **um elevado risco para os direitos e liberdades das pessoas singulares, o responsável pelo tratamento procede, antes de iniciar o tratamento, a uma avaliação de impacto das operações de tratamento previstas sobre a proteção de dados pessoais.** Se um

³⁴⁰ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 54.

³⁴¹ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

³⁴² CASEY, B.; FARHANGI, A.; VOGL, R. Rethinking Explainable Machines: The GDPR’s “Right to Explanation” Debate and the Rise of Algorithmic Audits in Enterprise. **Berkeley Technology Law Journal**, v. 34, p. 145-189, 2018, p. 19.

³⁴³ CASEY, B.; FARHANGI, A.; VOGL, R. Rethinking Explainable Machines: The GDPR’s “Right to Explanation” Debate and the Rise of Algorithmic Audits in Enterprise. **Berkeley Technology Law Journal**, v. 34, p. 145-189, 2018, p. 29.

conjunto de operações de tratamento que apresentar riscos elevados semelhantes, pode ser analisado numa única avaliação. (...)

3. A realização de uma avaliação de impacto sobre a proteção de dados a que se refere o n.º 1 é **obrigatória** nomeadamente em caso de:

a) Avaliação sistemática e completa dos aspetos pessoais relacionados com pessoas singulares, **baseada no tratamento automatizado, incluindo a definição de perfis, sendo com base nela adotadas decisões que produzem efeitos jurídicos relativamente à pessoa singular ou que a afetem significativamente de forma similar;**³⁴⁴ (grifo nosso)

De forma a suplementar tal previsão, os *Guidelines* do Conselho Europeu de Proteção de Dados estabelecem, como exemplo de medidas em um relatório de impacto, as seguintes salvaguardas:

As AIPD podem igualmente constituir um meio útil de o responsável pelo tratamento identificar as medidas a introduzir para fazer face aos riscos em matéria de proteção de dados resultantes do tratamento. Podem incluir-se entre estas medidas:

- **informar o titular dos dados da existência de um processo de decisão automatizada e da lógica subjacente;**
- **explicar a importância e as consequências previstas do tratamento para o titular dos dados;**
- **proporcionar ao titular dos dados meios de se opor à decisão;** e
- **permitir ao titular dos dados manifestar o seu ponto de vista.**³⁴⁵ (grifo nosso)

Em complemento, Lilian Edwards e Michael Vale, em célebre artigo *Slave to the Algorithm? Why a Right to an Explanation is Probably not the Remedy You are Looking For*, também ressaltam a importância do *Privacy by Design*, destrinchando instrumentos de efetivação desse modelo, pois, para os autores, “um direito desta natureza dificilmente representaria um remédio adequado para responder a todos os riscos apresentados por algoritmos e tecnologias correlatas”³⁴⁶, em uma insuficiência de mera existência do direito à explicação.

Nesse sentido, os Edwards e Vale entendem que, com o intuito de fomentar a construção de sistemas automatizados mais seguros para a privacidade, o RGPD traz novas disposições que não concedem direitos individuais de forma objetiva (como o direito à explicação), mas estabelecem um ambiente propício para assegurá-los por meio dessas medidas buscam integrar a privacidade desde o início do desenvolvimento dos sistemas, ao invés de adicioná-la posteriormente (PbD). Com a consciência de que os reguladores não podem controlar, regular e

³⁴⁴ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

³⁴⁵ EUROPEAN COMMISSION. **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679**. Disponível em: <https://ec.europa.eu/newsroom/article29/items/612053>

³⁴⁶ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 55.

fiscalizar de modo onisciente e onipresente, as disposições requerem que os controladores adotem “medidas técnicas e organizacionais apropriadas” durante o desenvolvimento e o processamento dos dados, assegurando a proteção dos direitos dos titulares. Ressaltam, ainda, que todas as autoridades públicas, grandes empresas do setor privado e controladores que lidam com categorias especiais de dados devem nomear um Encarregado de Proteção de Dados (DPO).³⁴⁷

Por fim, Edwards e Vale evocam a figura legal do art. 42 do RGPD, o qual prevê:

1. Os Estados-Membros, as autoridades de controlo, o Comité e a Comissão promovem, em especial ao nível da União, **a criação de procedimentos de certificação em matéria de proteção de dados, bem como selos e marcas de proteção de dados**, para efeitos de comprovação da conformidade das operações de tratamento de responsáveis pelo tratamento e subcontratantes com o presente regulamento. Serão tidas em conta as necessidades específicas das micro, pequenas e médias empresas. (...)
3. A certificação é **voluntária** e está disponível através de um processo transparente.³⁴⁸ (grifo nosso)

Assim, o dispositivo estabelece a possibilidade de uma “certificação” voluntária de controladores e processadores para demonstrar conformidade com o Regulamento, introduzindo “mecanismos de certificação” e “selos e marcas” promovidos pela União Europeia. Para os autores, a certificação de algoritmos pode ser realizada de duas maneiras: a) especificando diretamente as especificações de design ou o processo de desenvolvimento (padrões tecnológicos) e; b) estabelecendo requisitos relacionados aos resultados (padrões baseados em desempenho).³⁴⁹

Dessa forma, apresentou-se a discussão inicial acerca do ordenamento jurídico europeu quanto à existência ou não de um direito à explicação no contexto de decisões automatizadas previsto legalmente, e, como complementação, abordou-se contrapontos que extrapolam os textos produzidos à época da aprovação do RGPD, com um enfoque dinâmico e prático de meios de viabilizar não somente um direito à explicação, mas a explicabilidade de sistemas, por meio de abordagens sistêmicas de *accountability*, *compliance* e *enforcement* de entidades reguladoras. Portanto, convém racionalizar como esse processo ocorre no ordenamento jurídico brasileiro, quais são as semelhanças e distinções quanto ao posicionamento da União Europeia.

³⁴⁷ EDWARDS, L.; VEALE, M. Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. **Duke Law & Technology Review**, v. 16, p. 18-84, 2017, p. 60.

³⁴⁸ PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

³⁴⁹ EDWARDS, L.; VEALE, M. Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. **Duke Law & Technology Review**, v. 16, p. 18-84, 2017, p. 62.

3.3. O DIREITO À EXPLICAÇÃO NO CÓDIGO DE DEFESA DO CONSUMIDOR E NA LEI GERAL DE PROTEÇÃO DE DADOS BRASILEIRA

No ordenamento jurídico brasileiro, embora com elementos semelhantes à discussão do ordenamento europeu, a temática do direito à explicação em decisões automatizadas possui nuances próprias. Em verdade, trata-se, em termos doutrinários, de uma forte tendência a ser um consenso entre autores em defesa da existência, a partir de leitura histórica e setorial de garantias do setor consumerista e de abordagem sistemática da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709 ou LGPD), de forma similar a como Selbst e Powles constroem o raciocínio por meio de princípios, considerandos e outros direitos no escopo do RGPD. Por conseguinte, será realizada revisão bibliográfica dos principais artigos acadêmicos e livros de doutrina acerca da temática, em conjunto com o exame dos aspectos legislativos de dispositivos da LGPD.

Em um momento histórico de vácuo de uma regulação nacional unificadora da regulação de proteção de dados, debruça-se sobre a matéria do direito do consumidor, em que se fundam, como previamente apresentado nesta monografia, as fundações do paradigma de proteção de dados brasileiro. Nesse sentido, Renato Leite Monteiro constroi as bases da explicação em processos decisórios consumeristas por meio da análise do Código de Defesa do Consumidor (Lei nº 8.078/90 ou CDC) e a Lei do Cadastro Positivo, em que, juntos, compõem um microsistema explicativo na concessão de crédito.³⁵⁰ O CDC é uma regulação setorial que abrange as relações de consumo, online ou offline, estabelecendo transparência e boa-fé como princípios fundamentais. O STJ define a boa-fé objetiva como um modelo ideal de conduta para todos os participantes da relação obrigacional, visando o correto cumprimento das obrigações. A interpretação conjunta do CDC e decisões do STJ indicam o dever de informar o consumidor de maneira clara, abrangendo desde a formação da relação contratual até a máxima transparência dos registros de consumo.³⁵¹

Assim, depreende-se tanto um direito de acesso, quanto um direito à informações claras e compreensíveis ao consumidor. Isso pode ser visualizado no art. 43 da Lei 8.078/90:

³⁵⁰ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, p. 10.

³⁵¹ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, p. 8.

Art. 43. O consumidor, sem prejuízo do disposto no art. 86, **terá acesso às informações existentes em cadastros, fichas, registros e dados pessoais e de consumo arquivados sobre ele**, bem como sobre as suas respectivas fontes.

§ 1º Os cadastros e dados de consumidores devem ser **objetivos, claros, verdadeiros e em linguagem de fácil compreensão**, não podendo conter informações negativas referentes a período superior a cinco anos.³⁵²

Desse modo, pode-se inferir que o consumidor tem direito de acessar os dados que fundamentam a decisão automatizada no contexto de relações de consumo. A Lei do Cadastro Positivo (Lei 12.414/2011) segue essa lógica³⁵³, visando reduzir a assimetria informacional e proteger dados pessoais de consumo, garantindo diversos direitos referentes à “disciplina e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para a formação de histórico de crédito”³⁵⁴, inclusive um direito à explicação no art. 5º (IV)³⁵⁵:

Art. 5º São direitos do cadastrado:

IV - **conhecer os principais elementos e critérios considerados para a análise de risco**, resguardado o segredo empresarial;

V - **ser informado previamente** sobre o armazenamento, a identidade do gestor do banco de dados, o objetivo do tratamento dos dados pessoais e os destinatários dos dados em caso de compartilhamento;

VI - solicitar ao consulente **a revisão de decisão realizada exclusivamente por meios automatizados**; e

VII - ter os seus dados pessoais utilizados somente de acordo com a finalidade para a qual eles foram coletados.³⁵⁶ (grifo nosso).

Conforme Monteiro, esses quatro incisos da Lei do Cadastro Positivo, em conjunto com o art. 43 do CDC, “decorrem do princípio da transparência e não-discriminação e formam a espinha dorsal do direito à explicação de decisões automatizadas em relação de consumo.”³⁵⁷

³⁵² BRASIL. **Lei 8.078 de 11 de setembro de 1990 (Código de Defesa do Consumidor)**. Dispõe sobre a proteção do consumidor e dá outras providências. Disponível em: http://www.planalto.gov.br/ccivil_03/Leis/L8078.htm.

³⁵³ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, p. 9.

³⁵⁴ BRASIL. **Lei 12.414 de 9 de junho de 2011 (Lei do Cadastro Positivo)**. Disciplina a formação e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para formação de histórico de crédito. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112414.htm

³⁵⁵ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, p. 9.

³⁵⁶ BRASIL. **Lei 12.414 de 9 de junho de 2011 (Lei do Cadastro Positivo)**. Disciplina a formação e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para formação de histórico de crédito. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112414.htm

³⁵⁷ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, p. 10.

Ademais, corrobora-se tal pensamento a partir da jurisprudência estabelecida pelo Superior Tribunal de Justiça (STJ) com a súmula 550:

Súmula 550. A utilização de escore de crédito, método estatístico de avaliação de risco que não constitui banco de dados, dispensa o consentimento do consumidor, que terá o direito de solicitar esclarecimentos sobre as informações pessoais valoradas e as fontes dos dados considerados no respectivo cálculo.³⁵⁸

Por fim, para o autor, o STJ reconhece, no julgamento do caso que originou esta súmula, o direito à explicação de decisões automatizadas com impacto significativo na vida dos indivíduos. Ressalta que o tribunal afirmou o interesse do consumidor em acessar fontes de dados e a lógica do tratamento, desde que tenha sido negativamente impactado ao buscar crédito — com a ressalva de se respeitar o segredo de empresa — e permitiu o uso de dados pessoais sem consentimento para análise de risco de crédito, contanto que sejam observados os fatores limitadores e garantidos os direitos do consumidor, como o direito à explicação.³⁵⁹

Prossegue-se, portanto, para a análise dos elementos da LGPD. A discussão em torno do direito à explicação no Brasil converge com a previamente abordada na União Europeia nos seguintes pontos: (a) controvérsia sobre o direito à explicação no Brasil ocorre em paralelo ao debate na União Europeia, sendo amplamente influenciada por fundamentos teóricos e construções hermenêuticas que destacam as particularidades e desafios da LGPD. (b) uma interpretação sistemática da lei permite a derivação de novos direitos com base na lógica regulatória embora não haja menção expressa ao termo “explicação” e derivados na LGPD; e (c) persistem incertezas quanto à existência efetiva de um direito à explicação na LGPD brasileira.³⁶⁰

O debate remonta a tempos anteriores à promulgação da lei e perdurou após a lei ser publicada, em uma nítida ausência de consenso entre o Poder Executivo e o Poder Legislativo em como o art. 20 — o principal dispositivo para a discussão das decisões automatizadas no cenário brasileiro, assim como o art. 22 era o centro gravitacional no RGPD — deveria estar previsto. A doutrina identifica, ao menos, quatro redações legislativas distintas. A primeira e a original, oriunda do texto aprovado no ano de 2018, previa expressamente a revisão da decisão por uma “pessoa natural”. Contudo, o Presidente da República, Michel Temer, à época, editou a Medida

³⁵⁸ BRASIL. Superior Tribunal de Justiça. **Súmula 550**. Julgado em: 14/10/2015. Data de Publicação: 19/10/2015.

³⁵⁹ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, p. 10.

³⁶⁰ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, pp. 183-184.

Provisória 869/2018 com o objetivo (não único) de suprimir a participação humana expressa no art. 20. Em uma clara disputa, o Congresso Nacional emendou a MP para incluir, novamente, o termo “pessoa natural”. No fim, durante o processo conversão da medida provisória, o veto presidencial, em 2019, pelo presidente Jair Bolsonaro, retirou, de forma definitiva, a previsão legal da intervenção humana no procedimento de revisão de uma decisão automatizada³⁶¹, como pode ser visualizado no esquema comparativo a seguir:

Figura 6 - Quadro comparativo das redações legislativas do art. 20 da LGPD

Texto original 2018	MP 869	Texto Aprovado Congresso	Aprovado com vetos
Art. 20. O titular dos dados tem direito a solicitar revisão, por pessoa natural, de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, inclusive de decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.	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.	Adição: § 3º A revisão de que trata o caput deste artigo deverá ser realizada por pessoa natural, conforme previsto em regulamentação da autoridade nacional, que levará em consideração a natureza e o porte da entidade ou o volume de operações de tratamento de dados.	VETADO o § 3º.

Fonte: SOUZA, C. A.; PERRONE, C.; MAGRANI, E (2021, p. 395)

Em perspectiva comparada, afirma-se que a LGPD e o RGPD se baseiam em uma abordagem sistêmica de proteção à pessoa humana, com a garantia de um rol de direitos ao titular de dados. Essa é uma das grandes similaridades entre as leis, contudo, quando se trata da questão das decisões automatizadas, as distinções são evidentes. No contexto europeu, as decisões automatizadas são proibidas, como cláusula geral, e somente podem ser tomadas em três casos excepcionais.³⁶² Em contrapartida, no cenário brasileiro, caso um processo decisório automatizado trate dados pessoais, instantaneamente estará sob o escopo de aplicação da LGPD,

³⁶¹.SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021. pp. 394-395.

³⁶² SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021. pp. 395-396.

e, assim, por meio do art. 20, a lei brasileira “cria um amplo direito de revisão”³⁶³, em que o controlador tem como obrigação estabelecer um sistema revisional sob duas circunstâncias cumulativas: (a) decisões totalmente automatizadas e (b) afetem os interesses do titular³⁶⁴, como pode ser visualizado na redação legislativa do dispositivo:

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.³⁶⁵ (grifo nosso)

Para Souza, Perrone e Magrani, existem três elementos fundantes para defender a existência de um direito à explicação na LGPD: (a) o princípio da transparência³⁶⁶, ao estabelecer a obrigatoriedade do controlado conceder informações para reduzir a assimetria de poder informacional entre os agentes econômicos públicos e privados; (b) o direito de acesso a essas informações prestadas; e (c) a explicação é um pressuposto básico para o exercício de outros direitos, especialmente o de revisão das decisões automatizadas, pois sem esclarecimentos acerca dos critérios determinantes para o processo decisório alcançar determinado resultado, não é possível contestá-los, revisá-los e, eventualmente corrigi-los por erros, vieses e discriminação algorítmica.³⁶⁷ Os autores fundamentam, assim como fizeram Selbst e Powles com o GDPR, que

³⁶³ SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021. p. 396.

³⁶⁴ SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021, p. 396.

³⁶⁵ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

³⁶⁶ Art. 6º (...) 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;

³⁶⁷ SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021, pp. 396-398.

os arts. 9º, incisos I e II³⁶⁸, art. 18, inciso I e art. 20³⁶⁹, § 1.º da lei de proteção de dados brasileira “criam uma teia de direitos e obrigações que devem facilitar o acesso à informação”.³⁷⁰

Quanto às limitações e o conteúdo a estar contido nessas informações, principalmente no que tange à natureza jurídica da informação, a LGPD deixa em aberto para o entendimento posterior de entendimentos jurisprudenciais e da Autoridade Nacional de Proteção de Dados, de forma similar ao RGPD. Nesse sentido, para os autores, à medida que o RGPD traz expressões como “informação útil” sobre a “lógica subjacente” e a “importância e as consequências previstas”, a LGPD refere-se a “informações claras e adequadas” e “critérios e procedimentos utilizados”, indicando paralelos entre ambas as normas. Infere-se que o termo “informação útil” no GDPR possui o intuito de assegurar outras faculdades concedidas ao titular, similar ao propósito da “informação adequada” na LGPD. No entanto, o GDPR não especifica claramente as informações que devem ser fornecidas pelo controlador, assim como na LGPD.³⁷¹

Analogamente, Ana Frazão compreende que, por meio da redação legislativa do art. 20, é possível inferir um rol de direitos correlacionados ao princípio do devido processo legal diante de processos decisórios automatizados: (i) direito de acesso, e por conseguinte (ii) direito à informação acerca dos parâmetros do julgamento algorítmico; (iii) direito de contestação da decisão; (iv) direito de peticionamento de auditoria perante a ANPD.³⁷² Em sentido semelhante, com a ressalva de que um direito à explicação — com complexa exequibilidade — no contexto de automação decisória não deve ser fonte de confiança exacerbada para o cumprimento de fundamentos da LGPD sem a devida participação humana nos julgamentos algorítmicos a fim de torná-los mais diversificados por meio de métodos qualitativos ausentes em máquinas e presentes

³⁶⁸ Art. 9º O titular tem direito ao acesso facilitado às informações sobre o tratamento de seus dados, que deverão ser disponibilizadas de forma clara, adequada e ostensiva acerca de, entre outras características previstas em regulamentação para o atendimento do princípio do livre acesso: I - finalidade específica do tratamento; II - forma e duração do tratamento, observados os segredos comercial e industrial;

³⁶⁹ Art. 18. O titular dos dados pessoais tem direito a obter do controlador, em relação aos dados do titular por ele tratados, a qualquer momento e mediante requisição: I - confirmação da existência de tratamento;

³⁷⁰ SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021, p. 397.

³⁷¹ SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021, p. 398.

³⁷² FRAZÃO, Ana. O direito à explicação e à oposição diante de decisões totalmente automatizadas. **Jota**, 05 dez. 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-05122018?non-beta=1>

somente na capacidade humana³⁷³, Ana Frazão compreende, assim como o terceiro argumento de Souza, Perrone e Magrani, que o direito à explicação permite o exercício de outros:

Com efeito, **corolário lógico do direito de revisão de decisões automatizadas é o direito à explicação**, até porque o titular de dados não tem como contestar algo que ele não conhece ou não entende. Daí por que a compreensão das decisões algorítmicas é pressuposto do exercício de diversos direitos previstos pela LGPD, dentre os quais o da autodeterminação, o da transparência, o da não discriminação abusiva ou ilegal. Ademais, a explicabilidade de decisões algorítmicas é fundamental para assegurar o próprio devido processo legal³⁷⁴. (grifo nosso)

Desse modo, infere-se, a partir do exposto, que no ordenamento jurídico brasileiro, o direito à explicação decorre, portanto, mormente do princípio da transparência na LGPD, haja vista reger toda atividade de tratamento de dados pessoais (a conhecida propriedade “transversalidade”), e, portanto, pode ser usufruído em qualquer relação de titular e controlador — em uma nítida manifestação da autodeterminação informativa e do livre desenvolvimento da personalidade, fundamentos da LGPD — sem haver distinção acerca de setor da economia.³⁷⁵ Mas a origem da faculdade de se exigir explicações não se restringe à transparência, transcendendo outros institutos jurídicos. Dentre eles, destacam-se, mormente, o princípio do livre acesso (art. 6º, IV)³⁷⁶ e o princípio da não discriminação (art. 6º, IX).³⁷⁷ Em síntese, a lei de proteção de dados brasileiro carrega um robusto arcabouço instrumental e principiológico para sustentar a existência plena de um direito à explicação em decisões automatizadas.³⁷⁸

Em uma abordagem não convencional, sob a ótica da tutela coletiva do direito à proteção de dados pessoais e com base na premissa de que os direitos dos titulares são mais efetivamente assegurados em uma escala coletiva, diante da constatação que em um âmbito

³⁷³ FRAZÃO, Ana. Decisões algorítmicas e direito à explicação. *Jota*, 24 de nov. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/decisoes-algoritmicas-e-direito-a-explicacao-24112021>

³⁷⁴ FRAZÃO, Ana. Decisões algorítmicas e direito à explicação. *Jota*, 24 de nov. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/decisoes-algoritmicas-e-direito-a-explicacao-24112021>

³⁷⁵ MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: **Instituto Igarapé**, Artigo Estratégico, v. 24, dez. 2018, pp. 12-13.

³⁷⁶ Art. 6º (...) IV - **livre acesso**: garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais;

³⁷⁷ Art. 6º (...) IX - **não discriminação**: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos;

³⁷⁸ MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021, p. 183.

individual, podem não ser as ferramentas mais eficientes, Hosni e Martins trazem novas perspectivas acerca das possibilidades de um direito à explicação na LGPD.³⁷⁹

Conforme os autores, o termo “interesses” no caput do art. 20 permite que o direito seja mais abrangente do que o encontrado no RGPD, haja vista que não são requisitados “efeitos jurídicos relevantes” nem impactos nos direitos dos titulares, basta que a decisão automatizada interfira negativamente em qualquer aspecto individual. Assim, a LGPD assegura uma “tutela preventiva”, em que o titular pode prever um possível prejuízo e antecipadamente, tomar ações perante o controlador.³⁸⁰

A dimensão coletiva do direito à explicação, portanto, seria oriundo da redação legislativa do art. 22 da LGPD:

Art. 22. A defesa dos interesses e dos direitos dos titulares de dados poderá ser exercida em juízo, individual ou coletivamente, na forma do disposto na legislação pertinente, acerca dos instrumentos de tutela individual e coletiva.³⁸¹

Com a previsão explícita de que os direitos dos titulares (incluídos os direitos de informação e explicação em decisões automatizadas), em caso de processos decisórios com resultados enviesados e, por conseguinte, dotados de discriminação algorítmica nos termos já apresentados nesta monografia, ações coletivas poderiam ser utilizadas como instrumentos de contestação por meios jurisdicionais. Isto é, em casos concretos de minorias e grupos vulneráveis como mulheres e pessoas pretas sofrerem decisões que os afetem como um conjunto de titulares de dados, é possível que se unam para exercerem a garantia prevista no art. 22 da lei de proteção de dados brasileira. É possível, portanto, no âmbito da LGPD, que entidades da sociedade civil representem minorias sociais para exercerem um direito à explicação em decisões automatizadas, o que traz propriedades mais robustas e protetivas no ordenamento jurídico brasileiro, de forma diametralmente oposta às previsões unicamente individuais no RGPD.³⁸²

³⁷⁹ HOSNI, D. S. S.; MARTINS, P. B. L. Tomada de Decisão Automatizada e a Regulamentação da Proteção de Dados: Alternativas Coletivas Oferecidas Pela Lei Geral de Proteção de Dados. **Internet & Sociedade**, v. 1, n. 2, dez. 2020, p. 80.

³⁸⁰ HOSNI, D. S. S.; MARTINS, P. B. L. Tomada de Decisão Automatizada e a Regulamentação da Proteção de Dados: Alternativas Coletivas Oferecidas Pela Lei Geral de Proteção de Dados. **Internet & Sociedade**, v. 1, n. 2, dez. 2020, p. 90.

³⁸¹ BRASIL. Congresso Nacional. **Lei nº 13.709 de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

³⁸² HOSNI, D. S. S.; MARTINS, P. B. L. Tomada de Decisão Automatizada e a Regulamentação da Proteção de Dados: Alternativas Coletivas Oferecidas Pela Lei Geral de Proteção de Dados. **Internet & Sociedade**, v. 1, n. 2, dez. 2020, pp. 92-93.

Destarte, neste capítulo, buscou-se examinar a existência ou não de um direito à explicação de decisões tomadas por máquinas no ordenamento jurídico brasileiro a partir de uma leitura sistemática de legislações setoriais do setor consumerista e da Lei Geral de Proteção de Dados Pessoais. A conclusão é de que, diferentemente do cenário da União Europeia, há certo consenso doutrinário de que o direito existe no Brasil, embora não seja expresso em vocábulos exatos, sendo a grande questão como ele deverá ser assegurado em termos práticos. Nesse contexto, Caitlin Mulholland e Isabella Frajhof opinam que as previsões apresentadas (fundamentos, princípios e outros direitos) permitem reconhecer o direito à explicação, mas ainda há dúvidas sobre a implementação, conteúdo a ser fornecido pelo controlador e momento em que o titular pode pleitear esse direito. Será importante acompanhar o desenvolvimento da interpretação do direito à explicação pelas autoridades de proteção de dados pessoais brasileiras e europeias, bem como pelos tribunais. A definição sobre esses aspectos técnicos e legais ainda precisa ser amadurecida e esclarecida pelas instâncias responsáveis.³⁸³

O papel da ANPD será essencial para desenhar um quadro regulatório com segurança jurídica para agentes econômicos e corresponder às expectativas dos indivíduos e coletividades titulares dos dados pessoais. Como dissertam Ana Frazão, Angelo Carvalho e Giovanna Milanese “A ANPD tem o papel central de determinar o que pode ser considerado uma política adequada de proteção de dados e o que pode ser considerado uma avaliação idônea dos impactos e riscos para os usuários.”³⁸⁴

3.4 O DIREITO À EXPLICAÇÃO NO EU AI ACT E NOS PLs DE REGULAMENTAÇÃO DE IA NO BRASIL

De forma vanguardista, pretende-se, neste capítulo, explorar as delimitações técnicas, conceituais e implicações práticas de um possível direito à explicação no recém aprovado regulamento de inteligência artificial no ordenamento jurídico europeu e nas propostas de

³⁸³ MULHOLLAND, Caitlin; FRAJHOF, Isabella. Inteligência Artificial e a Lei Geral 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; MULHOLLAND, Caitlin (org.). **Inteligência Artificial e Direito - Ética, Regulação e Responsabilidade**. São Paulo, Thomson Reuters Brasil, 2019, pp. 288-289.

³⁸⁴ FRAZÃO, Ana; CARVALHO, Angelo Prata de; MILANEZ, Giovanna. **Curso de proteção de dados pessoais: fundamentos da LGPD**. Rio de Janeiro: Forense, 2022, p. 345.

regulação de IA no ordenamento jurídico brasileiro, no caso o PL 21/2020, o texto inicial do PL 2338/2023, PL 210/2024 e recente relatório final do PL 2338/2023 divulgado em junho de 2024.

Sob a égide do Brasil, serão utilizadas, mormente, o próprio texto legislativo das propostas de lei, contribuições da comissão de juristas e publicações da Autoridade Nacional de Proteção de Dados brasileira, primordialmente, devido à ausência de literatura acadêmica a versar sobre temática tão fluida e em atual debate legislativo — inclusive, nota-se a baixa iniciativa da comunidade científica jurídica em tratar da temática explicabilidade em comparação às produções europeias que se debruçam profundamente sobre a temática, com variedade de propostas e abordagens, como previamente apresentado.

Em contrapartida, devido ao fato de que a primeira versão de uma proposta de regulação na União Europeia ter sido apresentada logo em 2021, foi possível encontrar materiais acadêmicos que buscavam compreender como o AI Act aborda ou não um direito à explicação em sistemas de inteligência artificial. Ressalva-se, no entanto, que, ao fim de maio de 2024, fora aprovada uma versão distinta da inicial, cujas produções acadêmicas trazidas não utilizaram de marco teórico e possuem, portanto, esse recorte temporal delimitado. Assim, será feita uma abordagem comparando a redação legislativa da versão de 2021 e a recente de 2024 quanto à explicabilidade de modelos de IA e do direito à explicação em decisões automatizadas, haja vista a diferença da inclusão de um dispositivo legal específico para um “direito a explicações sobre as decisões individuais” no texto final aprovado.

3.4.1. O Direito à Explicação em Decisões Automatizadas no *EU AI ACT*

Em abril de 2021, a Comissão Europeia apresentou a proposta do Regulamento de Inteligência Artificial — denominado, oficial e popularmente como AI Act — com o objetivo de harmonizar as regras sobre IA no mercado europeu, de forma a equilibrar o respeito a direitos fundamentais e a inovação tecnológica. Durante o ano de 2022, a proposta foi discutida em várias versões, buscando criar um marco regulatório globalmente reconhecido para IA, com o Regulamento sobre a Proteção de Dados de modelo regulatório como inspiração. Em dezembro de 2023, após intensas negociações, o Conselho e o Parlamento Europeu chegaram a um acordo provisório, com um rascunho final sobre o Regulamento de IA com uma abordagem baseada em

riscos: quanto maior o risco de um sistema de IA, mais rigorosas são as regras aplicáveis.³⁸⁵ Em maio de 2024, o regulamento finalizado foi adotado pelo Conselho Europeu, tornando-o a primeira legislação mundial a estabelecer regras harmonizadas para a inteligência artificial, com entrada em vigor de determinadas disposições a partir de 2026, em uma abordagem baseada em riscos.³⁸⁶

Essa abordagem permite que o AI Act preveja critérios rigorosos para aplicações designadas como de alto risco ou risco inaceitável, e, assim, pode-se afirmar que a legislação adota uma abordagem assimétrica de regulação, com base em “uma matriz de avaliação de riscos”³⁸⁷. Em outros vocábulos, os sistemas de inteligência artificial são analisados sob a ótica do seu nível de risco, que está associado ao contexto de aplicação e ao impacto potencial na saúde, segurança ou direitos fundamentais de indivíduos ou coletividades.³⁸⁸ Nesse sentido, aplicações de risco extremamente alto são proibidas (em um nível não aceitável socialmente) por serem incompatíveis com direitos e valores fundamentais. Isso abrange técnicas de IA manipulativas, sistemas que exploram vulnerabilidades e *social scoring*. Já as IAs de alto risco se referem a aplicações em setores sensíveis que impactam diretamente em garantias fundamentais de pessoas afetadas como acesso à justiça, serviços públicos, oportunidades de trabalho e educação.³⁸⁹

Com efeito, para os fins da presente monografia, convém análise minuciosa dos considerados “sistemas de IA de alto risco”, haja vista a relação direta destes com um direito à explicação. Assim, os sistemas de IA de “risco elevado” são previstos no Capítulo III, especificamente no art. 6º, nos seguintes termos:

1. Independentemente de a colocação no mercado ou a colocação em serviço de um sistema de IA ser feita separadamente dos produtos a que se referem as alíneas a) e b), esse sistema de IA é **considerado de risco elevado sempre que se estejam preenchidas ambas as seguintes condições:**

³⁸⁵ CONSELHO EUROPEU. **Cronologia — Inteligência artificial**. 21 de mai. de 2024. Disponível em: <https://www.consilium.europa.eu/pt/policies/artificial-intelligence/timeline-artificial-intelligence/>

³⁸⁶ CONSELHO EUROPEU. **Regulamento Inteligência Artificial (IA): Conselho dá luz verde final às primeiras regras do mundo em matéria de IA**. 21 de mai. de 2024. Disponível em: <https://www.consilium.europa.eu/pt/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>

³⁸⁷ KIRA, Beatriz. COUTINHO, Diogo R. Regulamentação da inteligência artificial na Europa e reflexões para o Brasil. **Jota**, 20 de dez. de 2023. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/regulamentacao-da-inteligencia-artificial-na-europa-e-reflexoes-para-o-brasil-20122023?non-beta=1>

³⁸⁸ KIRA, Beatriz. COUTINHO, Diogo R. Regulamentação da inteligência artificial na Europa e reflexões para o Brasil. **Jota**, 20 de dez. de 2023. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/regulamentacao-da-inteligencia-artificial-na-europa-e-reflexoes-para-o-brasil-20122023?non-beta=1>

- a) O sistema de IA destina-se a ser utilizado como um componente de segurança de um produto ou o sistema de IA é, ele próprio, um produto abrangido pelos atos enumerados **na lista da legislação de harmonização da União constante do anexo I**;
- b) O produto cujo componente de segurança nos termos da alínea a) é o sistema de IA, ou o próprio sistema de IA enquanto produto, tem de ser sujeito a uma avaliação da conformidade por terceiros com vista à sua colocação no mercado ou colocação em serviço nos termos dos atos enumerados na lista da legislação de harmonização da União constante do anexo I.
2. Além dos sistemas de IA de risco elevado a que se refere o n.º 1, **os sistemas de IA a que se refere o anexo III são também considerados de risco elevado.**³⁹⁰ (grifo nosso)

Ademais, a listagem de quais são os sistemas considerados de risco elevado pelo AI Act são encontrados no Anexo I e Anexo III do documento, como pode ser visualizado nos grifos acima. Quanto ao Anexo I, refere-se à disposição do art. 6º (1), alínea “a” de quais são as legislações já existentes na União Europeia com as quais o AI Act deve ser harmonizado, principalmente àquelas referentes à mecanismos e disposições de segurança em diversos setores da economia. Já o Anexo III prevê a listagem de aplicações consideradas de alto risco:

1. **Dados biométricos**, na medida em que a sua utilização seja permitida ao abrigo do direito da União ou do direito nacional aplicável (...)
2. **Infraestruturas críticas**: Sistemas de IA concebidos para serem utilizados como componentes de segurança na gestão e no controlo de infraestruturas digitais críticas, do trânsito rodoviário ou das redes de abastecimento de água, gás, aquecimento ou eletricidade (...)
3. **Educação e formação profissional** (...)
4. **Emprego**, gestão de trabalhadores e acesso ao emprego por conta própria (...)
5. **Acesso a serviços privados essenciais e a serviços e prestações públicos** essenciais, bem como o usufruto dos mesmos (...)
6. **Aplicação da lei**, na medida em que a sua utilização seja permitida nos termos do direito da União ou do direito nacional aplicável (...)
7. **Gestão da migração, do asilo e do controlo das fronteiras**, na medida em que a sua utilização seja permitida nos termos do direito da União ou do direito nacional aplicável (...)
8. **Administração da justiça** e processos democráticos (...) ³⁹¹ (grifo nosso)

No texto inicial de 2021 do AI Act, o termo “explicação” e derivados como “explicabilidade” ou “explicável” é citado tão somente no Considerando 38 — e não se encontra mais nenhuma menção expressa em toda proposição — cujo objetivo é explicitar determinadas salvaguardas no cenário de usos de inteligência artificial em atividades policiais, com a seguinte redação legislativa:

³⁹⁰ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

³⁹¹ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

As ações das autoridades policiais que implicam certas utilizações dos sistemas de IA são caracterizadas por um grau substancial de desequilíbrio de poder e podem conduzir à vigilância, detenção ou privação da liberdade de uma pessoa singular, bem como ter outros impactos adversos nos direitos fundamentais garantidos pela Carta. Em particular, se não for treinado com dados de alta qualidade, não cumprir os requisitos adequados em termos de exatidão ou solidez ou não tiver sido devidamente concebido e testado antes de ser colocado no mercado ou em serviço, o sistema de IA pode destacar pessoas de uma forma discriminatória ou incorreta e injusta. Além disso, o exercício de importantes direitos fundamentais processuais, como **o direito à ação e a um tribunal imparcial, a presunção de inocência e o direito de defesa, pode ser prejudicado, em particular, se esses sistemas de IA não forem suficientemente transparentes, explicáveis e documentados.** Como tal, é apropriado classificar como de risco elevado um conjunto de sistemas de IA concebidos para serem utilizados no contexto da manutenção da ordem pública, no qual a exatidão, a fiabilidade e a transparência são particularmente importantes para evitar impactos adversos, reter a confiança do público e assegurar a responsabilidade e vias de recurso eficazes.³⁹²

Nesse contexto, ressaltam-se duas produções acadêmicas que buscaram avaliar sistematicamente o papel da IA explicável (“*Explainable AI*” ou “*XAI*”) sob o escopo de aplicação da versão inicial do AI Act. O primeiro — *The role of explainable AI in the context of the AI Act* — fora escrito em sua totalidade por membros da Comissão Europeia, em uma abordagem multissetorial e plural de nacionalidades, com contribuidores de Espanha, Bélgica e Itália. O conjunto de autores concluiu que o AI Act, em termos de linguagem, utiliza expressões comumente abordadas na literatura de IA explicável, como “transparência”, “opacidade” e “compreensão”, com a previsão de que os sistemas desenvolvidos pela União Europeia sejam capazes de serem suficientemente claros ao ponto de usuários interpretarem os resultados gerados, o que, em última instância, permite inferir que metodologias de *Explainable AI* e de “transparência” por design são identificadas ao longo do AI Act. Os autores sugerem que essa discussão, como hipótese, que essa discussão tenha se originado daquela previamente abordada nesta monografia: se existe ou não um direito à explicação previsto no RGPD.³⁹³ Para isso, destaca-se o art. 52, em que os usuários do sistema devem ser notificados quando determinado conteúdo gerado artificialmente se assemelhar substancialmente a pessoas, locais ou eventos reais, podendo induzir ao erro, e também ao interagir diretamente com um sistema de IA, isto é,

³⁹² EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

³⁹³ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FAcT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, p. 2.

estarem conscientes da ação de estar contato com uma máquina, a menos que isso seja evidente pelo contexto.³⁹⁴

1. Os fornecedores devem assegurar que **os sistemas de IA destinados a interagir com pessoas singulares sejam concebidos e desenvolvidos de maneira que as pessoas singulares sejam informadas de que estão a interagir com um sistema de IA**, salvo se tal se revelar óbvio dadas as circunstâncias e o contexto de utilização. Esta obrigação não se aplica a sistemas de IA legalmente autorizados para detetar, prevenir, investigar e reprimir infrações penais, salvo se esses sistemas estiverem disponíveis ao público para denunciar uma infração penal.³⁹⁵ (grifo nosso)

Os autores, por conseguinte, argumentam que o AI Act aborda a questão da opacidade em sistemas de risco elevado com um enfoque direcionado a dois elementos requisitos, primordialmente: (a) a transparência e (b) a supervisão humana. O primeiro pode ser encontrado no art. 13 e no Considerando 47, à medida que o segundo encontra-se no art. 14 e seu respectivo Considerando de suporte, o 48.³⁹⁶

O art. 13, referente à “*transparência e prestação de informações aos utilizadores*” prevê que sejam especificadas informações como a finalidade, características, capacidades, limitações, níveis de exatidão, solidez e cibersegurança, desempenho e, oportunamente, os inputs e dados de treinamento, nos seguintes termos, em uma espécie de direito à informações significativas:

1. Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de maneira que assegure que **o seu funcionamento seja suficientemente transparente para permitir aos utilizadores interpretar o resultado do sistema e utilizá-lo corretamente**. Deve ser garantido um tipo e um grau adequado de transparência, que permita cumprir as obrigações que incumbem ao utilizador e ao fornecedor por força do capítulo 3 do presente título.
2. Os sistemas de IA de risco elevado devem ser acompanhados de instruções de utilização, num formato digital ou outro adequado, que **incluam informações concisas, completas, corretas e claras que sejam pertinentes, acessíveis e compreensíveis para os utilizadores**.

Nesse sentido, o art. 13(1) estipula que sistemas de IA de risco elevado devem ser concebidos e desenvolvidos com transparência suficiente para que os usuários possam interpretar e utilizar corretamente seus resultados. O grau de transparência deve garantir o cumprimento das obrigações pertinentes tanto do usuário quanto do fornecedor. Em sentido similar, o art. 13(2) e

³⁹⁴ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FAcT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, p. 3.

³⁹⁵ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

³⁹⁶ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FAcT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, pp. 6-7.

(3) dispõe que o sistema de inteligência artificial deve ser acompanhado por instruções compostas de informações precisas, completas, corretas e claras (direcionadas ao usuário).³⁹⁷ Em complemento, o Considerando 47 prevê:

Para fazer face à **opacidade que pode tornar determinados sistemas de IA incompreensíveis ou demasiado complexos para as pessoas singulares, os sistemas de IA de risco elevado devem observar um certo grau de transparência**. Os utilizadores devem ser capazes de interpretar o resultado do sistema e utilizá-lo de forma adequada. Como tal, os sistemas de IA de risco elevado devem ser acompanhados de documentação pertinente e instruções de utilização e incluir informações concisas e claras, nomeadamente informações relativas a possíveis riscos para os direitos fundamentais e de discriminação, se for caso disso.³⁹⁸ (grifo nosso).

A fim de assegurar que os usuários possam utilizar adequadamente o sistema, o foco do AI Act está na disponibilização de documentação relevante e informações ao usuário (por exemplo, manual de uso), ao invés de impor modelos transparentes por design específicos ou o uso obrigatório de ferramentas *XAI*, e assim, considera-se que os agentes econômicos são livres para desenvolver seus modelos usando uma abordagem transparente por *design* ou técnicas *XAI*. A documentação e as informações fornecidas ao usuário são principalmente extraídas da documentação técnica, outro requisito para sistemas de IA de alto risco estabelecido pelo AI Act para fins de certificação e vigilância de mercado, previsto no art. 11 (o qual estabelece os critérios técnicos a serem disponibilizados nesses documentos).³⁹⁹

Analogamente, quanto ao aspecto da supervisão humana, o art. 14 estabelece que os sistemas de IA devem ser projetados para permitir uma supervisão eficaz por indivíduos durante o uso. Aqueles responsáveis pela supervisão devem compreender integralmente as capacidades e limitações do sistema de IA de alto risco, monitorar sua operação adequadamente, estar cientes da tendência de confiar excessivamente nos resultados (viés de automação) e interpretar corretamente os resultados do sistema, considerando suas características e ferramentas disponíveis⁴⁰⁰, como pode ser visto:

1. Os **sistemas de IA de risco elevado** devem ser concebidos e desenvolvidos de tal modo, incluindo com ferramentas de **interface homem-máquina apropriadas, que**

³⁹⁷ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, p. 6.

³⁹⁸ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

³⁹⁹ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, p. 6.

⁴⁰⁰ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FACCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, p. 6.

possam ser eficazmente supervisionados por pessoas singulares durante o período de utilização do sistema de IA.

2. **A supervisão humana deve procurar prevenir ou minimizar os riscos para a saúde, a segurança ou os direitos fundamentais que possam surgir quando um sistema de IA de risco elevado é usado** em conformidade com a sua finalidade prevista ou em condições de utilização indevida razoavelmente previsíveis, em especial quando esses riscos persistem apesar da aplicação de outros requisitos estabelecidos neste capítulo.⁴⁰¹ (grifo nosso).

O considerando 48 especifica que as medidas de supervisão humana devem assegurar que o sistema responda ao operador humano e que os supervisores tenham a competência, formação e autoridade necessárias para essa função, a ver:

Os sistemas de IA de risco elevado devem ser concebidos e desenvolvidos de maneira que permita a sua supervisão por pessoas singulares. Para o efeito, o fornecedor do sistema deve identificar medidas de supervisão humana adequadas antes da colocação no mercado ou da colocação em serviço do sistema. Em particular, se for caso disso, essas medidas devem garantir que o sistema integre restrições operacionais que não possam ser neutralizadas pelo próprio sistema e que respondam ao operador humano e que as pessoas singulares a quem foi atribuída a supervisão humana tenham as competências, a formação e a autoridade necessárias para desempenhar essa função.⁴⁰² (grifo nosso)

A conclusão final do trabalho é a de que o AI Act, embora apresente linguagem técnica comum à comunidade de inteligência artificial explicável, não impede a aplicação, no ordenamento jurídico europeu, de sistemas caixa preta, muito menos prevê que os agentes responsáveis pelo desenvolvimento e implementação dos modelos algorítmicos, obrigatoriamente, nem de forma recomendatória, utilizem ferramentas de *XAI*.⁴⁰³

O segundo artigo — intitulado como *A Survey on Methods and Metrics for the Assessment of Explainability Under the Proposed AI Act* — aborda os mesmos dispositivos legais apresentados, contudo, traz novas nomenclaturas *user-empowering* (empoderamento do usuário, em tradução livre) e *compliance-oriented* (orientado por conformidade, em tradução livre).⁴⁰⁴ Isso significa, em outros termos, que as propostas do AI Act para mitigar riscos de opacidade, a fim de tornar sistemas mais compreensíveis, interpretáveis e transparentes, são compostas de

⁴⁰¹ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

⁴⁰² EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

⁴⁰³ PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FAcCT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023, p. 2.

⁴⁰⁴ SOVRANO, Francesco; SAPIENZA, Salvatore; PALMIRANI, Monica; VITALI, Fabio. A Survey on Methods and Metrics for the Assessment of Explainability Under the Proposed AI Act. **Arxiv**, 21 de out. de 2021, p. 4.

salvaguardas direcionadas ao entendimento do usuário e devem ser implementadas em um contexto de compliance corporativo.

Ademais, os autores também preveem que o AI Act, ao estabelecer métricas de explicabilidade deve ser, no mínimo: i) orientada ao risco, ii) neutra ao modelo, iii) alinhada ao propósito, e iv) compreensível e acessível. Em síntese, a métrica orientada ao risco deve avaliar a extensão em que as explicações permitem uma análise dos riscos aos direitos e liberdades fundamentais dos indivíduos afetados. Neutra ao modelo significa que a métrica deve ser aplicável a todos os sistemas de IA regulamentados pelo AI Act. Alinhada ao propósito indica que a métrica deve ser flexível para as diversas necessidades dos destinatários das explicações. Compreensível e acessível implica que, se as informações sobre as métricas não forem acessíveis ou reproduzíveis, os destinatários enfrentarão incerteza, contrariando o princípio de minimização de riscos.⁴⁰⁵

Desse modo, pode-se identificar a evidente omissão do texto inicial do AI Act em prever explicações sobre sistemas (ao menos em circunstâncias delimitadas, em um escopo de aplicação recortado). Embora a Comissão Europeia tenha confiado que os requisitos de transparência e rastreabilidade garantiriam uma reparação efetiva para os afetados, reconhecendo que violações de direitos fundamentais poderiam ocorrer, a decisão de omitir um direito à explicação na proposta inicial do AI Act direciona a se perguntar: o que mudou entre 2021 e 2024? A Comissão parecia acreditar que as obrigações de testagem *ex ante*, gestão de riscos e supervisão humana poderiam minimizar o risco de decisões assistidas por sistemas de inteligência artificial que prejudicasse a esfera individual e coletiva de pessoas, principalmente em áreas críticas de acesso a direitos fundamentais.⁴⁰⁶ No entanto, a versão aprovada e que passará a ser vigente não considera tais aspectos suficientes para mitigar os riscos de sistemas em setores sensíveis da sociedade.

De maneira inovadora, e, silenciosamente discutida acerca de suas possíveis implicações⁴⁰⁷, a versão final de 2024, retificada e aprovada, prevê expressamente um direito à explicação em decisões automatizadas no caso de sistemas de risco elevado — em clara oposição

⁴⁰⁵ SOVRANO, Francesco; SAPIENZA, Salvatore; PALMIRANI, Monica; VITALI, Fabio. A Survey on Methods and Metrics for the Assessment of Explainability Under the Proposed AI Act. *Arxiv*, 21 de out. de 2021, pp. 5-6;

⁴⁰⁶ KELDE, Kätliin. On the Relative Importance of the AI Act Right to Explanation. *The Digital Constitutionalist*, 2 de mai. de 2024. Disponível em: <https://digi-con.org/on-the-relative-importance-of-the-ai-act-right-to-explanation/>

⁴⁰⁷ KELDE, Kätliin. On the Relative Importance of the AI Act Right to Explanation. *The Digital Constitutionalist*, 2 de mai. de 2024. Disponível em: <https://digi-con.org/on-the-relative-importance-of-the-ai-act-right-to-explanation/>

a sua versão nascitura, a palavra “explicação” e derivadas aparecem mais de uma dezena de vezes ao longo de todo o texto final. O art. 86 versa sobre “direito a explicações sobre as decisões individuais, da seguinte maneira:

1. **Qualquer pessoa afetada sujeita a uma decisão tomada pelo responsável pela implantação com base nos resultados de um sistema de IA de risco elevado enumerado no anexo III, com exceção dos sistemas enumerados no ponto 2 desse anexo, e que produza efeitos jurídicos ou analogamente afete num grau significativo essa pessoa, de forma que considere ter repercussões negativas na sua saúde, segurança ou direitos fundamentais, tem o direito de obter do responsável pela implantação explicações claras e pertinentes sobre o papel do sistema de IA no processo de tomada de decisão e sobre os principais elementos da decisão tomada.**
2. O n.º 1 não se aplica à utilização de sistemas de IA para os quais as exceções ou restrições à obrigação prevista nesse número decorram do direito da União ou do direito nacional em conformidade com o direito da União.
3. O presente artigo só é aplicável na medida em que o direito a que se refere o n.º 1 não esteja estipulado em contrário no direito da União.⁴⁰⁸ (grifo nosso)

Pode-se, inferir, a partir da redação legislativa, acentuadas limitações ao direito: (a) o sistema de IA precisa ser considerado de risco elevado; (b) deve produzir impactos jurídicos significativos na esfera individual da pessoa afetada pela decisão; (c) esses efeitos devem possuir efeitos negativos na saúde, segurança ou outros direitos fundamentais inerentes ao indivíduo; (d) o direito de Estados-membros pode permitir outras restrições ao direito, inclusive, quando a lei nacional dispor o contrário do previsto no art. 86. Como tais aceções serão interpretadas dependerá dos próximos passos de implementação e fiscalização do cumprimento do AI Act.

Dessa maneira, embora tenha se discutido, exaustivamente, acerca de um direito à explicação em decisões automatizadas tomadas por máquinas no Regulamento sobre a Proteção de Dados, o AI Act incrementa uma camada adicional de segurança jurídica ao finalmente findar, no mínimo, o debate da previsão legal expressa desse instituto jurídico no ordenamento da União Europeia. O direito, agora, existe nesses contextos (necessariamente com inteligência artificial envolvida no processamento), sem as ambiguidades do paradigma de proteção de dados. Contudo, é fundamental compreender que fora amplamente limitado de ser exercido pelas pessoas afetadas pela decisão conforme as disposições previstas no art. 86 (1), (2) e (3).

Kätliin Kelde, em texto publicado intitulado como *On the Relative Importance of the AI Act Right to Explanation*, entende que, não obstante as limitações consideráveis, o direito à explicação previsto no AI Act é de grande valia ao permitir que indivíduos afetados obtenham

⁴⁰⁸ EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

informações claras sobre decisões de IA sem necessitar de litígios, em uma base essencial para a proteção de outros direitos, especialmente em decisões administrativas (como no caso do INSS apresentado no início da presente monografia) e na otimização de serviços como gestão de trabalhadores e educação, não abrangidos pelo RGPD. Embora possa ser raramente utilizado futuramente devido às possibilidades de exercício somente serem nas aplicações de alto risco, para a autora, a garantia constitui uma medida crucial para acessar informações indispensáveis e assegurar um recurso jurídico eficaz.⁴⁰⁹

Desse modo, a partir de um estudo minucioso das previsões normativas do novo regulamento de inteligência artificial na União Europeia, inferiu-se que o AI Act não somente prevê, com dadas disposições limitadas de exercício, um direito à explicação em decisões automatizadas em sistemas de IA de risco elevado, como estabelece um regime centrado na pessoa humana e em medidas de transparência interligadas à *compliance*. Convém, portanto, examinar como o ordenamento jurídico brasileiro interpreta tais questões, ressaltando as semelhanças e diferenças do exposto neste subtópico.

3.4.2 O Direito à Explicação em Decisões Automatizadas nos PL nº 21/2020, 210/2024 e 2338/2023

No Brasil, a temática regulatória sobre a disciplina de temas relacionados à inteligência artificial fora proposta no Congresso Nacional, pela primeira vez, em 2020. Desde então, podem ser encontradas mais de uma dezena de propostas legislativas que visam normatizar uniformemente em escopo nacional a matéria, seja de forma ampla, seja de forma setORIZADA. Resultou-se, hodiernamente, em uma regulação policêntrica de IA, centrada em riscos, cuja abordagem envolve um esforço conjunto democrático que distribui responsabilidades entre diversos atores sociais, incluindo empresas e sociedade civil, a fim de mitigar riscos aos direitos fundamentais, em uma perspectiva de valorização da pessoa humana e regulação híbrida (autoridade central e autoridades setoriais).⁴¹⁰ Entretanto, é relevante retornar alguns passos no passado e compreender como chegou-se até aqui.

⁴⁰⁹ KELDE, Kätliin. On the Relative Importance of the AI Act Right to Explanation. **The Digital Constitutionalist**, 2 de mai. de 2024. Disponível em: <https://digi-con.org/on-the-relative-importance-of-the-ai-act-right-to-explanation/>

⁴¹⁰ BIONI, Bruno; RIELLI, Mariana; ZANATTA, Rafael A.F. Regulação de IA no Brasil: onde estávamos, onde estamos e onde podemos estar: **Data Privacy Brasil Research**, 30 de abr. de 2024. Disponível em: <https://www.dataprivacybr.org/regulacao-de-ia-no-brasil-onde-estavamos-onde-estamos-e-onde-podemos-estar/>

Em 2020, o Brasil se deparou com a primeira iniciativa de regulamentar a questão da inteligência artificial. O projeto de lei nº 21 de 2020 de autoria do Deputado Eduardo Bismarck e de relatoria da Dep. Luiza Canziani — PL 21/20 — estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências, conforme a própria proposta⁴¹¹, no entanto, faz-o de forma genérica, excessivamente principiológica, de forma pouco técnica e aprofundada nas nuances que a contemporaneidade exige de uma regulamentação para a IA. Com somente 10 artigos, ao invés de visar estabelecer um regime de segurança jurídica, o projeto regride quanto à garantia de direitos previstos na forma de recomendações ou conceituadas de forma acentuadamente ampla, notadamente por serem “desprovidas de força normativa”⁴¹². Para Laura Schertel, trata-se de “uma crise de identidade que aparenta rondar o PL nº 21/2020: queremos aprovar uma carta de intenções e princípios genérica ou uma regulação robusta sobre segurança e mitigação de riscos no desenvolvimento da inteligência artificial”?⁴¹³

Desse modo, por ser um documento curto e enxuto, inclusive para fins de análise, não se pretende destrinchá-lo, haja vista que o ponto central que convém, neste tópico, é a ausência explícita de qualquer previsão acerca de um direito à explicação em decisões automatizadas, diante da constatação que nenhum desses termos, nem sinônimos, nem outros derivados são citados em toda a redação legislativa do PL 21/20.

Nesse cenário de incertezas e ausência de robustez na proposta, em 2022, fora constituída no Senado Federal uma Comissão de Juristas com o fim de subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos PLs nº 5.051, de 2019, 21, de 2020, e 872, de 2021 (CJSUBIA), sob a Presidência do Ministro Ricardo Villas Bôas Cueva.⁴¹⁴ No relatório final da Comissão, apresentou-se um substitutivo preliminar, que visava estabelecer um novo

⁴¹¹ BRASIL, Congresso Nacional. **Projeto de Lei nº 21 de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/151547>.

⁴¹² MENDES, Laura Schertel. Projeto de Lei da Inteligência Artificial: armadilhas à vista. **O Globo**. 26 de novembro de 2021. Disponível em: <https://blogs.oglobo.globo.com/fumus-boni-iuris/post/laura-schertel-mendes-pl-da-inteligencia-artificial-armadilhas-vista.html>.

⁴¹³ MENDES, Laura Schertel. Projeto de Lei da Inteligência Artificial: armadilhas à vista. **O Globo**. 26 de novembro de 2021. Disponível em: <https://blogs.oglobo.globo.com/fumus-boni-iuris/post/laura-schertel-mendes-pl-da-inteligencia-artificial-armadilhas-vista.html>.

⁴¹⁴ SENADO FEDERAL. **Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial**. Disponível em: <https://legis.senado.leg.br/comissoes/comissao?codcol=2504>

marco legal capaz de conciliar a proteção à pessoa humana — o elo vulnerável da relação — segurança jurídica, sob a premissa de que não é necessário realizar um *trade off* entre assegurar garantias fundamentais e a inovação tecnológica em torno da ordem econômica. O objetivo, portanto, era criar um protótipo com potencial conciliatório: abordagem regulatória baseada em riscos, mesclada a de direitos⁴¹⁵. Nesse sentido, os juristas esclarecem:

Além de fixar direitos básicos e transversais para todo e qualquer contexto em que há interação entre máquina e ser humano (e.g., de **informação e transparência**), intensifica-se tal carga obrigacional quando o **sistema de IA produz efeitos jurídicos relevantes ou impactem os sujeitos de maneira significativa (e.g., direito de contestação e intervenção humana)**. Busca-se assegurar contraditório e ampla defesa – devido processo informacional – a quem possa ter direitos e **liberdades afetadas por uma decisão totalmente ou parcialmente automatizada**.⁴¹⁶ (grifo nosso)

O substitutivo formulado pelos especialistas tornou, posteriormente, o texto inicial do Projeto de Lei nº 2338, de 2023, de relatório do Senador Rodrigo Pacheco — PL 2338/23. Com efeito, ainda na exposição de motivos, encontram-se menções a direitos correlatos a um de explicação, com redação legislativa similar à escrita do RGPD e do AI Act, diante da expressão de “efeitos jurídicos relevantes” ou “impacto significativo” como gatilho de exercício dessas garantias. No segundo capítulo deste substitutivo, prevê-se, de forma ampla e geral, os direitos das pessoas afetadas por sistemas de inteligência artificial:

Art. 5º Pessoas afetadas por sistemas de inteligência artificial têm os seguintes direitos, a serem exercidos na forma e nas condições descritas neste Capítulo:

- I – **direito à informação prévia** quanto às suas interações com sistemas de inteligência artificial;
- II – **direito a explicação sobre a decisão, recomendação ou previsão** tomada por sistemas de inteligência artificial;
- III – **direito de contestar decisões ou previsões** de sistemas de inteligência artificial **que produzam efeitos jurídicos ou que impactem de maneira significativa os interesses** do afetado.⁴¹⁷ (grifo nosso)

Infere-se que existem três direitos previstos, separada e independentemente: (i) o direito à informação *ex ante* (informações sistêmicas); (ii) direito à explicação de decisões, recomendações e previsões e; (iii) direito à contestação/oposição de decisões ou previsões com efeitos jurídicos e impactos significativos. Nota-se como a limitação de impactos jurídicos

⁴¹⁵ COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022, pp. 9-10.

⁴¹⁶ COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022, p. 10

⁴¹⁷ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

relevantes somente é prevista no último direito, o que pavimentou-se distintamente do ordenamento europeu, o qual impunha esse limite, principalmente, no contexto de explicações.

Na Seção II do capítulo II (dos direitos) do substitutivo, apresenta-se, no art. 7º, trecho destinado a explicitar quais são as informações necessárias referentes ao primeiro direito estabelecido, nomeadamente “dos direitos associados à informação e compreensão das decisões tomadas por sistemas de inteligência artificial”, em que se inclui: (i) o fato da interação ser automatizado; (ii) descrições gerais do sistema e consequências possíveis de serem geradas; (iii) identificação dos operadores e as medidas de governança adotadas; (iv) distinção de papéis desempenhados pela máquina e por humanos no processo decisório; (v) quais dados pessoais da pessoa foram utilizados; (vi) medidas de segurança, não discriminação e elementos de confiabilidade (acurácia, precisão) e; (vi) outras informações a serem dispostas em regulamento no futuro.⁴¹⁸ Além disso, o § 3º desse dispositivo prevê, expressamente, a compreensibilidade dos parâmetros gerais desses sistemas no caso de grupos vulneráveis:

Os sistemas de inteligência artificial que se destinem a **grupos vulneráveis, tais como crianças, adolescentes, idosos e pessoas com deficiência**, serão desenvolvidos de tal modo que **essas pessoas consigam entender o seu funcionamento e seus direitos** em face dos agentes de inteligência artificial.⁴¹⁹ (grifo nosso)

Ainda na Seção II, o art. 8º do substitutivo busca tratar, explicitamente, de um direito à explicação de decisões (recomendações e previsões incluídas), e, de forma vanguardista, estabelece quais são os parâmetros a serem observados nessas explicações, incluindo informações sobre meios e procedimentos de se buscar outros direitos, como o de contestação e intervenção humana (nos incisos IV e V, respectivamente), de forma a correlacionar esse rol de garantias, com o direito à explicação de pressuposto para que os outros sejam efetivamente exercidos:

Art. 8º A pessoa afetada por sistema de inteligência artificial poderá solicitar **explicação sobre a decisão, previsão ou recomendação, com informações a respeito dos critérios e dos procedimentos utilizados**, assim como sobre os principais fatores que afetam tal previsão ou decisão específica, incluindo informações sobre:

I – **a racionalidade e a lógica do sistema**, bem como o significado e as consequências previstas de tal decisão para a pessoa afetada;

II – **o grau e o nível de contribuição do sistema** de inteligência artificial para a tomada de decisões;

III – **os dados processados e a sua fonte, bem como os critérios** para a tomada de decisão e, quando apropriado, a sua ponderação, aplicados à situação da pessoa afetada;

IV – **os mecanismos por meio dos quais a pessoa pode contestar a decisão**; e

⁴¹⁸ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁴¹⁹ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

V – a possibilidade de solicitar intervenção humana, nos termos desta lei.⁴²⁰ (grifo nosso)

Assim, pode-se afirmar que o substitutivo prevê tanto um direito à explicação *ex ante* acerca do funcionamento do sistema de IA, quanto um direito à explicação *ex post* da lógica envolvida no processo decisório — ambos autônomos de outras garantias — com determinados critérios a serem encontrados nessas informações, inclusive, de linguagem e de tempo (prazo) previstos no parágrafo único do art. 8º:

As informações mencionadas no caput serão fornecidas por **procedimento gratuito e facilitado, em linguagem que permita que a pessoa compreenda o resultado da decisão ou previsão em questão**, no prazo de até quinze dias a contar da solicitação, permitida a prorrogação, uma vez, por igual período, a depender da complexidade do caso.⁴²¹ (grifo nosso)

Conclui-se, a partir do exposto, que, diferentemente do AI Act da União Europeia, essa primeira versão de uma proposta de regulamentação da inteligência artificial no Brasil não impôs o direito à explicação de decisões automatizadas somente para sistemas de IA de risco elevado, muito menos limitou aos casos de efeitos jurídicos relevantes como na lei europeia, inclusive quando se inclui na análise o Regulamento sobre a Proteção de Dados.

A Coalizão Direitos Na Rede, organização da sociedade civil que reúne mais de cinquenta entidades de direitos digitais e governança da internet, divulgou nota técnica para avaliar o PL 2338/23 e recomendar alterações na redação legislativa.⁴²² No que tange a questão dos direitos relacionados à transparência, dissertam sobre a essencialidade da legislação refletir sobre a relevância da transparência não apenas para compreender o funcionamento técnico de um sistema, mas também outras questões relacionadas ao seu desenvolvimento e aplicação. A legalidade dos sistemas de IA pode ser avaliada mesmo sem identificar exatamente como chegaram a um resultado. Diferentes informações sobre um sistema de IA interessam a diversos

⁴²⁰ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁴²¹ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁴²² PICOLO, Cynthia; BUARQUE, Gabriela; PEREIRA, José Renato. **Nota Técnica sobre o PL n.º 2338/23. Coalizão Direitos na Rede**, 23 de ago. de 2023. Disponível em: <https://direitosnarede.org.br/2023/08/23/coalizao-direitos-na-rede-divulga-nota-tecnica-sobre-o-pl-2338-2023-que-bu-sca-regular-a-ia/>.

públicos, como cientistas, organizações da sociedade civil, indivíduos tentando entender decisões algorítmicas e reguladores interessados em práticas de coleta de dados.⁴²³

Em complemento, a Autoridade Nacional de Proteção de Dados (ANPD) brasileira, elaborou uma Análise Preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial, com o intuito primordial de traçar paralelos entre a proposição legislativa e disposições da vigente LGPD. Na avaliação da Autoridade, o PL 2338/2023 interage significativamente com a Lei Geral de Proteção de Dados Pessoais, em três eixos principais, sendo o primeiro deles o mais relevante para o presente objeto de estudo, embora os outros tenham a sua devida relevância: (i) na tutela de direitos; (ii) na classificação de sistemas de IA de alto risco e; (iii) nos mecanismos de governança. Ademais, a ANPD argumenta que deveria atuar como entidade reguladora central nesta matéria, defendendo um posicionamento de disciplinar tanto a proteção de dados, quanto a inteligência artificial.⁴²⁴

Quanto ao primeiro eixo, a autoridade explicitamente menciona o direito à explicação como um direito relacionado à matéria disciplinada e prevista na LGPD:

Dentre os direitos dispostos no PL, chamam atenção os **direitos à informação, à explicação**, e à contestação e de solicitar revisão. Trata-se de direitos que se destinam à tutela das pessoas naturais afetadas por sistemas de IA e que possuem similaridades com alguns dispositivos da LGPD relativos aos direitos dos titulares.⁴²⁵ (grifo nosso)

Ademais, em pronunciamento sólido e, a primeira vez que a Autoridade cita em documento oficial o direito à explicação, o qual não é consenso de que exista ou não na LGPD (como previamente abordado em tópico para tal), diante do processo de tomada de subsídios sobre norma dos direitos do titulares ainda estar em processo de consolidação e ser necessário aguardar o documento final. Em específico, citam:

Percebe-se, ainda, **notória convergência entre o dever imposto no artigo 20, § 1º, da LGPD, aos controladores no contexto de tomada de decisão automatizada, e o**

⁴²³ PICOLO, Cynthia; BUARQUE, Gabriela; PEREIRA, José Renato. **Nota Técnica sobre o PL n.º 2338/23. Coalizão Direitos na Rede**, 23 de ago. de 2023, p. 22. Disponível em: <https://direitosnarede.org.br/2023/08/23/coalizacao-direitos-na-rede-divulga-nota-tecnica-sobre-o-pl-2338-2023-que-bu-sca-regular-a-ia/>.

⁴²⁴ BRASIL, Autoridade Nacional de Proteção de Dados. **Análise Preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial**. Brasília, 6 de jul. de 2023, p. 1. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial>

⁴²⁵ BRASIL, Autoridade Nacional de Proteção de Dados. **Análise Preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial**. Brasília, 6 de jul. de 2023, p. 3. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial>

direito à explicação sobre a decisão, previsão ou recomendação, com informações a respeito dos critérios e dos procedimentos utilizados, assim como sobre os principais fatores que afetam tal previsão ou decisão específica. **Desse modo, a ANPD possui papel relevante na regulação do direito à explicação.**⁴²⁶

Portanto, infere-se que a ANPD reconhece expressamente, ainda que de forma rudimentar e inicial, um direito à explicação em decisões automatizadas no art. 20, § 1º. Trata-se de um posicionamento de extrema relevância da entidade reguladora, principalmente diante do papel desempenhado na construção da interpretação e aplicação dos dispositivos da LGPD e do poder de *enforcement* que exerce sobre empresas e entes da Administração Pública.

Nessa senda, ressalta-se que relatório final da Comissão de Juristas, junto à redação do substitutivo que viera a ser o texto preliminar de nova propositura de lei, foi fruto de uma série de de audiências públicas e um seminário internacional, ouvindo mais de setenta especialistas no tema, incluindo representantes da sociedade civil organizada, governo, academia e setor privado. Além disso, abriu espaço para a participação de quaisquer interessados por meio de contribuições escritas, recebendo 102 manifestações que foram analisadas individualmente e organizadas conforme suas propostas.⁴²⁷ Convém, portanto, trazer pontos favoráveis e contrários a existir ou não o direito à explicação de decisões automatizadas na discussão legislativa brasileira do Painel 8 “Atributos do design sócio-técnico de confiabilidade da IA: segurança, acurácia, transparência, rastreabilidade e monitoramento” e Painel 9 “Direitos e deveres: transparência e explicabilidade; revisão e o direito à intervenção humana; correção de vieses”⁴²⁸ das audiências públicas que ensejaram como resultado o texto minuciosamente investigado previamente.

Em posicionamento comum, Renato Leite Monteiro, Tainá Junquillo e Ana Paula Bialer esclarecem que o direito à explicação não se deve confundir com uma abertura total da “caixa preta” nem ter acesso direto aos códigos-fonte ou todas as camadas de processamento de algoritmos e softwares de empresas de tecnologia, confusão comumente ocorrida quando se trata de explicabilidade de sistemas decisórios automatizados. Em verdade, Monteiro e Junquillo

⁴²⁶ BRASIL, Autoridade Nacional de Proteção de Dados. **Análise Preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial.** Brasília, 6 de jul. de 2023, p.4. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial>.

⁴²⁷ BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023.** Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁴²⁸ COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil.** Brasília, 2022, p. 40.

colocam a explicabilidade como um princípio relacionado ao indivíduo, a pessoa humana como fator determinante: é ela quem compreende e deve estabelecer os limites.⁴²⁹

De forma a equilibrar o debate, Alexandre Pacheco da Silva, Dora Kaufman, Caroline Tauk, Edson Prestes e Bruno Miragem ponderam acerca dos limites das explicações, principalmente no que se refere ao segredo de negócios (industrial e comercial), destacando os formatos que a transparência pode assumir, principalmente diante de quem a explicação vai ser direcionada: público geral e leigo ou o público técnico.⁴³⁰ Além disso, fora trazida a questão da opacidade inerente aos modelos de redes neurais e a dificuldade de se estabelecer formas de auditar tais tecnologias.⁴³¹

Em contrapartida, Nina da Hora, Diego Machado e Jamila Venturi trazem a questão de documentação para cumprimento do requisito da explicabilidade, com visões distintas. Para Da Hora, os documentos sempre são constantemente dirigidos a públicos específicos, ressaltando que se costuma afirmar que explicação para cada setor significa uma coisa diferente, mas, na verdade, esquece-se que a sociedade como um todo também merece explicações para além de “documentação”. Machado destaca a importância de registrar etapas do desenvolvimento do sistemas, desde o momento da criação até a execução. Venturi traz a ideia dos relatórios de riscos como mecanismos eficazes de mitigação.⁴³²

Segue-se, portanto, para o projeto de Lei nº 210 de 2024 de relatoria do Senador Marcos Pontes — PL 210/24 — o qual fora proposto em um contexto de discussão do PL 2338/23, veio com o intuito de ser uma possibilidade distinta da apresentada pela proposição anterior. Nesse sentido, de forma diametralmente oposta à versão inicial do PL 2338/23, o direito à explicação não é previsto de forma separada e autônoma do direito à informação significativa e ambos estão com redações legislativas intimamente interligadas, misturados em um só dispositivo legal com uma série de parâmetros de documentação. Trata-se, em verdade, de uma grande mistura entre princípios de transparência e *accountability* (principalmente ao se referir a documentações) e os direitos de informação e explicação em sistemas automatizados, a ver:

⁴²⁹ COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022, p. 108

⁴³⁰ COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022, p. 109-111

⁴³¹ COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022, p. 113.

⁴³² COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022, p. 112-113

Art. 7º Sempre que usar um sistema com tecnologia de inteligência artificial, o indivíduo tem **direito a receber informações completas sobre seu funcionamento, a fim de compreender os resultados que o impactem.**

§ 1º Os sistemas disponibilizarão documentação em linguagem simples, incluindo descrições claras sobre seu funcionamento geral e do papel que a automação desempenha, aviso de que os sistemas estão em uso, a denominação do responsável pelo sistema e **explicações claras, oportunas e acessíveis sobre os resultados gerados.**⁴³³ (grifo nosso)

Sob a ótica do mesmo dispositivo legal, também são previstos critérios do formato da explicações que devem ser: tecnicamente válidas, significativas e úteis:

§ 2º As notificações serão mantidas atualizadas.

§ 3º Os indivíduos afetados pelo sistema serão alertados sobre alterações significativas nas funcionalidades.

§ 4º O sistema informará como seus resultados impactam o indivíduo.

§ 5º Os sistemas automatizados fornecerão **explicações que sejam tecnicamente válidas, significativas e úteis para qualquer indivíduo, operador ou pessoa** que precise compreender o sistema.

§ 6º É obrigatório o fornecimento de relatórios que incluam informações resumidas sobre os sistemas em linguagem simples.

§ 7º Os **resultados sobre avaliações da clareza e da qualidade da notificação e das explicações** serão tornados públicos sempre que possível.⁴³⁴

Assim, com redação legislativa distinta, o PL 210/24 mantém a lógica do texto inicial do PL 2338 de não limitar um direito à explicação em decisões automatizadas somente em casos de efeitos jurídicos relevantes ou em aplicações de IA consideradas de alto risco.

Por fim, em um cenário recente, nos meses de abril e junho, novas versões do relatório foram apresentadas, e, em julho de 2024, foi apresentada a complementação do relatório final do PL 2338/23 (a versão mais recente até o momento desta pesquisa) pelo novo relator, o Senador Eduardo Gomes. O texto de abril seguiu lógica de abordagem de riscos da versão inicial do PL 2338/23, contudo a alteração de maior relevância se trata da retirada da taxonomia explícita dos sistemas considerados de alto risco, transferindo a responsabilidade de delimitar hipóteses de alto risco somente quando houver ao SIA (Sistema Nacional de Regulação e Governança de Inteligência Artificial), conforme determinados critérios descritos no art. 15 do novo texto.⁴³⁵A

⁴³³ BRASIL, Congresso Nacional. **Projeto de Lei nº 210 de 2024**. Dispõe sobre os princípios para uso da tecnologia de inteligência artificial no Brasil. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/161980>

⁴³⁴ BRASIL, Congresso Nacional. **Projeto de Lei nº 210 de 2024**. Dispõe sobre os princípios para uso da tecnologia de inteligência artificial no Brasil. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/161980>

⁴³⁵ FRAZÃO, Ana. Marco da inteligência artificial e o substitutivo do relator. **Jota**, 15 de mai. de 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/marco-da-inteligencia-artificial-e-o-substitutivo-do-relator-15052024?non-beta=1>.

versão inicial do PL 2338/23, proposta pela Comissão de Juristas, estabelecia, de forma semelhante ao AI Act, uma lista de setores e aplicações de IA consideradas de alto risco, em seu art. 17. Dentre elas, destacam-se: infraestruturas críticas, saúde, educação, acesso a oportunidades de trabalho, acesso a serviços essenciais públicos e privados, concessão de crédito, administração da justiça, veículos autônomos, biometria, investigação criminal, segurança pública e gestão de migração e controle de fronteiras.⁴³⁶ Na versão de abril, o PL 2338/23 transferia, portanto, a definição de altos riscos para o SIA, adotando uma abordagem procedimental em vez de material⁴³⁷, como pode ser visualizado no art. 15 do novo texto:

Caberá ao SIA regulamentar a classificação da lista dos sistemas de inteligência artificial de alto risco, bem como identificar novas hipóteses, levando em consideração a probabilidade e a gravidade dos impactos adversos sobre pessoa ou grupos afetados, e com base em, pelo menos, um dos seguintes critérios: (...)⁴³⁸

Na opinião de Ana Frazão, o parágrafo único do art. 63 dessa versão determina que a regulamentação da lista de alto risco será feita pelo SIA após consulta pública, refletindo a definição de risco como um conceito científico, social, político e cultural. Essa abordagem procedimental, em vez de material, considera os usos antes listados como de alto risco apenas como critério, resolvendo problemas da taxonomia anterior e criando um regime mais flexível, adaptável e aberto à participação pública e à *accountability*.⁴³⁹ No entanto, em contramão a esse entendimento, os relatórios publicados em junho e julho trouxeram de volta às hipóteses de risco ao texto, novamente, engessando a regulamentação e criando novas barreiras para a aprovação para setores específicos.⁴⁴⁰

A polêmica entre como definir a abordagem de riscos entre a versão inicial e final é de suma importância para os fins desta monografia, pois outra grande mudança aconteceu da primeira versão até a versão atual (jun. de 2024) do PL 2338/23: o direito à explicação em

⁴³⁶ BRASIL, Senado Federal. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁴³⁷ FRAZÃO, Ana. Marco da inteligência artificial e o substitutivo do relator. **Jota**, 15 de mai. de 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/marco-da-inteligencia-artificial-e-o-substitutivo-do-relator-15052024?non-beta=1>.

⁴³⁸ BRASIL. Congresso Nacional. **Relatório Final do Substitutivo ao Projeto de Lei nº 2338/2023, junho de 2024**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

⁴³⁹ FRAZÃO, Ana. Marco da inteligência artificial e o substitutivo do relator. **Jota**, 15 de mai. de 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/marco-da-inteligencia-artificial-e-o-substitutivo-do-relator-15052024?non-beta=1>.

⁴⁴⁰ BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline.

decisões automatizadas tomadas por máquinas, antes consideravelmente ilimitado, agora, passa a ser bem mais restrito de ser exercido somente mediante duas possibilidades: quando (a) a decisão for produzida por sistemas de alto risco ou (b) produza efeitos jurídicos relevantes. Poderá ser uma, ou outra, conforme a redação estabelece conjunção de alternância.⁴⁴¹

Além disso, o texto traz uma diferenciação intencional e expressa de quais são os direitos assegurados em todas as aplicações de inteligência artificial, e quais são em circunstâncias específicas, haja vista o Capítulo II “Dos Direitos” ser dividido entre a Seção I “Dos Direitos da Pessoa e do Grupo Afetados por Sistema de IA” e a Seção II “Dos Direitos da Pessoa e Grupos Afetados por Sistema de IA que Produz Efeitos Jurídicos Relevantes ou de Alto Risco”.⁴⁴² Na Seção I, somente é assegurado, em todos os sistemas de IA, o direito à informação prévia, nos seguintes termos:

Art. 5º A pessoa ou o grupo afetado por sistema de inteligência artificial, independentemente do seu grau de risco, têm os seguintes direitos, a serem exercidos na forma e nas condições descritas neste Capítulo:

I – **direito à informação prévia quanto às suas interações com sistemas de inteligência artificial**, de forma acessível, gratuita e de fácil compreensão, inclusive sobre caráter automatizado da interação, exceto nos casos em que se trate de sistemas de IA dedicados única e exclusivamente à cibersegurança e à ciberdefesa; (grifo nosso)⁴⁴³

Na Seção II, o direito à explicação, em sua nova configuração mais limitada e específica, apresenta-se da seguinte maneira:

Art. 6º A pessoa ou o grupo afetado por sistema de IA que **produza efeitos jurídicos relevantes ou de alto risco** tem os seguintes direitos:

I – **direito à explicação sobre a decisão, recomendação ou previsão feitas pelo sistema**;

Parágrafo único. A explicação solicitada no âmbito do inciso I, **respeitando o segredo comercial e industrial**, incluirá informações suficientes, adequadas e inteligíveis, nos termos do Regulamento.⁴⁴⁴ (grifo nosso)

⁴⁴¹ BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline.

⁴⁴² BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline.

⁴⁴³ BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline.

⁴⁴⁴ BRASIL, Senado Federal. **Complementação de Voto PL 2338/2023**. Brasília, 4 de jul. de 2024. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=9683716&ts=1720545993533&rendition_principal=S&disposition=inline.

Por conseguinte, a redação do relatório de complementação de voto do PL 2338/23 se diferencia profundamente da versão inicial no que se refere às especificidades de um direito à explicação em decisões automatizadas. Ele continua a existir, mas não tão livremente, nem mais tão facilmente exercido. Agora, torna-se extremamente limitado, principalmente com a redação do parágrafo único do art. 6º que prevê o segredo de empresa como elemento de embate direto no momento de assegurar a garantia nas restritas situações em que poderá ser invocada. Como o *tradeoff* entre o segredo comercial e industrial vs explicabilidade decisória será realizado, para que lado irá pesar, dependerá de como a regulamentação seja aprovada e posteriormente interpretada pelo SIA (caso a sua existência e previsão legal perdure em uma eventual promulgação) e pela ANPD. Ainda é cedo para assumir o relatório de complementação de voto como texto definitivo, haja vista ter um longo caminho a ser percorrido de discussões de diversos pontos da proposta e a tramitação ainda estar na sua casa de origem (Senado) e com um mar de possibilidades de alterações quando passar a tramitar na Câmara dos Deputados.

Buscou-se, destarte, examinar minuciosamente as principais propostas de regulamentação de inteligência artificial no Brasil, com especial enfoque para a compreensão das facetas de um direito à explicação em decisões automatizadas se mostrava em cada uma. Não se pode afirmar uma trajetória linear, pois as propostas de lei, em momentos temporais distintos, avançavam em conceder maior possibilidades para a garantia, e, depois, voltam passos para trás a fim de limitá-lo, à semelhança do AI Act do ordenamento jurídico europeu. Atualmente, pode-se inferir que a atual proposição brasileira, na temática explicabilidade, aproximou-se da regulação recém-aprovada na União Europeia, possivelmente, diretamente influenciada.

CONSIDERAÇÕES FINAIS

A presente monografia, buscou, portanto, examinar, minuciosamente, três dimensões de um direito à explicação em decisões automatizadas: (i) a dimensão técnica, referente à viabilidade de explicações serem ou não possíveis de serem geradas, inferidas e concedidas às pessoas afetadas pelos processo decisórios, com especial enfoque aos limites da explicabilidade de sistemas de inteligência artificial de redes neurais, devido à sua capacidade e acurácia os tornem os modelos mais utilizados para tomadas de decisões (em suma, a exequibilidade do direito); (ii) a dimensão existencial com argumentação principiológica, constitucional e teórica dessa garantia, a partir do estudo de institutos jurídicos como o devido processo legal (e a sua formatação no mundo digital

no devido processo informacional), o livre desenvolvimento da personalidade e a autodeterminação informativa em conjunto dos fenômenos da contemporaneidade moldada pelo capitalismo de vigilância e pela sociedade algorítmica e; (iii) a dimensão legislativa por meio da análise pormenorizada dos principais dispositivos legais que regem a explicação em processos decisórios, por meio da leitura sistemática do RGPD, da LGPD, do AI Act na União Europeia e dos PLs de regulamentação de IA no Brasil.

Conclui-se, a partir do exposto no primeiro capítulo, que máquinas têm o poder e a capacidade de decidir, e, no mínimo, são ferramentas de auxílio à tomada de decisões que influenciam direta e fortemente no resultado final. Nesse sentido, as explicações emergem como formas de compreensão desses processos decisórios. Elas devem ser cognoscíveis e transbordam a concepção tradicional de transparência e não devem ser padronizadas, estáticas e conceitualmente pré-definidas. Em verdade, devem ser dinâmicas, flexíveis e variar conforme contexto e indivíduos e grupos afetados pelas decisões, principalmente no que tange à aplicações de IA em setores sensíveis da sociedade, como àqueles que impactam diretamente em garantias fundamentais dos cidadãos, como acesso à justiça, a oportunidades de trabalho, a concessão de benefícios sociais e tantos outros citados nas classificações de risco elevado dos modelos de IA nos projetos de regulação.

Desse modo, a participação humana em processos decisórios (interação humano-máquina) é essencial para permitir que as explicações possam ter as propriedades apresentadas. Inferiu-se, também no primeiro capítulo, que a explicabilidade é um pressuposto, não uma utopia, intrinsecamente associada às noções de transparência e *accountability* de máquinas. Se não é possível assegurá-la, ao menos em termos mínimos (como explicações *ex ante* ou explicações *ex post* coletivas), deve-se repensar se o sistema deve continuar a tomar decisões que interfiram na esfera coletiva e individual de pessoas. Uma aplicação de IA relacionada ao acesso a garantias fundamentais incapaz de gerar uma explicação ao menos ampla, geral e minimamente útil para a efetivação de outros direitos não segue os padrões éticos e regulatórios impostos aos modelos de inteligência artificial. Nessa senda, a explicabilidade e a interpretabilidade, embora não sejam consensos técnicos, devem ser almejados e buscados pela comunidade acadêmica e por desenvolvedores e aplicadores da tecnologia.

Para além da conclusão de ser possível assegurar com limitações inerentes à tecnologia, o direito à explicação, como conclusão desta monografia, deve existir como instrumento de defesa

do indivíduo. Por conseguinte, o devido processo legal como princípio constitucional consolidado, o livre desenvolvimento da personalidade e autodeterminação informativa, como fundamentos do paradigma de proteção de dados, e por consequência, da ótica regulatória de inteligência artificial, ensejam tal direito como manifestação prática delimitada desses institutos principiológicos. Em um vácuo da legislação brasileira, é fundamental buscar fontes para além dos textos de lei. Assim, um direito à explicação é um mecanismo — não único nem suficiente de maneira isolada — para combater práticas abusivas de vigilância em massa e o perfilamento eletrônico são identificados como ameaças significativas à liberdade individual típicas de uma contemporaneidade moldada pelos fenômenos do capitalismo de vigilância. A personalidade, portanto, não pode se desenvolver livre e devidamente em ambientes com tais propriedades sem ferramentas efetivas de caráter individual e coletivo — como o direito à explicação — de compreender que de forma decisões, previsões e recomendações são feitas para moldar o presente e o futuro de pessoas e como corrigi-las quando são ilicitamente discriminatórias, enviesadas e precipitadas.

Por fim, sob a égide comparada, chegou-se à conclusão que o ordenamento jurídico brasileiro e europeu possuem evidentes semelhanças quanto aos modelos regulatórios para disciplinar o direito à explicação em decisões automatizadas. Mas também se distinguem essencialmente, devido à natureza jurídica dos dispositivos do RGPD e da LGPD não serem idênticas. Na União Europeia, não se tem consenso se a lei de proteção de dados prevê o direito, sendo um verdadeiro embate acadêmico com três correntes principais: (a) não existe um direito à explicação no RGPD (ii) existe um direito à explicação no RGPD a partir de uma interpretação sistemática) e (iii) existir expressamente não é relevante sem considerar aspectos práticos de compliance e *accountability* com ferramentas de *Privacy By Design*, relatórios de impacto, e possibilidade de auditorias. Com o advento do AI Act, é possível afirmar que, agora, o direito existe — embora consideravelmente limitado — no contexto de automação algorítmica de sistemas de IA na União Europeia: somente em hipóteses de aplicação de alto risco que produzam efeitos jurídicos de maneira a haver repercussões negativas na sua saúde, segurança ou direitos fundamentais, desde que não exista previsão contrária no direito interno dos países membros.

No Brasil, a doutrina se estabelece de maneira diametralmente oposta às contraposições explícitas da UE, pois se posiciona de modo consensual ao concordar em uma abordagem

sistêmica para inferir a existência do direito à explicação a partir do disposto no art. 20 e art. 20, parágrafo 1º da LGPD, com o especial destaque à inovação brasileira de se criar um instrumento de tutela coletiva para violações de direito da LGPD no seu art. Nesse sentido, ressaltou-se que a própria Autoridade Nacional de Proteção de Dados reconheceu, em documento público, que tal direito existe no ordenamento de proteção de dados brasileiro ao citá-lo expressamente como um dos direitos do titular. Por fim, em sentido bastante similar ao contexto europeu, os PLs de regulamentação de IA no Brasil incrementam uma camada adicional para além da LGPD ao prever expressamente o direito à explicação em hipóteses de alto risco, com menos limitações em comparação às previsões do AI ACT.

Destarte, reconhece-se que um direito à explicação não é uma solução milagrosa, muito menos solitária. Trata-se de uma garantia que compõe uma complexa rede de direitos de titulares de dados e pessoas afetadas por decisões de máquinas: a sua relevância depende de um sistema regulatório que o torne possível de ser exercido em conjunto a outras faculdades, pois permite que indivíduos prejudicados e grupos vulneráveis possam exigir de *Big Techs* e entes da Administração Pública informações que lhes são devidas. Mas isso não é suficiente. Desenvolver a personalidade livremente e se autodeterminar informativamente é muito além disso, mas, para que se permita um paradigma moldado pela proteção de dados pessoais, são necessários uma série de mecanismos, e um direito à explicação mínimo, em situações de alto risco a garantias fundamentais, é somente um desses relevantes instrumentos.

Por fim, em um panorama inicial de aplicação de leis de proteção de dados por autoridades nacionais e tribunais brasileiros e europeu e de regulamentação de inteligência artificial ainda nos primeiros passos, não se pretendeu responder como tal direito será limitado efetivamente e os limites da sua exequibilidade. Dependerá da aplicação em casos concretos de que forma o segredo de empresa poderá esbarrar nas explicações exigidas por titulares (em âmbito individual e coletivo) e de que modo as previsões que, sistematicamente, ensejam o direito à explicação no ordenamento jurídico europeu e brasileiro serão ampliadas ou restringidas pelas autoridades de proteção de dados dos Estados-membros e pela ANPD.

REFERÊNCIAS

ALI, Sajid et al. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. **Information Fusion**, v. 99, 2023.

AMARAL, Francisco. **Direito Civil: Introdução**. São Paulo: Saraiva Educação, 2018.

ANANNY, M.; CRAWFORD, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. **New Media & Society**, v. 20, n. 3, p. 973–989, 2016.

AVELINO, Rodolfo; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu da. **A Sociedade do Controle: manipulação e modulação nas redes digitais**. São Paulo: Hedra, 2018.

AWS. **Aprendizado supervisionado versus não supervisionado**. Disponível em: <https://aws.amazon.com/pt/compare/the-difference-between-machine-learning-supervised-and-un-supervised/>. Acesso em 5 de mar de 2024.

BENTES, Anna et al. **Para além da proteção de dados: uma coletânea**. São Paulo: Data Privacy Brasil Ensino, 2023.

BIDDLE, Sam; RIBEIRO, Paulo Vitor; DIAS, Tatiana. CENSURA INVISÍVEL: TikTok escondeu ‘feios’ e favelas para atrair novos usuários e censurou posts políticos. **The Intercept**, 16 de mar. de 2020. Disponível em: <https://www.intercept.com.br/2020/03/16/tiktok-censurou-rostos-feios-e-favelas-para-atrair-novos-usuarios/>

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

BIONI, Bruno. **Regulação e proteção de dados pessoais: o princípio da accountability**. Rio de Janeiro: Forense, 2022.

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

BIONI, Bruno; RIELLI, Mariana; ZANATTA, Rafael A.F. Regulação de IA no Brasil: onde estávamos, onde estamos e onde podemos estar: **Data Privacy Brasil Research**, 30 de abr. de 2024. Disponível em:

<https://www.dataprivacybr.org/regulacao-de-ia-no-brasil-onde-estavamos-onde-estamos-e-onde-podemos-estar/>

BIONI, Bruno. MARTINS, Pedro. Devido processo informacional: um salto teórico-dogmático necessário? **Jota**, 15 de jul. de 2015. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/devido-processo-informacional-um-salto-teorico-dogmatico-necessario-15072020>

BOBBIO, Norberto. **Teoria do Ordenamento Jurídico**. São Paulo: Edipro, 2 ed., 2014.

BOHLENDER, Dimitri; KÖHL, Maximilian A. Towards a Characterization of Explainable Systems. **ArXiv** [cs], [S.l.], 2019.

BOUCHER, Philip. Artificial Intelligence: how it works? Why does it matter and what we can do about it? **European Parliamentary Research Service**, 2020.. Disponível em: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU\(2020\)641547_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf).

BRAGA, Carolina. Discriminação nas decisões por algoritmos: polícia preditiva. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil Conteúdo e Tecnologia Ltda, 2019.

BRANCO, Ana Paula. Entenda a nova inteligência artificial do INSS para detectar fraude em atestados médicos. **Folha de São Paulo**, São Paulo, 15 de jan. de 2024. Disponível em: <https://www1.folha.uol.com.br/mercado/2024/01/entenda-a-nova-inteligencia-artificial-do-inss-para-detectar-fraude-em-atestados-medicos.shtml?origin>

BRASIL, Autoridade Nacional de Proteção de Dados. **Análise Preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial**. Brasília, 6 de jul. de 2023, p.4. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial>.

BRASIL. Congresso Nacional. **Constituição (1988)**. Constituição da República Federativa do Brasil. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm

BRASIL. Congresso Nacional. Lei nº 13.709 de 14 de agosto de 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

BRASIL, Congresso Nacional. **Projeto de Lei nº 21 de 2020**. Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil; e dá outras providências. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/151547>.

BRASIL. Congresso Nacional. **Projeto de Lei nº 2338, de 3 de maio de 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>

BRASIL. Congresso Nacional. **Projeto de Lei nº 210 de 2024**. Dispõe sobre os princípios para uso da tecnologia de inteligência artificial no Brasil. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/161980>

CASSELA, Vinícius; PARREIRA, Marcelo. CGU: análise automática de benefícios do INSS tem mais negativas e risco de decisões indevidas. **G1**, Brasília, 9 de nov. de 2023. Disponível em: <https://g1.globo.com/economia/noticia/2023/11/09/cgu-analise-automatica-de-beneficios-do-inss-tem-mais-negativas-e-risco-de-decisoes-indevidas.ghtml>

CASSINO, João Francisco. Modulação deleuzeana, modulação algorítmica e manipulação midiática. In: AVELINO, Rodolfo; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu da. **A Sociedade do Controle: manipulação e modulação nas redes digitais**. São Paulo: Hedra, 2018.

CASEY, B.; FARHANGI, A.; VOGL, R. Rethinking Explainable Machines: The GDPR’s “Right to Explanation” Debate and the Rise of Algorithmic Audits in Enterprise. **Berkeley Technology Law Journal**, v. 34, 2018, p.145-189.

COALIZÃO DIREITOS NA REDE. **Nota Técnica sobre o PL n.º 2338/23**. Disponível em: <https://direitosnarede.org.br/2023/08/23/coalizacao-direitos-na-rede-divulga-nota-tecnica-sobre-o-pl-2338-2023-que-busca-regular-a-ia/>.

COMISSÃO DE JURISTAS. **Relatório Final da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil**. Brasília, 2022. Disponível em: <https://legis.senado.leg.br/comissoes/mnas?codcol=2504&tp=4>

CONFESSORE, Nicholas. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. **New York Times**, New York, 4 abr. 2018. Disponível em: <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>

CORTE INTERAMERICANA DE RECECHOS HUMANOS. **Comunicado: A Colômbia é responsável internacionalmente por haver realizado atividades arbitrárias de inteligência contra defensores de direitos humanos, que também foram vítimas de atos de violência e estigmatização por parte das autoridades estatais.** 2024. Disponível em https://www.corteidh.or.cr/docs/comunicados/cp_16_2024_port.pdf

DIAKOPOULOS, Nicholas. Algorithmic accountability reporting: on the investigation of black boxes. [S. l.]: **Tow Center for Digital Journalism**, 2014.

DOMINGOS, Pedro. **O Algoritmo Mestre**. São Paulo, Novatec, 2017.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: elementos da formação da Lei Geral de Proteção de Dados**. 2. ed. Rio de Janeiro, Thomson Reuters Revista dos Tribunais, 2020.

DONEDA, Danilo; MENDES, Laura Schertel; SOUZA, Carlos Afonso Pereira de; ANDRADE, Norberto Nuno Gomes de. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar**, Fortaleza, v.23, n.4, 1-17, out./dez. 2018.

EDWARDS, L.; VEALE, M. Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. **Duke Law & Technology Review**, v. 16, p. 18-84, 2017.

EUROPEAN COMMISSION. **Article 29 Working Party**. Guidelines on Automated Individual Decision-Making and Profiling for the Purpose of Regulation 2016/679. Bruxelas, 2017.

EUROPEAN PARLIAMENT. **Artificial Intelligence Act (AI Act)**. Bruxelas, 2024. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf

FEFERBAUM, Marina; et. al. **Ética, governança e inteligência artificial**. São Paulo: Almedina, 2023.

FRAJHOF, Isabella Z. O papel dos mecanismos de compliance para operacionalização do direito à explicação de decisões totalmente automatizadas. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas (coord.). **Compliance e políticas de proteção de dados**. São Paulo: Revista dos Tribunais, 2022.

FRAZÃO, Ana. Controvérsias sobre direito à explicação e à oposição diante de decisões automatizadas. **Jota**, 12 de dez. 2018. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/controversias-sobre-direito-a-explicacao-e-a-oposicao-diante-de-decisoes-automatizadas-12122018>.

FRAZÃO, Ana. Decisões algorítmicas e direito à explicação. **Jota**, 24 de nov. 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/decisoes-algoritmicas-e-direito-a-explicacao-24112021>.

FRAZÃO, Ana. Marco da inteligência artificial e o substitutivo do relator. **Jota**, 15 de mai. de 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/marco-da-inteligencia-artificial-e-o-substitutivo-do-relator-15052024?non-beta=1>.

FRAZÃO, Ana. O ChatGPT e voz de Scarlett Johansson. **Jota**, 30 de mai. de 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/o-chatgpt-e-a-voz-de-scarlett-johansson-29052024?non-beta=1>. Acesso em: 30 de mai. de 2024.

FRAZÃO, Ana. Obstáculos para a consideração de questões éticas nos julgamentos algorítmicos. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial**, São Paulo: Almedina, 2023, p. 37.

FRAZÃO, Ana. Transparência de algoritmos x segredo de empresa. **Jota**, 9 de jun. de 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/transparencia-d-e-algoritmos-x-segredo-de-empresa-09062021>

FRAZÃO, Ana; CARVALHO, Angelo Prata de; MILANEZ, Giovanna. **Curso de proteção de dados pessoais: fundamentos da LGPD**. Rio de Janeiro: Forense, 2022.

FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil Conteúdo e Tecnologia Ltda, 2019.

GERCINA, Cristiane. INSS aumenta análise de aposentadorias por robôs e nega benefício em seis minutos. **Folha de São Paulo**, São Paulo, 31 de jul. de 2023. Disponível em:

<https://www1.folha.uol.com.br/mercado/2023/07/inss-aumenta-analise-de-aposentadorias-por-robos-e-nega-beneficio-em-seis-minutos.shtml>

GERCINA, Cristiane. Órgão que julga recursos de aposentadoria levará 16 anos para zerar fila, estima TCU. **Folha de São Paulo**, São Paulo, 30 de mar. de 2023. Disponível em: <https://www1.folha.uol.com.br/mercado/2023/03/orgao-que-julga-recursos-de-aposentadorias-levara-16-anos-para-zerar-fila-estima-tcu.shtml>

GOODMAN, Bryce; FLAXMAN, Bryce. European Union regulations on algorithmic decision-making and a “right to explanation”. **ArXiv**, 31 de aug. de 2016.

HANSSON, Jamie Heinemeier. I Applied for an Apple Card. What they offered was a sexist insult. **Fast Company**, 11 de nov. de 2019. Disponível em: <https://www.fastcompany.com/90429224/i-applied-for-an-apple-card-what-they-offered-was-a-sexist-insult>.

HILL, Kashmir. How Target Figured Out a Teen Girl Was Pregnant before Her Father Did. **Forbes**, 16 de fev. de 2012. Disponível em: <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-fat-her-did/>

HORA, Nina da. **My News Explica! Algoritmos**. São Paulo: Edições 70, 2023.

HOSNI, D. S. S.; MARTINS, P. B. L. Tomada de Decisão Automatizada e a Regulamentação da Proteção de Dados: Alternativas Coletivas Oferecidas Pela Lei Geral de Proteção de Dados. **Internet & Sociedade**, v. 1, n. 2, dez. 2020.

JUNQUILHO, Tainá. **Inteligência Artificial no Direito: limites éticos**. Salvador, Juspodivm, 2022.

KAMINSKI, Margot. The right to explanation, explained. **Berkeley Technology Law Journal**, v. 34, n. 1, May 2019.

KASPERKEVIC, Jana. Google says sorry for racist auto-tag in photo app. **The Guardian**. New York. 1 de jul. 2015. Disponível em: <https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app>.

KAUFMAN, Dora. Decisão automatizada por inteligência artificial. In: FEFERBAUM, Marina. et al (org). **Ética, Governança e Inteligência Artificial**, São Paulo: Almedina, 2023.

KAUFMAN, Dora; JUNQUILHO, Tainá; REIS, Priscila. Externalidades negativas da inteligência artificial: conflitos entre limites da técnica e dos direitos humanos. **Revista Direitos e Garantias Fundamentais**, Vitória, v. 24, n. 3, p. 43-71, set./dez. de 2023.

KELDE, Kätliin. On the Relative Importance of the AI Act Right to Explanation. **The Digital Constitutionalist**, 2 de mai. de 2024. Disponível em: <https://digi-con.org/on-the-relative-importance-of-the-ai-act-right-to-explanation/>

KIM, Tae Wan; ROUTLEDGE, Bryan. Why a right to an explanation of algorithmic decision-making should exist: a trust-based approach. **Business Ethics Quarterly**, p. 1-28, 2021.

KIRA, Beatriz; COUTINHO, Diogo R. Regulamentação da inteligência artificial na Europa e reflexões para o Brasil. **Jota**, 20 de dez. de 2023. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/regulamentacao-da-inteligencia-artificial-na-europa-e-reflexoes-para-o-brasil-20122023?non-beta=1>.

KREMER, Bianca. **Direito e tecnologia em perspectiva amefricana : autonomia, algoritmos e vieses raciais**. Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Direito, 2021.

KREMER, Bianca et al. **Racismo Algorítmico**. Rio de Janeiro: O Panóptico, 2023.

LEE, Kai-fu. **Inteligência artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos**. Rio de Janeiro: Globo Livros, 2019.

LIMA, Taisa Maria Macena; SÁ, Mária de Fátima Freire. 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, p.227-246, out./dez. 2020.

LINDOSO, Maria Cristine Branco. **Discriminação de gênero em processos decisórios automatizados**. Dissertação (Mestrado - Mestrado em Direito) - Universidade de Brasília, 2019.

LUDWIG, Marcos de Campos. O direito ao livre desenvolvimento da personalidade na Alemanha e possibilidades de sua aplicação no Direito privado brasileiro. **Revista da Faculdade de Direito da UFRGS**, v. 19, mar. de 2001, p. 261.

MAC, Ryan. Facebook apologizes after A.I. puts ‘primates’ label on videos of black men. **The New York Times**, New York, 3 de set. de 2021. Disponível em: <https://www.nytimes.com/2021/09/03/technology/facebook-ai-race-primates.html>

MACLURE, Jocelyn. AI, Explainability and Public Reason: The argument from de Limitations of the Human Mind. **Mind and Machines**, v. 31, p.1 421-438, 2021.

MAGENTA, Matheus et al. Como telefones de usuários do Facebook foram usados por campanhas em 'disparos em massa' no WhatsApp. **BBC News Brasil**, 20 out. 2018. Disponível em: <https://www.bbc.com/portuguese/brasil-45910249>.

MARQUES, Ana Luiza Pinto Coelho. **Erosão constitucional no Brasil e tecnologias disruptivas: análise a partir do processo eleitoral de 2018**. Dissertação (Pós Graduação em Direito - Mestrado Acadêmico), Belo Horizonte, 2022.

MAYBIN, Simon. Sistema de algoritmo que determina pena de condenados cria polêmica nos EUA. **BBC News Brasil**, 31 de out. 2016. Disponível em: <https://www.bbc.com/portuguese/brasil-37677421>.

MEAKER, Morgan. Drivers Are Rising Up Against Uber’s ‘Opaque’ Pay System. **Wired**, London, 23 de Mai. de 2024. Disponível em: <https://www.wired.com/story/drivers-are-rising-up-against-ubers-opaque-pay-system/>

MENDES, Gilmar; BRANCO, Paulo. **Curso de Direito Constitucional**. São Paulo: Saraiva Educação (Série IDP), 2021.

MENDES, Laura Schertel. Decisão histórica do STF reconhece direito fundamental à proteção de dados pessoais. **Jota**, Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/decisao-historica-do-stf-reconhece-direito-fundamental-a-protecao-de-dados-pessoais-10052020>.

MENDES, Laura Schertel. Projeto de Lei da Inteligência Artificial: armadilhas à vista. **O Globo**, 26 de nov. de 2021. Disponível em: <https://blogs.oglobo.globo.com/fumus-boni-iuris/post/laura-schertel-mendes-pl-da-inteligencia-artificial-armadilhas-vista.html>.

MENDES, Laura Schertel. **Série IDP - Linha de pesquisa acadêmica - Privacidade, proteção de dados e defesa do consumidor: linhas gerais de um novo direito fundamental**. 1ª Edição. São Paulo: Editora Saraiva, 2014.

MENDES, Laura Schertel; FUJIMOTO, Mônica (org.). **Painel LGPD nos Tribunais**. Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, 2024.

MENDES, Laura Schertel. Autodeterminação informativa: a história de um conceito. **Pensar**, Fortaleza, v. 25, n.4, p. 1-18, out./dez. 2020.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **RDU**, v. 16, n. 90, p. 39-64, nov.-dez. 2019.

MICHEL, François; SILVA, Carlos A. da. A ameaça aos direitos fundamentais na era da inteligência artificial. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas (orgs.). **Compliance e Políticas de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2019.

MICKLITZ, Hans. W et al. Constitutional Challenges in The Algorithmic Society. Cambridge, **Cambridge University Press**, 2022.

MIRANDA, Felipe Arady. O direito fundamental ao livre desenvolvimento da personalidade. **RIDB**, v.2, n° 10, 2013.

MOOR, James H. Are There Decisions Computers Should Never Make?. **Nature and System**, [S. l.], v. 1, 1985.

MONTEIRO, Renato Leite. **Desafios para a efetivação do direito à explicação na Lei Geral de Proteção de Dados do Brasil**. 2021. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade de São Paulo, São Paulo, 2021.

MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Rio de Janeiro: Instituto Igarapé, **Artigo Estratégico**, v. 24, dez. 2018.

MULHOLLAND, Caitlin; GOMES, Rodrigo Dias de Pinho. Inteligência Artificial e seus principais desafios para os programas de compliance e as políticas de proteção de dados. In: FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas. **Compliance e Políticas de Proteção de Dados**. São Paulo: Thomson Reuters Brasil, 2021.

O'NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. São Paulo: Editora Rua do Sabão, 2020.

OLIVEIRA, Carla. Aprendizado de máquina e modulação do comportamento humano. In: AVELINO, Rodolfo; SOUZA, Joyce; SILVEIRA, Sérgio Amadeu da. **A Sociedade do Controle: manipulação e modulação nas redes digitais**. São Paulo: Hedra, 2018.

OLIVEIRA, Carlos Elias; NETO, João Costa. Direito Civil. Rio de Janeiro: Forense, 2022.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO - OCDE. **Recommendation of the Council on Artificial Intelligence**. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

PANIGUTTI et al. The Role of Explainable AI in the context of the AI Act. **FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, pp. 1139-1150, 12 de jun. de 2023.

PARLAMENTO EUROPEU. **Regulamento Geral sobre a Proteção de Dados (RGPD)**. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679>.

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

RODOTÀ, Stefano. **A vida na sociedade da vigilância: a privacidade hoje**. Rio de Janeiro: Renovar, 2008.

ROSAS, Eduarda Moraes Chacon. **Inteligência artificial: regulação ética a partir das regras de proteção de dados pessoais**. Rio de Janeiro: Editora GZ, 2023.

RUSSELL, Stuart. **Inteligência artificial a nosso favor: como manter o controle sobre a tecnologia**. São Paulo: Companhia das Letras, 2021.

SAMEK, Wojciech; WIEGAND; MULLER, Klaus-Robert. Explainable Artificial Intelligence: understanding, visualizing and interpreting deep learning models. **Cornell University Press**, 28 de Aug. de 2017.

SELBST, Andrew; POWLES, Julia. Meaningful information and the right to explanation. **International Data Privacy Law**, v. 7, issue 4, p. 233-242, nov. 2017.

SILVA, Elisa de Macedo da. Redes Sociais, **Eleição Presidencial de 2018 e Questões de Credibilidade da Mídia**. 2019. Monografia (Pós Graduação em Comunicação) - Mestrado Acadêmico, [S. l.], 2019. Disponível em: <https://repositorio.ufjf.br/jspui/bitstream/ufjf/12040/1/elisademacedodasilva.pdf>.

SIMONCINI, Andrea; LONGO, Erik. Fundamental Rights and the Rule of Law in the Algorithmic Society. *Constitutional Challenges in the Algorithmic Society*. **Cambridge: Cambridge University Press**, 2021.

SOLOVE, Daniel J.; MATSUMI, Hideyuki. AI, Algorithms, and Awful Humans. **Fordham Law Review**, vol. 92, 6 de nov.. de 2023.

SOUZA, Carlos Affonso et al. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: DONEDA, Danilo et al. (coord.). **Tratado de proteção de dados pessoais**. Rio de Janeiro: Forense, 2021.

SOUSA, Bruno. **Panóptico: reconhecimento facial renova velhas táticas racistas de encarceramento**. Rede Observatórios de Segurança, 22 de abr. de 2022. Disponível em: <https://observatorioseguranca.com.br/panoptico-reconhecimento-facial-renova-velhas-taticas-racistas-de-encarceramento/>

SOVRANO, Francesco; SAPIENZA, Salvatore; PALMIRANI, Monica; VITALI, Fabio. A Survey on Methods and Metrics for the Assessment of Explainability Under the Proposed AI Act. **Arxiv**, 21 de out. de 2021

TEIXEIRA, Pedro S. Reconhecimento facial erra gênero de pessoa trans. São Paulo, **Folha de São Paulo**, 5 de mai. de 2024. Disponível em: <https://www1.folha.uol.com.br/tec/2024/05/reconhecimento-facial-erra-genero-de-pessoa-trans.shtml>

UNESCO. **Recomendação sobre a Ética de Inteligência Artificial**. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por

VÉLIZ, Carissa. **Privacidade é Poder: porque e como você deveria retomar o controle de seus dados**. São Paulo: Editora Contracorrente, 2021.

VINHOLES, Thiago. Caixa-preta de avião é, na verdade, laranja; entenda por que e como ela funciona. **CNN Brasil**, São Paulo, 09 de out. de 2024. <https://www.cnnbrasil.com.br/economia/caixa-preta-de-aviao-e-na-verdade-laranja-entenda-por-que-e-como-ela-funciona/>. Acesso em 9 de mar. de 2024.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision Making Does Exist in the General Data Protection Regulation. **International Data Privacy Law**, vol. 7, n. 2, pp. 76–99, mai. de 2017

WARREN, Samuel; BRANDEIS, Louis. The right to privacy. **Harvard Law Review**, v. IV, n. 5, p. 195, 1890.

WIMMER, Miriam; DONEDA, Danilo. Falhas de IA e a Intervenção Humana em Decisões Automatizadas: Parâmetros para a Legitimação pela Humanização. **Revista Direito Público**, Brasília, v. 18, n. 100, 374-406, out./dez. 2021.

WYLIE, Christopher. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. **The Guardian**, London, 17 mar. 2018. Disponível em: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

ZANATTA, Rafael A.F. A centralidade da pessoa humana entre a filosofia e a regulação. **Jota**, 6 de dez, de 2023. <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/a-centralidade-da-pessoa-humana-entre-a-filosofia-e-a-regulacao-06122023>

ZUBOFF, Shoshana. **A Era do Capitalismo de Vigilância**. Rio de Janeiro: Intrínseca, 2021.