

**UNIVERSIDADE PRESBITERIANA MACKENZIE**

MARCELO FONSECA SANTOS

DIREITO DE EXPLICAÇÃO EM INTELIGÊNCIA ARTIFICIAL

SÃO PAULO

2025

MARCELO FONSECA SANTOS

DIREITO DE EXPLICAÇÃO EM INTELIGÊNCIA ARTIFICIAL

Dissertação de Mestrado apresentada ao programa de Pós-Graduação *Stricto Sensu* em Direito Político e Econômico da Universidade Presbiteriana Mackenzie, como requisito parcial à obtenção do título de Mestre em Direito Político e Econômico.

ORIENTADOR: Prof. Dr. Felipe Chiarello de Souza Pinto

SÃO PAULO

2025

S237d Santos, Marcelo Fonseca.

Direito à explicação em inteligência artificial [recurso eletrônico]. / Marcelo Fonseca Santos.

322 KB

Dissertação (Mestrado em Direito Político e Econômico)  
Universidade Presbiteriana Mackenzie, São Paulo, 2025.

Orientador: Prof. Dr. Felipe Chiarello de Souza Pinto.

Referências bibliográficas: f. 147-161.

1. Inteligência artificial. 2. Neurociências. 3. Decisões.  
4. Explicação. 5. Explicabilidade. I. Pinto, Felipe Chiarello  
Souza, *orientador*. II. Título.

CDDir 340.0285

Bibliotecário Responsável: Luciano Francisco da Silva - CRB 8/11287

MARCELO FONSECA SANTOS

DIREITO DE EXPLICAÇÃO EM INTELIGÊNCIA ARTIFICIAL

Dissertação de Mestrado apresentada ao programa de Pós-Graduação *Stricto Sensu* em Direito Político e Econômico da Universidade Presbiteriana Mackenzie, como requisito parcial à obtenção do título de Mestre em Direito Político e Econômico.

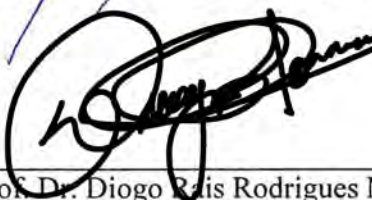
Aprovado em: 10/02/2025.

BANCA EXAMINADORA



---

Orientador: Prof. Dr. Felipe Chiarello de Souza Pinto



---

Examinador: Prof. Dr. Diogo Reis Rodrigues Moreira



---

Examinador: Prof. Dr. Irineu Francisco Barreto Júnior

## Folha de Identificação da Agência de Financiamento

**Autor:** MARCELO FONSECA SANTOS

**Programa de Pós-Graduação *Stricto Sensu* em** Direito Político e Econômico

**Título do Trabalho:** DIREITO À EXPLICAÇÃO

O presente trabalho foi realizado com o apoio de <sup>1</sup>:

- CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico
- FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo
- Instituto Presbiteriano Mackenzie/Isenção integral de Mensalidades e Taxas
- MACKPESQUISA - Fundo Mackenzie de Pesquisa
- Empresa/Indústria:
- Outro:

<sup>1</sup> **Observação:** caso tenha usufruído mais de um apoio ou benefício, selecione-os.

## AGRADECIMENTOS

Ao meu orientador Prof. Felipe Chiarello pelo carinho com que me acolheu como seu orientando, e por quem tenho imensa admiração e carinho, pois acreditou em mim e na minha pesquisa, me inspirou e despertou o estudo de neurociências, cuja contribuição para essa dissertação sem dúvida compõe um fator diferencial, e minha mais enorme gratidão.

Gratidão aos professores Diogo Rais e Irineu Barreto por suas contribuições e pela avaliação na banca, professores que admiro e que muito me ensinaram e com quem tenho a honra de ser avaliado.

À professora Juliana Abrusio Florêncio, que na escolha do tema e carinho, e certamente pelo exemplo de carreira acadêmica e profissional.

E aos meus amigos, que sempre acreditaram e me motivaram, em especial a Mauro Antonio Lauria, Alexandro Romeira e Thiago Viola pelas contribuições na ciência da computação, Stela Chaves Rocha Sales, e seu esposo Timóteo Coelho Sales, bem como aos queridos Professores Marco Aurélio Florêncio Filho, Giovani A. Saavedra, Orlando Villas Bôas Filho, Vicente Bagnoli e Luciane Cristina Petean Duarte pelos conselhos valiosos, amizade fraterna e acolhedora.

Aos amigos do Programa de Pós-Graduação *Strito Sensu* em Direito Político e Econômico da Universidade Presbiteriana Mackenzie, todos vocês estarão sempre em minha memória e história.

## **DEDICATÓRIA**

Dedico toda a pesquisa, estudo e amor à minha esposa Bruna (amor da minha vida) e meu filho Gustavo (parte do meu ser, e meu grande amor), que sempre me apoiaram e incentivaram em todos os meus projetos e sonhos acadêmicos, com paciência e confiança que tudo que faço é pensando no amor que tenho por eles (e pela Gaia, minha cachorrinha que amo demais!), vocês são a minha motivação, alegria, amor!

Aos meus pais e irmãs, que me transmitiram os valores, e importância, do estudo e da educação, do propósito, do trabalho e da perseverança, vocês são meu fundamento.

## RESUMO

A Inteligência Artificial (IA) é um campo que visa criar sistemas capazes de realizar tarefas que normalmente requerem inteligência humana, como reconhecimento de padrões, raciocínio lógico e tomada de decisão automatizada. A pesquisa analisa como os dados de forma exponencial se correlacionam com as aplicações envolvendo Inteligência Artificial. As decisões automatizadas proferidas por esses sistemas são muitas vezes denominadas e consideradas como caixas pretas, pois não fornecem uma explicação clara e compreensível sobre como chegaram às suas decisões automatizadas, as denominadas *Black Box*. A pesquisa aborda os sistemas de IA, como eles são afetados pelas decisões humanas e perfilamento de dados de comportamento, e como podem gerar problemas éticos, jurídicos, econômicos e sociais, bem como na economia de mercado. O objetivo é apresentar uma visão geral sobre a Inteligência Artificial, as decisões humanas, e como influenciam as decisões automatizadas, como regulações contemporâneas e *ex ante* podem auxiliar na concretização do Direito à Explicação em Inteligência Artificial, e melhorar sobremaneira os mecanismos de Explicabilidade. A pesquisa e dissertação utilizam metodologia baseada em leitura e análise bibliográfica e de artigos científicos, ou seja, método Hipotético-Dedutivo: “o pesquisador elege o conjunto de proposições hipotéticas que acredita serem viáveis como estratégia de abordagem para se aproximar de seu objeto”. A análise crítica permite examinar os desafios e as limitações das técnicas existentes, bem como as implicações éticas, jurídicas e sociais do Direito à Explicação em Inteligência Artificial. A pesquisa tem desafios, como: (i) Decisões Automatizadas e relação das Decisões Humanas, discutindo os benefícios e os riscos de delegar ou compartilhar dados pessoais e comportamentais com as máquinas e seus sistemas; (ii) a Neuroeconomia em virtude do Perfilamento de Dados; (iii) Regulamentações *Ex Ante* e Mecanismos Positivos; (iv) Existência e concretização do Direito à Explicação; (v) Explicabilidade em Inteligência Artificial. Dessa maneira, a pesquisa procurou responder ao problema: Existe Direito à Explicação em Inteligências Artificiais? Os capítulos foram desenvolvidos com o escopo de expor ao leitor compreensão sobre o que são, como funcionam e aprendem os sistemas de Inteligência Artificial, como humanos tomam decisões, e essas Decisões influenciam (Direitos de Personalidade) o Aprendizado e as Decisões da Inteligência Artificial. Como os seres humanos são influenciados pela neuroeconomia, e posteriormente a dissertação trata da importância das regulações *Ex Ante* como governança de dados, o *compliance* digital e a governança de inteligência artificial como mitigadores de risco, atuando na prevenção e produção de documentos que permitem maior capacidade de explicar (Direito



à Explicação) a automatização decisória, permitindo sua revisão ou confirmação, e também tratando dos sistemas específicos possam atuar na Explicabilidade da Inteligência Artificial, proporcionando maior confiança, ética e transparência para a inteligência artificial, e, também, a Inteligência Artificial Explicável (XAI).

**Palavras-chave:** Inteligência Artificial; Neurociências; Decisões; Explicação; Explicabilidade.

## ABSTRACT

Artificial Intelligence (AI) is a field that aims to create systems capable of performing tasks that typically require human intelligence, such as pattern recognition, logical reasoning, and automated decision-making. The research analyzes how data exponentially correlates with applications involving Artificial Intelligence. The automated decisions made by these systems are often referred to and considered as black boxes, as they do not provide a clear and understandable explanation of how they arrived at their automated decisions, the so-called Black Box. The research addresses AI systems, how they are affected by human decisions and behavioral data profiling, and how they can generate ethical, legal, economic, and social problems, as well as in the market economy. The objective is to present an overview of Artificial Intelligence, human decisions, and how they influence automated decisions, how contemporary and "ex ante" regulations can assist in the realization of the Right to Explanation in Artificial Intelligence, and significantly improve the mechanisms of Explainability. The research and dissertation utilize a methodology based on reading and analyzing bibliographic sources and scientific articles, that is, the Hypothetico-Deductive method: "the researcher selects the set of hypothetical propositions that they believe are viable as a strategy to approach their object." The critical analysis allows for the examination of the challenges and limitations of existing techniques, as well as the ethical, legal, and social implications of the Right to Explanation in Artificial Intelligence. The research has challenges, such as: (i) Automated Decisions and the relationship with Human Decisions, discussing the benefits and risks of delegating or sharing personal and behavioral data with machines and their systems; (ii) Neuroeconomics due to Data Profiling; (iii) "Ex Ante" Regulations and Positive Mechanisms; (iv) Existence and realization of the Right to Explanation; (v) Explainability in Artificial Intelligence. In this way, the research sought to answer the question: Is there a Right to Explanation in Artificial Intelligences? The chapters were developed with the aim of providing the reader with an understanding of what Artificial Intelligence systems are, how they function and learn, how humans make decisions, and how these decisions influence (Personality Rights) the Learning and Decisions of Artificial Intelligence. As human beings are influenced by neuroeconomics, and subsequently, the dissertation addresses the importance of "Ex Ante" regulations such as data governance, digital compliance, and artificial intelligence governance as risk mitigators, acting in the prevention and production of documents that allow for a greater capacity to explain (Right to Explanation) decision automation, enabling its review or confirmation, and also addressing specific systems

that can act on the Explainability of Artificial Intelligence, providing greater trust, ethics, and transparency for artificial intelligence, and also, Explainable Artificial Intelligence (XAI).

Keywords: Artificial Intelligence; Neurosciences; Decisions; Explanation; Explainability.

## **LISTA DE ABREVIATURAS E SIGLAS**

AI – Artificial Intelligence

API – Application Programming Interface IA – Inteligência Artificial

LGPD – Lei Geral de Proteção de Dados Pessoais

PPGDPE – Programa de Pós-graduação em Direito Político e Econômico UPM –

Universidade Presbiteriana Mackenzie

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	12
<b>1 RELAÇÃO ENTRE INTELIGÊNCIAS ARTIFICIAIS E DECISÕES HUMANAS</b> ..	18
1.1 SISTEMAS INTELIGENTES E DECISÕES .....	18
1.2 DECISÕES E NEUROECONOMIA .....	47
<b>2 MECANISMOS DE FORMAÇÃO DA DECISÃO ALGORÍTMICA</b> .....	58
2.1 PERSONALIZAÇÃO E ETICIDADE DA INTELIGÊNCIA ARTIFICIAL.....	58
2.2 RESULTADOS E DECISÕES ARTIFICIAIS .....	75
<b>3 INSTRUMENTOS DE REGULAÇÃO CONTEMPORÂNEA DA INTELIGÊNCIA ARTIFICIAL</b> .....	81
3.1 <i>COMPLIANCE</i> DIGITAL.....	81
3.2 MECANISMOS POSITIVOS .....	99
<b>4 EXPLICAÇÃO E EXPLICABILIDADE DA INTELIGÊNCIA ARTIFICIAL</b> .....	120
4.1 DIREITO À EXPLICAÇÃO .....	120
4.2 EXPLICABILIDADE – XAI .....	134
<b>CONCLUSÃO</b> .....	144
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	147

## INTRODUÇÃO

A “inovação pode ser alavanca de desenvolvimento econômico nacional se fomentada com o propósito final de estimular a economia”.<sup>1</sup>

Tratando da influência da inteligência artificial na economia, a pesquisa tem como objetivo demonstrar como nossas decisões, processadas e transformadas em dados estruturados e perfilados para inserção de sistemas automatizados precisam ser explicados quando nos deparamos com decisões automatizadas que direta ou indiretamente influenciam a sociedade.

O Direito de Explicação das Decisões proferidas por Inteligência Artificial tem como escopo a obtenção de informações sobre a tomada de decisões automatizadas expressas por sistemas de Inteligência Artificial que possam ocasionar vieses éticos, discriminação de indivíduos ou grupos, bem como manipulação de agentes de mercado decorrente da imputação e manejo das informações nos sistemas computacionais de Inteligência Artificial, utilizando perfis comportamentais, necessitando, portanto, de pesquisa sobre a influência da tomada de decisões humanas e a tomada de decisão automatizada, os meios para prestação de contas (*accountability*) e mitigação de riscos, revisão das decisões automatizadas e formas de prevenção ou resposta para os indivíduos ou grupos, por meio do *Compliance*, da Governança de Dados, Relatórios, Governança da I.A., descritas no presente trabalho, pois o uso de Inteligências Artificiais que aprendem de forma profunda, e podem conduzir e modular a oferta e demanda de produtos e serviços aos cidadãos, e que têm no Direito de Explicação e na Explicabilidade um mecanismo de salvaguarda e segurança jurídica.

Seres humanos criam informações comportamentais resultantes de suas escolhas e decisões que: conscientes ou subconscientes, rápidas ou vagarosas, tornam tais decisões como parte do seu ser, ou seja, são seus Direitos de Personalidade que constituem na Sociedade da Informação os dados utilizados e constantes do Big Data.

Designadamente decorre da Computação em Nuvem, a qual passou a possibilitar o armazenamento de volumes crescentes de dados, com disponibilidade permanente e uma fiabilidade assegurada pela redundância, tudo isso com custos cada vez menores, que se juntaram às comunicações de banda muito larga, em fibra óptica e ponto a ponto, com velocidades de acesso tais que foi desnecessário manter centros de dados próprios, também com custos decrescentes. A ambas acresceram algoritmos de análise assentes em Inteligência

---

<sup>1</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Inovação e sua Principal Barreira Jurídica. **Revista Eletrônica do Curso de Direito da UFSM (Universidade Federal de Santa Maria)**, v. 15, n. 2, p. 2, 2020. Disponível em: <https://periodicos.ufsm.br/revistadireito/article/view/41419/pdf>. Acesso em: 24 abr. 2024.

Artificial, mais do que em força computacional, ainda que distribuída, pelo menos na pendência da computação quântica, os quais vieram a acrescentar a viabilidade de gerir pacotes cada vez maiores de dados em tempo real. Finalmente, a proliferação de sensores interligados, que se dá o nome de Internet das Coisas, ou de Tudo, conduziu ao multiplicar a informação disponível respeitando sempre e definitivamente os cidadãos-consumidores.<sup>2</sup>

Tais dados e informações de personalidade produzem uma infinidade de informações objeto de cadastros, transmissão de sistemas computacionais, aplicativos e outras relações jurídicas (físicas e digitais) que estão a todo instante sendo capturadas pelas instituições. O uso das tecnologias necessita de instrumentos que garantam aos indivíduos certo grau de proteção e resposta: meios de contenção de riscos, boas práticas éticas e confiáveis para utilização da Inteligência Artificial que buscam prevenir e diminuir fragilidades, bem como conferir explicabilidade das decisões tomadas por seres autômatos, máquinas que aprendem e produzem resultados são de suma importância.

Na Sociedade da Informação e do Cansaço (livro: “Sociedade do Cansaço”)<sup>3</sup>, a confluência entre comportamento do Ser Humano e os resultados sociais advindos da tecnologia em um mundo cada vez mais globalizado e conectado se faz presente.

A presente pesquisa tem como premissa demonstrar que temos condições de estabelecer boas práticas e modos de agir positivos para melhorar, utilizar e compreender a Inteligência Artificial, mitigando riscos e possibilitando mais robustez e confiança nas decisões (e consequências) proferidas pelos sistemas computacionais inteligentes.

Estudar o problema e as medidas afirmativas em várias frentes proporcionará uma melhor absorção de como esses sistemas interagem com nosso ambiente e vida.

Desse modo, diminuindo o temor da inovação em Inteligência Artificial, vivemos a Cibercultura:

Sociedade da Informação e suas características de interatividade, instantaneidade e simultaneidade, disruptivas do espaço-tempo físico e linear, propiciam o surgimento de um ambiente, o ciberespaço para onde fluem inúmeras relações econômico-sociais, agora leves líquidas, digitais ou virtuais, em contínuo inter-relacionamento com mundo concreto, pesado, que desapareceu, mas transformou-se.<sup>4</sup>

---

<sup>2</sup> MASSENO, Manuel David. Como a União Europeia procura proteger os cidadãos-consumidores em tempos de Big Data. In: MARTINS, Guilherme Magalhães; LONGHI, João Victor Rozatti. **Direito Digital: Direito Privado e Internet**. 4.<sup>a</sup> ed. Indaiatuba: Editora Foco, 2021, p. 496.

<sup>3</sup> HAN, Byung-Chul. **Sociedade do Cansaço**. Tradução: Ênio Paulo Giachini. 2.<sup>a</sup> ed. Petrópolis: Vozes, 2017.

<sup>4</sup> BARRETO JÚNIOR, Irineu Francisco; UELZE, Hugo Barroso. Cibercultura e o Poder Informacional na Esfera Pública: Crise do Paradigma Positivista no Direito Tributário. **Revista Prima Facie**, Universidade Federal da Paraíba, João Pessoa, v. 19, n. 40, p. 1-29, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/primafacie/article/view/51047>. Acesso em: 24 abr. 2024.

O desenvolvimento da autodeterminação, da proteção da privacidade e dados, a segurança da informação são alguns dos mecanismos da explicabilidade aos cidadãos, e permitem o exercício do Direito à Explicação e revisões dos resultados e decisões automatizadas em casos de vieses ou explicações que não atendam aos enunciados normativos e éticos da sociedade, que certamente avançará na utilização das novas tecnologias.

A relevância da pesquisa tem como escopo demonstrar a denominada Ética Algorítmica, pesquisando meios de aferição para que tais decisões possam evitar, na medida do possível, vieses discriminatórios ou alterações decisórias, diminuindo o grau de risco mediante medidas afirmativas e preventivas.

Frise-se que as decisões automatizadas influenciam até mesmo outros agentes de mercado que podem ser afetados por decisões anticoncorrenciais, prejudiciais para toda a sociedade e economia.

A tomada de decisão por parte dos sistemas de Inteligência Artificial precisa ser transparente para que, uma vez averiguada, comprove os critérios e dados utilizados para obtenção de resultados (*outputs*).

A Inteligência Artificial toma decisões com base no design computacional e em dados, e as questões relacionadas ao desenvolvimento, monitoramento e atualização da linguagem de programação.

O aprendizado de máquina da Inteligência é relevante na tomada de decisões, uma vez que precisamos, na medida do possível, entender como as máquinas aprendem, e de acordo com o aprendizado (da mesma forma que aprendem seres humanos), os sistemas computacionais inteligentes podem expressar vieses, discriminações.

A pesquisa apresenta medidas como a governança (de dados e de Inteligência Artificial), prestação de contas por meio de relatórios e auditoria (*accountability*) podem auxiliar no Direito à Explicação e à Explicabilidade.

A pesquisa também analisa a utilização de Inteligência Artificial como diferencial de mercado, seu poder econômico para domínio de mercado por empresas que utilizam esses sistemas computacionais, sem responder e explicar as decisões, e eventuais vieses ou discriminações decorrentes das tomadas de decisões da Inteligência Artificial que aprendem em profundidade (*Deep Learning*).

A partir da pesquisa será possível verificar a possibilidade de tomarmos decisões humanas melhores, conferir maior importância à proteção da privacidade e à proteção dos dados, obter informação sobre o objeto de inserção (*Inputs*), para tornar as decisões automatizadas explicáveis, conferindo maior transparência e aferição.



O presente trabalho enquadra-se na linha do Poder Econômico e seus limites jurídicos, considerando que a pesquisa abordará os aspectos legais e econômicos decorrentes do Direito à Explicação e a Explicabilidade em sistemas computacionais de Inteligência Artificial, e como essas decisões devem cumprir princípios de Direito, éticos e legais para explicação, possível revisão e para evitar riscos de vieses algorítmicos como discriminação e distorções de mercado por parte de empresas, e de que forma o resultado de suas conclusões para tomada de decisões são utilizadas por agentes de mercado.

A questão que orienta a pesquisa é saber se as decisões automatizadas podem ser objeto do Direito à Explicação e a Explicabilidade, bem como a revisão dessas decisões por seres humanos baseados em instrumentos de governança e relatórios para essa finalidade, aferindo-se o cumprimento de normas de Direitos Fundamentais, promovendo a ética e confiança algorítmica?

A pesquisa investiga se realmente tais decisões automatizadas, principalmente em aprendizado profundo (*Deep Learning*), viabilizam a ocorrência de vieses algorítmicos discriminatórios ou antiéticos prejudicando seres humanos individualmente ou em grupo.

O viés algorítmico causado pelo descumprimento do Princípio da Não Discriminação no campo da privacidade e da proteção de dados como base e fundamento da Inteligência Artificial, tal como exposta na Lei Geral de Proteção de Dados Pessoais e no Marco Regulatório da Inteligência Artificial, inspirado, também, no marco regulatório europeu, pode trazer entraves para a possibilidade de riscos sociais e econômicos no uso da Inteligência Artificial na medida que a coleta, o tratamento e o compartilhamento de dados são ativos das instituições ou se o “Respeito à Privacidade”, tal como exposto no art. 2º, inciso I da LGPD, estão em total equilíbrio com o desenvolvimento econômico da tecnologia, portanto, os impactos econômicos serão integralmente potencialmente positivos.

Analisando a bibliografia, artigos, pesquisas doutrinárias e empíricas disponíveis sobre Black Box, Opacidade e Direito de Explicação em Inteligência Artificial.

A metodologia da pesquisa buscou atingir os objetivos propostos utilizando-se do método Hipotético-Dedutivo: “o pesquisador elege o conjunto de proposições hipotéticas que acredita serem viáveis como estratégia de abordagem para se aproximar de seu objeto”<sup>5</sup>, ou seja, a presente pesquisa pretende fazer uso do método teórico e referencial dogmático (artigos científicos e livros) de abordagem, partindo do estudo de teorias e proposições hipotéticas que acredita serem viáveis como estratégia de abordagem para se aproximar do objeto (decisões

---

<sup>5</sup> MEZZARROBA, Orides; MONTEIRO, Cláudia Servilha. **Manual de Metodologia da Pesquisa no Direito**. 8.ª ed. São Paulo: SaraivaJur, 2019, p. 90.

automatizadas) e conclusão da forma (de que maneira) o Direito à Explicação, nos sistemas de Inteligência Artificial, auxilia na averiguação de como são tomadas as decisões automatizadas.

A pesquisa pretende verificar se as decisões automatizadas são ou não transparentes, e como poderemos torná-las objeto de explicação, evitando vieses discriminatórios e éticos, causando prejuízos ao mercado e aos usuários (finais e/ou indiretos) como utilização dessas decisões automatizadas como forma de desrespeito aos Direitos de Privacidade e Proteção de Dados (Direitos de Personalidade), uso dos dados como forma de poder dominante em Direito da Concorrência, bem como a segurança jurídica para as hipóteses concretas relacionadas à esfera da Inteligência Artificial.

A pesquisa tem desafios como: (i) informar e melhorar a explicabilidade e o Direito à Explicação em Inteligência Artificial; (ii) o papel e a responsabilidade exercida pelos humanos na decisão automatizada, na revisão e supervisão das decisões da IA, discutindo os benefícios e riscos de delegar ou compartilhar a tomada de decisão com máquinas; (iii) o uso de instrumentos de *compliance* e governança de Inteligência Artificial como método *ex ante* permitindo tornar decisões automatizadas mais transparentes e éticas; (iv) melhorar questões éticas, jurídicas e sociais da explicabilidade em Inteligência Artificial.

O estudo justifica-se pela importância do exercício do Direito à Explicação como direito de acesso à informação, garantindo maior transparência e controle dos atos praticados por instituições que se utilizam de Inteligência Artificial em suas decisões perante o mercado e os indivíduos, bem como estabelece diretrizes para desenvolvimento e utilização de tais sistemas computacionais com maior segurança jurídica e mercadológica.

O Objetivo da pesquisa, portanto, é entender a relação humano e sistemas computacionais inteligentes (IA) na formação das decisões automatizadas, riscos, benefícios e ética das decisões da máquina.

A hipótese da pesquisa é se decisões humanas influenciam o aprendizado das máquinas e suas decisões, bem como se a aplicação de regulação *ex ante* de governança, e a instituição de estatal formal por meio de Lei, Decretos, Resoluções) – hetero regulação, tem o escopo de estabelecer regras que assegurem o cumprimento da ética e transparência algorítmica como forma de evitar vieses e discriminações.

Referida hipótese permitirá tornar as tomadas de decisões automatizadas mais acessíveis e “confiáveis” em sistemas de Inteligência Artificial, principalmente respondendo indagações de usuários e instituições da sociedade, o que traz importantes reflexos econômicos, comerciais e aos usuários finais.

A pesquisa tem como marcos teóricos os professores Luciano Floridi<sup>6</sup>, cuja obra “*Etica dell'intelligenza artificiale: Sviluppo, opportunità, sfida.*” trata de questões éticas no uso e aplicação dos sistemas de Inteligência Artificial e sua usabilidade no desenvolvimento dos negócios, comércio, indústrias e sociedade, e Frank Pascale<sup>7</sup>, autor do livro: “*The Black Box Society*”, que aborda a questão da falta de transparência em algoritmos e sistemas de tomada de decisão automatizados usados em áreas como finanças, seguros, saúde e justiça. Pascale aborda a ausência de transparência em algoritmos de decisão utilizados em diversas áreas, desde finanças, saúde e até justiça criminal, e como isso pode perpetuar vieses discriminatórios e de desigualdade social e, ainda, explora a maneira como essas caixas-pretas são (ou devem ser) regulamentadas, pois a falta de supervisão e explicabilidade podem permitir abusos e fraudes.

Como marco teórico em ciência da computação, a pesquisa de do professor Stuart J. Russell e Peter Norvig, cuja obra “*Inteligência Artificial - Uma Abordagem Moderna*”<sup>8</sup> é referência técnica.

A pesquisa pretende alertar a necessidade de transparência e responsabilidade no uso de sistemas de Inteligência Artificial e como a regulação das caixas-pretas deve envolver partes interessadas da sociedade e usuários finais que são diretas - ou indiretamente - atingidos por decisões automatizadas não transparentes, e destaca a necessidade de criar mecanismos que garantam o desenvolvimento da Inteligência Artificial justa e transparente (ética algorítmica).

---

<sup>6</sup> Professor da Universidade de Oxford na Inglaterra. FLORIDI, Luciano. **Etica dell'intelligenza artificiale: Sviluppo, opportunità, sfida.** Milão-Itália: Raffaello Cortina Editore, 2022.

<sup>7</sup> Professor da Brooklyn Law School – New York/USA. PASQUALE, Frank. **The Black Box Society. The Secret Algorithms That Control Money and Information.** Cambridge-MA: Harvard University Press, 2015.

<sup>8</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna.** Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022.

# 1 RELAÇÃO ENTRE INTELIGÊNCIAS ARTIFICIAIS E DECISÕES HUMANAS

## 1.1 SISTEMAS INTELIGENTES E DECISÕES

A Inteligência Artificial, para McCarthy, é: “A inteligência é a parte computacional da capacidade de alcançar objetivos no mundo. Diversos tipos e graus de inteligência ocorrem em pessoas, muitos animais e algumas máquinas.”<sup>9</sup>

Inicialmente pensado por Turing, em um artigo científico de 1950<sup>10</sup>, trago a pesquisa a definição de Russell e Norvig, para inteligência artificial, que é mais abrangente que entender como pensamos e agimos, sendo:

O campo da inteligência artificial, ou IA, vai mais além: ele tenta não apenas compreender, mas também construir entidades inteligentes – máquinas que conseguem computar como agir de modo eficaz e seguro em uma grande variedade de novas situações.<sup>11</sup>

Para essa construção, a IA precisa de dados para aprender, pensar e proferir decisões sobre o mundo, e claro, sobre as pessoas que nele habitam.

A aplicação e importância à Lei Geral de Proteção de Dados (de forma física e digital), à Segurança da Informação (*off line* e *online*), bem como as normas que enunciam respeito e diretrizes sobre proteção aos Direitos de Personalidade como, por exemplo, o Código de Defesa do Consumidor, a Lei de Sigilo Bancário, o Sigilo Fiscal para a utilização de sistemas de Inteligência Artificial – e Softwares com aplicações gerenciais – que utilizam informações inseridas (por humanos ou por captação de outras bases de dados por interoperabilidade), pois são fundamentais para o uso ético da I.A.

Os dados desempenham um papel essencial no funcionamento da linguagem matemática da Inteligência Artificial. Eles são a base para treinar modelos matemáticos e estatísticos que permitem às máquinas aprenderem padrões, realizar previsões e tomar decisões.

---

<sup>9</sup> MCCARTHY, John. **Wha its Artificial intelligence?** Computer Science Department. Stanford: Stanford University CA 94305, 12 nov. 2007, p. 1-15. Available at: <https://www-formal.stanford.edu/jmc/whatisai.pdf>. Access in: 2 dez. 2024.

<sup>10</sup> TURING, Alan M. **Mind a Quarterly Review of Psychology and Philosophy**. Disponível em: <https://academic.oup.com/mind/article/LIX/236/433/986238>. Acesso em: 2 dez. 2024.

<sup>11</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.ª ed. São Paulo: LTC, 2022, p. 1.

De acordo com Russell e Norvig: "a matemática fornece os fundamentos para a construção de algoritmos que transformam grandes volumes de dados em conhecimento acionável, por meio de técnicas como álgebra linear, cálculo e teoria das probabilidades".<sup>12</sup>

Principalmente porque o Direito à Proteção de Dados é um Direito Fundamental expresso na Constituição Federal (art. 5, inciso LXXIX), diretamente vinculado à dignidade humana e aos demais Direitos Constitucionais e Humanos.

Was Rahman, físico e cientista da computação, diz: "o conceito de que inteligência é um termo complexo e composto por até oito características, como raciocínio, percepção, linguagem natural, mobilidade, aprendizado, representação de conhecimento, planejamento e consciência social".<sup>13</sup>

Nota-se que de acordo com o citado autor, a inteligência possui oito características, e algumas delas são humanas, ainda não incorporadas aos sistemas de inteligência artificial.

A I.A. precisa de problemas definidos, com delimitações específicas, porque são programadas para atingir resultados previsíveis, até certo contexto ou propósito controláveis, resultados matemáticos (de acordo com a programação da lógica algorítmica escrita) e, portanto, ainda há um longo caminho a ser percorrido para que os sistemas de Inteligência Artificial possam ser comparados – em termos de completude do resultado – para serem efetivamente "inteligências" como a dos seres humanos.

Esse longo caminho ainda passa pelas decisões automatizadas tomadas por Inteligência Artificial da mesma forma que um humano, uma vez que o Direito é ciência social aplicada (que une outras ciências como sociologia, antropologia, geografia, engenharia na tomada de decisão), fundamental – portanto – entendermos o aprendizado e a decisão da I.A.

Os processos decisórios tratam de questões individuais e corporativas – e particularmente – argumentadas, provadas e fundamentadas para casos concretos (fatos, direito e provas) entre as partes envolvidas por um contexto ou por outras características e elementos dele constantes. No caso da automação por sistemas inteligentes, as questões devem ser devidamente informadas, classificadas e tratadas para melhorar o *output*, o resultado do processamento que a rede neural estruturada e arquitetada para determinada solução proferirá.

Isso confere não só devido enunciado normativo para a aplicação efetiva da Lei Geral de proteção de dados pessoais e demais leis que tratam de dados e informações, mas que também promovem a proteção dos Direitos de Personalidade, como acima citado.

---

<sup>12</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.ª ed. São Paulo: LTC, 2022, p. 117.

<sup>13</sup> RAHMAN, Was. **Inteligência artificial e aprendizado de máquina**. 1.ª ed. São Paulo: SENAC, 2022.

Quanto aos dados (corporativos e pessoais) e sua relação como um dos fundamentos para tais tecnologias, entende-se que:

A utilização de dados pessoais para alimentar os novos sistemas de inteligência artificial e a sua utilização para tomar decisões proporcionam uma acurácia bastante significativa para um número crescentes de aplicações. Isto abre espaço para, ao menos, dois temas centrais para os debates sobre autonomia e direitos fundamentais nos próximos anos: os efeitos que a utilização desses sistemas causará para a pessoa e sua autonomia pessoal, bem como a necessidade de qualificar a natureza desses instrumentos e sistemas de inteligência artificial. Nesse debate, a necessidade de que sejam proporcionadas soluções que preservem os direitos fundamentais, dentro de um quadro de intenso desenvolvimento tecnológico e mesmo de questionamento de alguns institutos centrais do ordenamento jurídico, sugere a necessidade de recorrer à ética como instrumento capaz de encaminhar soluções que, eventualmente, e se for o caso, possam consolidar-se em alternativas legislativas posteriormente.<sup>14</sup>

Os dados são um ativo importante para a atividade nuclear e funcional da Inteligência Artificial, ou seja, está no cerne, e na alma, da atividade a ser executada e no resultado da execução. Seu resultado é o processamento desses dados e informações pela lógica matemática (algoritmos).

Nas Inteligências Artificiais Generativas, como diz Tom Taulli: “[...] necessitamos muito de infraestrutura de dados sofisticada”<sup>15</sup>, que demandam ainda mais dados e atenção ao uso.

Isto torna clara a dependência da inteligência artificial com dados:

[...] sendo que o volume (*big data*), pode acelerar a programação. Por consequência, a capacidade de processamento da máquina está relacionada, como o próprio título sugere, ao tratamento destes dados de forma a complementá-los com resultados efetivos.<sup>16</sup>

Desta forma, será necessário um esforço enorme para que os dados imputados no sistema de Inteligência Artificial (ou software a ele integrado) sejam devidamente mapeados,

---

<sup>14</sup> DONEDA, Danilo Cesar Maganhoto *et al.* Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar – Revista de Ciências Jurídicas**, Fortaleza, v. 23, n. 4, p. 3, 2018. Disponível em: <https://ojs.unifor.br/rpen/article/view/8257/pdf>. Acesso em: 24 abr. 2024.

<sup>15</sup> Tradução livre: “[...] the needs to be a highly sophisticated data infrastructure.” TAULLI, Tom. **Generativa AI: How ChatGPT and Other AI Tools Will Revolutionize Business**. New York: Apress, 2023, p. 21.

<sup>16</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Desafios Internacionais da Aplicação da Inteligência Artificial no Direito. **Revista Justiça de Direito**, v. 35, n. 2, p. 10, mai./ago. 2021. Disponível em: <https://seer.upf.br/index.php/rjd/article/view/13040/114116181>. Acesso em: 24 abr. 2024.

classificados, inseridos, e assim governados para que a I.A. possa – em sua saída (*output*) – dar resultados com o maior grau de assertividade e segurança jurídica para o caso concreto.

A matemática escrita em forma de linguagem em múltiplas vertentes, demandando estudo e análise linguística profunda para seu desenvolvimento e disponibilização:

A linguística computacional é estudada por linguistas e por cientistas da computação. É comumente chamada de Processamento de Linguagem Natural, de onde a sigla PLN (ou NLP, em inglês). É estudada, também, como parte das chamadas ciências cognitivas, um conjunto de disciplinas que reúne domínios como neurociências, a filosofia, a psicologia e, mais uma vez, a ciência da computação e a linguística, todas voltadas ao estudo da mente humana. Com esse objetivo em foco, a investigação tem sempre de levar em consideração a linguagem humana. É nesse contexto que a linguagem computacional participa das pesquisas sobre inteligência artificial. Através delas, vamos imediatamente ampliar o alcance do que acabamos de dizer sobre a antiga ambição de criar máquinas capazes de compreender e produzir expressões em língua humana. A essas habilidades, soma-se a intenção de criar máquinas capazes de aprender línguas.<sup>17</sup>

O comentário acima colacionado pela pesquisa demonstra que o desenvolvimento, treinamento e aprendizado das máquinas, tornando-as de certa forma inteligentes artificialmente, passa pelo processo de linguagem natural ao qual são utilizadas – como exemplo – a programação de algumas linguagens como Python, Análise Quantitativa de Corpus, Modelos de N-Gramas e Classificadores Bayesianos Ingênuos.

Python é uma linguagem de programação de propósito geral, isto é, que serve para diferentes finalidades”, porque ela pode “incorporar módulos externos aos programas, isto é, recursos específicos criados por outros programadores, o que permite utilizar funcionalidades avançadas com mínimo esforço.<sup>18</sup>

Através da inserção de dados no sistema que compõe a base (dados estes inseridos por humanos ou coletados no Big Data), a Inteligência Artificial utilizar-se-á para aprender e para alcançar os resultados de sua programação, o que se espera, que estejam adequados ao serviço prestado, no qual as partes esperam a melhor e mais adequada solução de um problema, cujo algoritmo é construído para solucioná-lo.

Os dados externos que são imputados na programação, tornando-se códigos, uma vez que:

---

<sup>17</sup> FERREIRA, Marcelo; LOPES, Marcos. **Para Conhecer Linguística Computacional**. São Paulo: Contexto, 2023, p. 10.

<sup>18</sup> *Ibidem*, p. 16.

Os computadores são máquinas que precisam se comunicar por meio de informações gravadas ou transitando por meio de cabos, discos, ondas de rádio (como redes de Wi-Fi) etc. Os dados físicos presentes nesses diferentes meios são convertidos para valores numéricos binários, e, estes, para qualquer outra forma de representação: letras de um alfabeto quaisquer números, marcas de início e fim de um arquivo. [...].<sup>19</sup>

Acima citado, um modelo N-Grama é uma sequência de “n” elementos em um determinado nível para análise (letras, morfemas, palavras, imagens, etc.), portanto:

São modelos probabilísticos em que se estima a probabilidade de uma sequência qualquer a partir dos n-gramas que a compõe. Assim, a **probabilidade associada a um n-grama** é expressa por  $P(w_n | w_1 w_2 \dots w_{n-1})$ . A ideia central é que a ocorrência de uma palavra em uma sequência é condicionada somente pelas n-1 palavras que a antecedem imediatamente. Desprezam-se influências de palavras mais distantes, no que se conhece como **Assunção de Markov**.<sup>20</sup>

A probabilidade é medir a ocorrência, avaliando-a.

Nos modelos probabilísticos, os desenvolvedores utilizam-se da perplexidade como medida inversamente proporcional à probabilidade que o modelo atribuí ao corpus de teste, ou seja, “quanto maior a probabilidade, menor a perplexidade”.<sup>21</sup>

Desenvolvedores utilizam modelos probabilísticos para prever eventos futuros ou inferir informações desconhecidas com base em dados observados. Esses modelos são fundamentais em sistemas de IA, especialmente no processamento de linguagem natural (PLN), onde auxiliam na previsão da próxima palavra ou sequência de palavras em um texto.

A perplexidade é uma métrica comum nesses contextos, sendo usada para avaliar a qualidade de um modelo probabilístico. É utilizada para avaliar a qualidade de modelos de linguagem natural em inteligência artificial (IA). Ela quantifica o grau de incerteza que o modelo possui ao prever a próxima palavra em uma sequência de texto. Matematicamente, a perplexidade é definida como a exponencial da entropia cruzada entre a distribuição real das palavras e a distribuição prevista pelo modelo.

Como destacado por Jurafsky e Martin, "a perplexidade mede a incerteza de um modelo em prever um conjunto de dados, sendo diretamente relacionada ao seu desempenho".<sup>22</sup>

<sup>19</sup> FERREIRA, Marcelo; LOPES, Marcos. **Para Conhecer Linguística Computacional**. São Paulo: Contexto, 2023, p. 16.

<sup>20</sup> *Ibidem*, p. 130.

<sup>21</sup> *Ibidem*, p. 145.

<sup>22</sup> URAFASKY, Daniel; MARTIN, James H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3.<sup>a</sup> ed. Hoboken: Pearson, 2023, p. 142.



Para que seja possível analisar padrões e probabilidades são necessários critérios de classificação:

De forma geral, classificar é agrupar os dados, objetos ou observações em categorias predeterminadas. Um classificador pode ser visto como um dispositivo que mapeia esses dados (*inputs*) em determinada classe, e a mesma, parte de um conjunto de classes (*outputs*). Várias tarefas de processamento de linguagem natural encaixam-se nessa definição. Eis exemplos diferentes de níveis de análise linguística: (i) detecção de spams, em que mensagens de e-mails são classificadas como spam ou não-spam, podendo, a partir dessa classificação, serem filtradas e talvez eliminadas da caixa de entrada de um usuário; (ii) etiquetagem gramatical (POS *tagging*), em que se atribui uma classe morfosintática (substantivo, verbo, preposição, etc.) a ocorrências de palavras, geralmente como parte de um processo de análise sintática de sentenças; (iii) desambiguação de sentido de palavras (WSD, Word Sense Desambiguation), em que ocorrências de palavras ambíguas em um texto são relacionadas a um dos sentidos dicionarizados da palavra (“manga”, por exemplo, como fruta ou como parte de roupa). O Mapeamento de um dado em uma classe baseia-se sempre numa lista de atributos, ou características (*features*), desse dado. [...] Uma seleção adequada de atributos é sempre parte fundamental em uma tarefa de classificação. Nunca é demais lembrar que mesmo o que parece óbvio e trivial a um ser humano precisa ser “dito” ou “ensinado” a uma máquina se quisermos que ela execute automaticamente uma tarefa.<sup>23</sup>

Thomas Bayes – criador do classificador bayesiano ingênuo (Naïve Bayes classifier) – foi um matemático presbiteriano inglês conhecido por seu teorema que exerceu muita influência nos campos da estatística e da teoria da probabilidade, tendo impacto particularmente na filosofia.

O uso classificatório da palavra “ingênuo” é devido pela simplificação na aplicação da regra que facilita sobremaneira o cálculo das probabilidades.<sup>24</sup>

A importância dessas explicações para o tema da pesquisa é relevante, uma vez que para entendermos e concluirmos sobre Direito à Explicação e a Explicabilidade dos modelos de Inteligência Artificial, foi necessário pesquisar e entender sobre como as máquinas absorvem as informações transformando, interpretando e aprendendo para posteriormente informar sua conclusão ou decisão sobre aquele certo dado ou informação solicitada.

A pesquisa demonstra que o tema é complexo, instigante e explica como os dados são importantes para o Processamento de Linguagem Natural, uma vez que desde sua coleta,

---

<sup>23</sup> FERREIRA, Marcelo; LOPES, Marcos. **Para Conhecer Linguística Computacional**. São Paulo: Contexto, 2023, p. 161-162.

<sup>24</sup> *Ibidem*, p. 164.

classificação, tratamento, inserção em outros sistemas de software tornar-se-ão insumo da Inteligência Artificial que aprende e toma decisões.

Para conseguir resultados claros e mais aproximados com o pretendido, é preciso atender aos ditames linguísticos expostos acima como pressupostos fundamentais aliados aos princípios da precaução, transparência, tratamento de dados, etc.

Estudando os dados – informações – como objetos fundamentais para análise preditiva pode informar determinados comportamentos, condições e acontecimentos em escala, que serão objeto de uma análise e direcionamento dentro do sistema da Inteligência Artificial, alterando seus resultados.

A formação de uma decisão automatizada passa pelos trâmites da proteção de dados, ou seja, pelo cumprimento dos princípios da adequação, finalidade, prevenção, etc., tendo como um desses princípios, o da Qualidade dos Dados, assim necessária classificação de dados (n-gramas) corretos para formação linguística de aprendizado do sistema computacional.

As instituições usam dados para gestão, analisando predições e os padrões para – sob certas e determinadas condições – que quando inseridas na Inteligência Artificial causam decisões de máquina, e é preciso saber como coletar e tratar os dados antes de imputá-los, garantindo, assim, que a máquina aprenda de forma mais eficiente, garantindo uma maior assertividade de resultados e maior transparência.

Desse modo, a pesquisa verificou o fortalecimento – para decisões de agentes autônomos inteligentes, a Lei Geral de Proteção de Dados, leis de sigilo, de acesso, bem como a Constituição Federal preveem a proteção dos indivíduos, devendo ser mantidas sob o ideal de afirmação dos Direitos e Garantias Fundamentais e dos Direitos de Personalidade.

Observa-se que o estágio atual da sociedade e da economia ao qual:

[...] se observa uma nova gama de atividades sintetizadas pela denominação de economia digital, cujo crescimento, de início, se verificou de maneira indireta, porque relacionada aos bens e produtos corpóreos e tangíveis, porém, comercializados através de Tecnologias da Informação e Comunicação (TICs), e depois, pelo aparecimento de bens e serviços com contornos, conteúdos ou formatos, eterizados, virtuais e, mesmo, líquidos, daí a dificuldade da descrição de fatos ou eventos [...].<sup>25</sup>

---

<sup>25</sup> BARRETO JÚNIOR, Irineu Francisco; UELZE, Hugo Barroso. Cibercultura e o Poder Informacional na Esfera Pública: Crise do Paradigma Positivista no Direito Tributário. **Revista Prima Facie**, Universidade Federal da Paraíba, João Pessoa, v. 19, n. 40, p. 3, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/primafacie/article/view/51047>. Acesso em: 24 abr. 2024.

O comércio informacional (descrito também como eletrônico) tornou os dados um objeto econômico de muita importância, preponderando sobre outros fatores até então mais significativos para a economia e pensados como imbatíveis, como petróleo, aço e ouro, fazendo com que:

A compreensão desse novo paradigma, o do poder comunicacional exige, pois, uma abordagem que alcance as diferentes formas de relação humana e de organização social, daí suas implicações transdisciplinares – aspectos antropológicos, econômico-sociais, sociológica, político-jurídico, etc. – cujos desdobramentos não podem ser solucionados sem uma análise mais ampla, cientes de que a Ciência e o Direito não devem ser separados do meio em que se vêm chamados a atuar.<sup>26</sup>

Como as inteligências artificiais demandam que este poder lhes seja transferido para que aprendam e decidam sobre consumo, geolocalização, cultura, política, preferências, viagens, etc.

Esse poder informacional citado acima um ponto fundamental de pesquisa de como nossas decisões transformadas em dados para ensino da rede neural artificial seja um fator não somente social, mas econômico de muita importância.

A aprendizagem por meio da utilização e exteriorização dos dados tornam a Inteligência Artificial relevante para o cenário mundial, conforme informam e explicam Russel e Norvig:

Um agente aprenderá a melhorar seu desempenho após fazer observações sobre o mundo. A aprendizagem pode variar do corriqueiro, como anotar uma lista de compras, até o mais profundo, como mostrado por Albert Einstein, que deduziu uma nova teoria para o universo. Quando o agente é um computador, nós o chamamos aprendizado de máquina: um computador observa alguns dados, monta um modelo baseado nos dados e usa o modelo como uma hipótese sobre o mundo e um software que pode resolver problemas. [...] A tecnologia de aprendizado de máquina se tornou uma parte padrão da engenharia de software. Sempre que você estiver criando um sistema de software, mesmo que não pense nele como um agente de IA, os componentes do sistema podem ser potencialmente melhorados com o aprendizado de máquina.<sup>27</sup>

---

<sup>26</sup> BARRETO JÚNIOR, Irineu Francisco; UELZE, Hugo Barroso. Cibercultura e o Poder Informacional na Esfera Pública: Crise do Paradigma Positivista no Direito Tributário. **Revista Prima Facie**, Universidade Federal da Paraíba, João Pessoa, v. 19, n. 40, p. 3, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/primafacie/article/view/51047>. Acesso em: 24 abr. 2024.

<sup>27</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.ª ed. São Paulo: LTC, 2022, p. 590-591.

As decisões pós-aprendizado podem ser simples ou complexas, sendo que a **tomada de decisão simples** escolhe de acordo com a teoria da probabilidade, levando em conta redes de decisão e valoração da informação (valoração da informação pensada como o resultado da sua correspondência e utilizada em cada instituição que dela se serve), todos estes componentes da programação escrita em linguagem computacional que, após os dados serem informados a I.A., resultam de uma decisão simples e binária (bom e ruim, por exemplo).

A **tomada de decisão complexa** leva em consideração os aspectos do parágrafo anterior, adicionando-se os processos de decisão de Markov, que se definem por um modelo de transição com especificações probabilísticas de ações somadas as formas de recompensa, cujos algoritmos de interação de valor resolvem um conjunto de equações que se relacionam com as utilidades encontradas para execução – e aqui encontramos uma complexidade na aplicação dos sistemas para decisões judiciais explicáveis, bem como utiliza uma rede de decisão dinâmica para atualizar suas crenças e projetar cenários para possíveis ações futuras, tudo isso mediante a imputação de informações de variadas fontes (que são, ou não, constituídas por outras bases de dados e informações adicionais – até mesmo necessárias ao deslinde da questão posta em juízo).

A **tomada de decisão em ambientes multiagentes** surge em circunstâncias que são compostas por vários atores, no qual o tomador de decisão (no caso a IA) utiliza-se de teoria dos jogos, seus conceitos de solução que integram sua programação para o equilíbrio decisório e atingimento do resultado que se espera.

Portanto, como reflete João Sérgio dos Santos Souza Pereira:

Tomar os padrões como textos, ainda que com a finalidade de classificações e estruturações de dados, a partir de análise semântica, para posterior tomada de decisões, por meio das informações obtidas, nos leva ao nosso primeiro e principal parâmetro ou condição de possibilidade proposto: que os procedimentos e os eventuais modelos tecnológicos estejam adaptados e permeados por garantia, resguardados por meio do devido processo legal.<sup>28</sup>

Não podemos deixar de fora dessas linhas a questão atinente ao Aprendizado Profundo (*Deep Learning*), que são algoritmos com complexos circuitos algébricos e intensidades de conexão ajustáveis, ou seja, camadas e redes neurais complexas para que a Inteligência Artificial possa adquirir mais aprendizado diverso e com maior complexidade (a profundidade

---

<sup>28</sup> PEREIRA, João Sérgio dos Santos Soares. **A padronização decisória na era da inteligência artificial: uma possível leitura hermenêutica e da autonomia do direito**. Belo Horizonte: Casa do Direito, 2021, p. 374. (Coleção Direito Processual na Ordem Constitucional).

da denominação chamamos de Aprendizado Profundo), algo cada vez menos explicável dada sua complexidade e infinidade de equações e dados, as Caixas Pretas.

O processo de decisão:

Acontece no processo conhecido como if/then (se/então) que significa programar, em símbolos lógicos, por exemplo, a seguinte frase “se chover, então abra o guarda-chuva. Se não chover, então use a piscina.” Isso não é inteligência, é apenas lógica direta.<sup>29</sup>

Levando-se em conta que decisões (humanas e de máquinas) carregam muitas variantes, questões de fato, o fato descrito na norma jurídica e sua enunciação, interpretação jurídica adequada, possíveis súmulas, doutrina, precedentes judiciais e entendimentos etc., a alocação dos dados do fato de forma correta para que a Inteligência Artificial possa utilizar a informação de maneira adequadamente classificada e inserida como tal para que sua decisão complexa seja aferível.

Diversas são as versões sobre o que é Inteligência Artificial, todavia, de acordo com a pesquisa seria a capacidade de um computador processar dados e informações coletados para inserção em sistema contendo algoritmos para obtenção de resultados e aprendizado por padrões silogísticos, lógicos e cujas probabilidades preenchem lacunas para aprendizado e raciocínio de tais sistemas.<sup>30</sup>

A Inteligência Artificial idealizada por Alan Turing quando esboçou o projeto de experimento que permitiria ocupar o espaço de questionamento da filosofia sobre a possibilidade de uma máquina pensar como um ser humano somente seria validado quando sua aprovação no teste, mas que não levou em consideração outros aspectos relacionados ao pensamento, à cognição, ao raciocínio, às crenças, aos subjetivismos, às culturas, etc.

A Inteligência Artificial dos “mais interessantes e de mais rápido crescimento, já conseguindo gerar mais de um trilhão de dólares por ano em receitas”.<sup>31</sup>

O impacto da tecnologia em nossas vidas é claro, estando em uma infinidade de aplicativos, softwares, sistemas computacionais, empresas, instituições públicas e privadas, pesquisa científica, sistema financeiro, segurança, securitização, saúde, educação, marketing, escritórios e profissões diversas, que não conseguiremos descrever aqui porque ela está inserida

---

<sup>29</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Desafios Internacionais da Aplicação da Inteligência Artificial no Direito. **Revista Justiça de Direito**, v. 35, n. 2, p. 10, mai./ago. 2021. Disponível em: <https://seer.upf.br/index.php/rjd/article/view/13040/114116181>. Acesso em: 24 abr. 2024.

<sup>30</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022, p. 1.

<sup>31</sup> *Ibidem*, p. 1.

em quase toda parte ou relação de nossas vidas, mesmo quando não estamos utilizando tecnologias como computadores, tablets, smartphones e outros gadgets/devices.

A afirmação de que “as benesses da IA vêm acompanhadas de muitas incertezas, posto que o impacto dela na sociedade tão profundo que, em pouco tempo, todas as áreas da vida humana são atingidas, em menor ou maior grau”.<sup>32</sup>

Evitando uma conceituação formal e estática, entendemos que Inteligência Artificial é o conjunto de disciplinas (matemática, estatística, engenharia, computação, psicologia, Direito) que se conectam e interagem para a obtenção de resultados ou decisões que buscam proximidade com as decisões humanas.

Fruto do conjunto de informações ou dados inseridos em determinado computador, considerando-se as linguagens desenvolvidas para obtenção dos resultados, a IA consegue obter um aprendizado por representação e repetição de padrões, similares ao aprendizado de conhecimento que o humano adquire ao viver.

A reflexão trazida na pesquisa tem relação com a denominada BIAS (é como o intercepto adicionado em uma equação linear, ou seja, é o parâmetro adicional na Rede Neural) usada para ajustar a saída de informações (decisões) junto da soma ponderada das entradas (*inputs*) para o neurônio (humano) e “*neuron*” (Inteligência Artificial) é requisito para compreensão do tema.

O peso padrão que a Inteligência Artificial recebe durante esse processo de aprendizado, cujos resultados decisórios não aferíveis ou passíveis de modificação (após aferição de opacidade, por exemplo), tornam a utilização da Inteligência Artificial um risco para a sociedade e a economia para casos de não explicação das metodologias aplicáveis a tais tecnologias computacionais.

Em termos de mercado, as *Data Brokers* e as *Big Techs* – atuam proeminentemente e diretamente com inovação digital – na posse de uma infinidade de dados (e de sua utilização indevida para inserção em sistemas de IA), podem manipular mercados e indivíduos ou grupos, ofertar produtos e serviços com preços dispare, segregando dessa forma indivíduos ou grupos de maneira a aplicar as tecnologias em favor da desigualdade humana e econômica.

A coleta e tratamento de dados, com inserção em softwares inteligentes ou softwares que contenham integração com Inteligência Artificial são denominadas API’s (sigla para

---

<sup>32</sup> PICAZIO, Joseph Rodrigo Amorim. **Inteligência Artificial e seus Impactos nos Direitos Sociais: um panorama da convergência e dos desafios da IA para a efetividade dos direitos e garantias fundamentais**. São Paulo: Dialética, 2024, p. 13.

*Application Programming Interface*), que em português significa Interface de Programação de Aplicações.

APIs (*Application Programming Interfaces*) são conjuntos de definições e protocolos que permitem que diferentes softwares se comuniquem entre si. Elas funcionam como intermediárias, facilitando a interação entre sistemas, aplicações e dispositivos, ao expor funcionalidades de um sistema para que possam ser utilizadas por outros. No contexto de Inteligência Artificial, as APIs desempenham um papel crucial ao oferecer interfaces padronizadas para acessar modelos treinados, serviços de aprendizado de máquina ou bancos de dados, promovendo escalabilidade e interoperabilidade.<sup>33</sup>

A API pode ser usada para acessar serviços ou recursos de Inteligência Artificial como reconhecimento de voz, processamento de linguagem natural, visão computacional ou aprendizado de máquina, sem ter que desenvolver essas funcionalidades do zero, portanto, aprendendo com o que lhes são inseridos em termos de informações. Por exemplo, uma empresa pode usar uma API de Inteligência Artificial para integrar um *chatbot* em seu site, ou um aplicativo pode usar uma API de IA para oferecer recursos de tradução ou de edição de fotos aos seus usuários – e seus reflexos na sociedade – se faz necessário o tratamento com segurança jurídica suficiente para garantia da privacidade de cada indivíduo.

Os sistemas de Inteligência Artificial possuem diferentes formas de atuação e expressão, ou seja, algumas aprendem, outras apenas auxiliam de forma a cumprir funções predeterminadas em seu desenvolvimento e programação.

Essa compreensão das diferentes espécies de Inteligência Artificial (IA) é crucial para o avanço tecnológico e a implementação ética destas tecnologias. A seguir, a pesquisa descreve as principais categorias e exemplos de IA baseados em suas capacidades e funcionalidades.

Controles preventivos, boas práticas e implementação de programas (*Frameworks*), governança na utilização dessas ferramentas computacionais, iniciando tal tarefa quando o uso de dados coletados da vida privada, dados pessoais de identificação e dados corporativos que subsidiam as decisões não constituam máculas aos direitos fundamentais, sociais e econômicos, cada vez mais utilizados no perfilamento individual de grupos e de empresas, e que predeterminam os resultados exarados (decisões automatizadas, *Outputs*) pela Inteligência Artificial, reforçando linhas atrás as leis de proteção e privacidade.

---

<sup>33</sup> FIELDING, Roy Thomas. **Architectural Styles and the Design of Network-based Software Architectures**. Dissertação (Doutorado em Ciência da Computação) – University of California, Irvine, 2000, p. 1-24. Disponível em: <https://www.proquest.com/openview/fc2d064044b971dda476dfb429a2b344/1?pq-origsite=gscholar&cbl=18750&diss=y>. Acesso em: 2 dez. 2024.

Os meios e características utilizados pela Inteligência Artificial (por meio de dados obtidos com vigilância e invasão de privacidade, como exemplo), bem como características físicas e comportamentais informadas em cadastros de coleta, tratamento de dados que são posteriormente disponibilizados e imputados na Inteligência Artificial poderão ocasionar vieses discriminatórios, modulação de comportamento, etc.

E a pesquisa não deixou de fora as Inteligências Artificiais Generativas, que conforme elucida Tom Taulli: “[...] necessitamos muito de infraestrutura de dados sofisticada”.<sup>34</sup> Ou seja, precisam de dados para sua base, aprendizado e resposta gerada.

As inteligências artificiais generativas tiveram como ponto de partida o artigo “Attention Is All You Need”, de autores cuja pesquisa em conjunto<sup>35</sup>, descrevem no resumo do artigo que:

Os modelos de transdução de sequência dominantes são baseados em redes neurais recorrentes ou convolucionais complexas que incluem um codificador e um decodificador. Os modelos de melhor desempenho também conectam o codificador e o decodificador através de um mecanismo de atenção. Propomos uma nova arquitetura de rede simples, o Transformer, baseada exclusivamente em mecanismos de atenção, dispensando completamente a recorrência e as convoluções. Experimentos em duas tarefas de tradução automática mostram que esses modelos são superiores em qualidade, ao mesmo tempo em que são mais paralelizáveis e requerem significativamente menos tempo para treinar. Nosso modelo alcança 28,4 BLEU na tarefa de tradução do WMT 2014 de inglês para alemão, melhorando os melhores resultados existentes, incluindo conjuntos, em mais de 2 BLEU. Na tarefa de tradução do inglês para o francês do WMT 2014, nosso modelo estabelece uma nova pontuação BLEU de estado da arte de 41,0 em um único modelo, após 3,5 dias de treinamento em oito GPUs, uma fração pequena dos custos de treinamento dos melhores modelos da literatura.<sup>36</sup>

A pesquisa traz, ainda, um fator de muitos questionamentos e variados conceitos, temos no Algoritmo um dos pilares fundamentais da Inteligência Artificial. Com efeito, é preciso entender o que seriam algoritmos, ainda mais porque vivemos na era dos algoritmos ou como denomina Juliana Abrusio, Algocracia.<sup>37</sup>

---

<sup>34</sup> TAULLI, Tom. **Generativa AI: How ChatGPT and Other AI Tools Will Revolutionize Business**. New York: Apress, 2023, p. 21.

<sup>35</sup> *Ibidem*, p. 21.

<sup>36</sup> VASWANI, Ashish *et al.* Attention is all you need. **31st Conference on Neural Information Processing Systems (NIPS)**. Long Beach – CA: 2017, p. 1. Available at: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Access in: 25 jan. 2025.

<sup>37</sup> ABRUSIO, Juliana. **Proteção de Dados na Cultura do Algoritmo**. 1.ª ed. São Paulo: D’Plácido, 2020, p. 204. (Coleção Direito privado, v. 1).



Algoritmos não se explicam de forma assim tão simples, Russell e Norvig descrevem-nos como pseudocódigo, escritos em linguagem computacional como Java, C++, e em especial Python, mas que podem ser expressos também como fórmulas matemáticas ou linguagem comum para diminuir a complexidade de sua descrição podendo ser observadas como: variáveis persistentes; funções como valores, recuo, atribuição de desestruturação, valores padrão para parâmetros, produz (como função geradora); laços; listas; conjuntos; Arrays começam em 1.38

Algoritmos não são simples de explicar, mas a compreensão do que são algoritmos é fundamental para entender Inteligência Artificial, sendo:

O processo ou conjunto de regras a serem seguidas em cálculos ou outras operações de solução de problemas, especialmente, por um computador. O objetivo de um algoritmo é resolver um problema específico, geralmente, definido por alguém como uma sequência de instruções.<sup>39</sup>

Diogo Rais define algoritmos de forma didática:

Um algoritmo nada mais é que um conjunto de instruções ou passos a serem seguidos, especificando o que fazer para cada situação. Qualquer receita de pão é um exemplo de algoritmo. Transportando essa noção para a ciência da computação, fica mais fácil compreender que os algoritmos são instruções para o computador, quando se depare com determinada circunstância, execute as ações previstas para tal. Esses algoritmos podem ser programados para evoluir a partir de uma base de dados com que se caracterizariam como inteligência artificial.<sup>40</sup>

Sicsu, Samartini e Barth<sup>41</sup> entendem que “*Algoritmo* é o conjunto de procedimentos a serem executados, em determinada sequência, a fim de transformar um conjunto de dados de entrada em um ou mais valores.”

Em termos de tamanho de conjuntos, procedimentos, números, tarefas que um sistema de Inteligência Artificial é ilimitado, conforme expõe Penrose:

O algoritmo – ou o procedimento de cálculo geral – é exatamente o mesmo, não importa quão grandes os números sejam. Para números grandes, o procedimento pode de fato levar muito tempo para terminar, e uma quantidade bem grande de “papel rascunho” pode ser necessária para efetuar os cálculos

<sup>38</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022, p. 935.

<sup>39</sup> LAGE, Fernanda de Carvalho. **Manual de Inteligência Artificial no Direito Brasileiro**. 1.<sup>a</sup> ed. Salvador: Juspodivm, 2021, p. 37.

<sup>40</sup> RAIS, Diogo; PRADO FILHO, Francisco Octávio de Almeida. **Direito público digital: o Estado e as novas tecnologias – desafios e soluções**. São Paulo: Revista dos Tribunais, 2020, p. 65.

<sup>41</sup> SICSU, Abraham Laredo; SAMARTINI, André; BARTH, Nelson Lerner. **Técnicas de Machine Learning**. São Paulo: Blucher, 2023, p. 16.

em si. Porém, o algoritmo é o mesmo conjunto finito de instruções, não importa o quão grandes sejam, os números. [...] Já que nosso aparato possui somente um número finito de estados internos distintos, não podemos esperar que ele “internalize” todos os estados externos, nem todos os resultados de seus próprios cálculos. Em vez disso, ele deve examinar somente as partes dos dados ou cálculos prévios com os quais está lidando naquele instante e então efetuar qualquer operação que deva.<sup>42</sup>

Na obra “Inteligência Artificial – Uma Abordagem Moderna”<sup>43</sup>, os autores conceituam a Inteligência Artificial como um grande campo de pesquisa em Ciência da Computação que busca criar sistemas capazes de realizar tarefas que normalmente requerem inteligência humana.

Essas tarefas incluem aprendizado, raciocínio, percepção, linguagem natural, entre outros, como por exemplo o **Aprendizado de Máquina**, um subcampo da IA, que tem como escopo o estudo de algoritmos e modelos estatísticos que os sistemas de computador usam para realizar tarefas com ou sem instruções explícitas, confiando em padrões e inferência que podem ser descritos como: (i) **Aprendizado Profundo (*Deep Learning*)**: como uma subcategoria do aprendizado de máquina que cria redes neurais com muitas camadas (daí o ‘profundo’ em aprendizado profundo) e dessa maneira consegue aprender representações de dados com vários níveis de abstração; (ii) **Programação Probabilística**: como abordagem para modelar e inferir incerteza em dados complexos e domínios de conhecimento; (iii) **Sistemas Multiagente**: cujo objetivo é estudar como múltiplos agentes inteligentes interagem para alcançar seus objetivos; (iv) **Aprendizado por Reforço**: um tipo de aprendizado de máquina onde um agente aprende a se comportar em um ambiente realizando ações que maximizam algum conceito de recompensa cumulativa; (v) **Visão Computacional**: que é a capacidade de processar imagens e vídeos para obter informações úteis; e (vi) **Processamento de Linguagem Natural (PLN)**: capacidade e concentração para entender a interação entre computadores e humanos usando a linguagem natural. O objetivo final do PLN é ler, decifrar, entender e fazer sentido da linguagem humana de uma maneira valiosa para que possa se aproximar do humano cada vez mais. No livro, Russell e Norvig discutem vários aspectos da IA incluindo o PLN. Eles descrevem alguns componentes chave da IA que incluem: (a) **Processamento de Linguagem Natural**: para permitir que a IA se comunique com sucesso em um idioma natural; (b) **Representação de Conhecimento**: para armazenar o que a IA sabe ou ouve; (c) **Raciocínio**

---

<sup>42</sup> PENROSE, Roger. **A mente do novo imperador: Sobre computadores, mentes e as leis da física**. Tradução: Gabriel Cozzella. São Paulo: UNESP, 2023, p. 77.

<sup>43</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022.

**Automatizado:** para usar as informações armazenadas com a finalidade de responder a perguntas e tirar novas conclusões.

Os componentes acima citados criam juntos Inteligências Artificiais que podem interagir de maneira eficaz e inteligente (na forma lógica, não emocional) com o mundo ao seu redor. A **Inteligência Artificial Generativa (IA Generativa)** é uma tecnologia que se baseia em modelos de *Deep Learning* treinados por meio de grandes conjuntos de dados do Big Data ou por meio de coleta automática (e nem sempre autorizada) para criar conjuntos de dados que criam e geram novos conteúdos ou dados (os dados sintéticos). Temos três tipos a seguir: (i) **Redes Generativas Adversariais (GANs):** compostas por duas redes neurais competindo entre si. Uma é a rede geradora, que possui a capacidade de geração de dados, e a segunda é a rede discriminadora, que faz uma avaliação dos dados da rede geradora; (ii) **Redes Neurais Recorrentes (RNNs):** tipos de redes que são capazes de gerar sequências de dados como textos, música ou código, utilizando a aprendizagem de padrões em dados existentes para prever e gerar novas sequências; (iii) **Redes Neurais de Transformadores (ex: GPT):** semelhante às RNNs, essas redes também são capazes de gerar sequências de dados, mas utilizam uma arquitetura diferente.

Além dos Algoritmos temos o Aprendizado de Máquina e as Redes Neurais, que conforme lecionam Melo e Oliveira:

O nome “Rede Neural” deriva da estrutura neural do cérebro, na terceira página do trabalho de McCulloch e Pitts de 1943 podemos encontrar a expressão “nervous nets”, portanto, é uma expressão inventada por neurofisiologistas e matemáticos que tentaram simular e modelar os neurônios do cérebro.<sup>44</sup>

Esse trabalho citado, de 1943, denominado:

*A Logical Calculus of Ideas Immanent in Nervous Activity* de autoria de Walter Harry Pitts Jr. E Warren Sturgis McCulloch “é considerado tanto o primeiro trabalho de IA, como também a primeira proposta de modelar matematicamente o neurônio, ou seja, este é o trabalho seminal (inaugural) sobre Redes Neurais (RNA).<sup>45</sup>

---

<sup>44</sup> MELO JÚNIOR, Fábio Antero Pulpa; OLIVEIRA, Mauri Aparecido de. **Backpropagation e Redes Neurais – Volume 1: MLP, RPROP, QPROP, LMBP, DELTA-BAR-DELTA, MATLABR E PYTHONTM.** Rio de Janeiro: Editora Ciência Moderna Ltda, 2024, p. 7.

<sup>45</sup> *Ibidem*, p. 7.

É complexo desenvolver uma IA, e a pesquisa necessita expor basicamente uma explanação de questões básicas da parte técnica ao leitor para que se entenda como desenvolver mecanismos de explicação e explicabilidade se faz imprescindível para essa pesquisa expor alguns fatores fundamentais.

Um deles é a lógica básica expressa na forma algébrica que:

Em 1847 por George Boole publicou o livro com título *The Mathematical Analysis of Logic*, que desempenhou papel importante na construção da lógica matemática moderna. Em 1854 Boole publica o trabalho que é considerado sua obra-prima a *An Investigation os tje Laws of the Thought*, em que desenvolve uma estrutura para representar a lógica na forma algébrica. Uma proposição simples não possui qualquer outra proposição como uma parte componente de si mesma e é representada por uma letra minúscula, por exemplo:

**p:** Aristóteles é escravo

**q:** O homem é amoral

**r:** O número 274.207.281-1 é primo ...

Sejam **p** e **q** proposições, em que cada uma pode ser verdadeira ou falsa. Ou seja, o valor lógico da proposição **p** é verdade se **p** é verdadeira. O valor lógico de **p** é falso se **p** é uma falsa. O mesmo acontece com **q**, o valor lógico da proposição **q** é verdade se **q** é verdadeira. O valor lógico da proposição **q** é falso se **q** é uma falsa.<sup>46</sup>

Para entender a Rede Neural Artificial, antes será preciso trazer para essa dissertação o conceito de Perceptron, como sendo “a representação mais simples de uma rede neural, de tal forma que incorpora as pares básicas do neurônio em sua arquitetura”.<sup>47</sup>

Os autores explanam sobre a utilização do Perceptron, afirmam que:

Um perceptron pode ser utilizado para classificar padrões de entrada em duas classes. Sendo a principal tarefa a ser realizada no treinamento do perceptron em seu processo de aprendizagem constituir-se em ajustar o vetor de pesos  $w$  utilizando uma regra de aprendizagem adaptativa, denominada algoritmo de convergência do perceptron.<sup>48</sup>

A obra supramencionada<sup>49</sup> nos expõe que “o procedimento de aprendizagem consiste em duas partes, a propagação adiante e a propagação para trás.”, sendo a propagação o fluxo do algoritmo na utilização dos dados que foram imputados, e:

<sup>46</sup> MELO JÚNIOR, Fábio Antero Pulpa; OLIVEIRA, Mauri Aparecido de. **Backpropagation e Redes Neurais – Volume 1: MLP, RPROP, QPROP, LMBP, DELTA-BAR-DELTA, MATLAB E PYTHON**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2024, p. 92.

<sup>47</sup> *Ibidem*, p. 95.

<sup>48</sup> *Ibidem*, p. 152.

<sup>49</sup> HAYKIN, Simon. **Redes Neurais: princípios e prática**. Tradução: Paulo Martins Engel. 2.ª ed., Porto Alegre: Bookman, 2001, p. 28.

O algoritmo de *backpropagation* é utilizado para aprender quais são os pesos de uma rede neural de múltiplas camadas. Para isso é utilizado o método do gradiente descendente para minimizar a soma dos erros quadrados entre valores de saída da rede e os valores desejados. Conceitualmente, ocorre uma propagação de ativações ao longo das sinapses para produzir uma saída e os erros são retropropagados para realizar as mudanças dos pesos.

A explanação acima tem o objetivo de expor não somente como é complexo o aprendizado de máquina, mas como a matemática envolvida na linguagem permite entendermos a imitação do humano pela máquina.

Com efeito, as Redes Neurais Artificiais exercem um papel significativo na condição de aprender e pensar para tomar decisões da Inteligência Artificial, sendo “também referidas na literatura como neurocomputadores, redes conexionistas, processadores distribuídos, etc.” <sup>46</sup>, e possuem algumas capacidades cuja lista essa pesquisa informa para melhor entendimento do que uma Rede Neural pode oferecer, como: não-linearidade ou linearidade; mapeamento de entrada e saída de acordo com sua arquitetura; adaptabilidade resposta a evidências; informação contextual; tolerância a falhas; implementação a VLSI (*very-large-scale-integration*); universalidade de análise e projeto; analogia neurobiológica.<sup>50</sup>

E ensina que:

Existem dois modos de treinar um perceptron de múltiplas camadas ordinário (estático): o modo por lote e o modo sequencial. No modo por lote, a sensibilidade da rede é calculada para o conjunto inteiro de treinamento antes de ajustar os parâmetros livres da rede. No modo sequencial, por outro lado, os ajustes dos parâmetros são feitos após a apresentação de cada padrão do conjunto de treinamento.<sup>51</sup>

A relevância desse conceito para a pesquisa é sabermos o que são Redes Neurais e como são treinados seus perceptrons reside no conceito de que para que se criem parâmetros de governança da IA para fins e desenvolvimento, monitoramento e atualização de pesos e parâmetros, é preciso entender minimamente de questões básicas sobre os sistemas de Inteligência Artificial.

Cientes sobre como os dados para funcionamento dentro das Redes Neurais, ou seja, como o sistema ao receber os dados efetua o treinamento e aprendizado para proferimento da decisão, será possível com maior assertividade governar o uso da IA, e dar a Explicação solicitada por um indivíduo ou agente institucional (mercado ou governo).

---

<sup>50</sup> HAYKIN, Simon. **Redes Neurais: princípios e prática**. Tradução: Paulo Martins Engel. 2.ª ed., Porto Alegre: Bookman, 2001, p. 29-31.

<sup>51</sup> *Ibidem*, p. 28.

A pesquisa citou sobre as tarefas e aprendizado da máquina, e passo a tratar do tema das técnicas de Aprendizado da Máquina (*Machine Learning*) como uma das necessárias informações para governança da Inteligência Artificial, vez tais técnicas fazem, parte do desenvolvimento e a utilização em Inteligência Artificial.

Com efeito, a ciência das técnicas de aprendizado são parte da governança da IA, pois permitem análises e possibilidades de aferição para obtenção de decisões automatizadas mais confiáveis e transparentes, uma vez que as técnicas permitem monitorar, avaliar, atualizar e documentar respostas para concretização do Direito à Explicação.

O aprendizado de máquina (*machine learning*) é uma subárea da inteligência artificial que se dedica a desenvolver algoritmos que permitem que sistemas aprendam a partir de dados e tomem decisões com base neles.<sup>52</sup>

As principais técnicas de machine learning podem ser divididas em três categorias: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço.

No aprendizado supervisionado, os algoritmos aprendem a partir de um conjunto de dados de treinamento rotulados, onde a resposta correta é conhecida.

Essa técnica é amplamente utilizada em tarefas de classificação e regressão. Por exemplo, algoritmos como as máquinas de vetores de suporte (*Support Vector Machines – SVM*) e as redes neurais artificiais são comumente aplicados em reconhecimento de padrões e análise preditiva.

O aprendizado não supervisionado, por sua vez, lida com dados que não têm rótulos, buscando encontrar padrões ou agrupamentos dentro dos dados. Técnicas como a análise de agrupamento (*clustering*) e a decomposição de valor singular (SVD) são utilizadas para segmentação de clientes e redução de dimensionalidade, respectivamente.

O aprendizado por reforço é uma técnica onde os algoritmos aprendem a tomar decisões por meio da interação com o ambiente, buscando maximizar uma recompensa cumulativa. Essa abordagem tem se mostrado eficaz em jogos e sistemas autônomos, como na famosa aplicação do algoritmo AlphaGo, que venceu campeões humanos no jogo de Go.<sup>53</sup>

Inteligências Artificiais estão embarcadas em infinidade de aplicativos, softwares, sistemas computacionais, empresas, instituições públicas e privadas, pesquisa científica, sistema financeiro, segurança, securitização, saúde, educação, marketing, escritórios e

---

<sup>52</sup> HAYKIN, Simon. **Redes Neurais: princípios e prática**. Tradução: Paulo Martins Engel. 2.<sup>a</sup> ed., Porto Alegre: Bookman, 2001, p. 29-31.

<sup>53</sup> SILVER, David *et al.* Mastering the game of Go with deep neural networks and tree search. **Nature**, v. 529, n. 7587, p. 484-489, 2016.

profissões diversas, que não conseguiremos descrever aqui por que ela está inserida em quase toda parte ou relação de nossas vidas, mesmo quando não estamos utilizando tecnologias como computadores, tablets, smartphones e outros devices ou gadgets.

A Inteligência Artificial tem, portanto, redes neurais e aprendizado de máquina, o que permite que ela possa pensar, sendo essa habilidade denominada de pensamento computacional, conforme:

O pensamento computacional representa uma habilidade fundamental do século XXI, que vai além da programação, envolvendo decomposição de problemas, reconhecimento de padrões e abstração – elementos essenciais para o desenvolvimento de sistemas de IA.<sup>54</sup>

Alguns estudiosos esclarecem que “A integração entre pensamento computacional e IA ocorre quando problemas complexos são decompostos em subproblemas gerenciáveis, permitindo a criação de soluções algorítmicas escaláveis”.<sup>55</sup>

Outros firmam que: “O pensamento computacional na IA manifesta-se através de três dimensões principais: conceitos computacionais (sequências, loops, paralelismo), práticas computacionais (iteração, depuração) e perspectivas computacionais (expressão, conexão)”.<sup>56</sup>

Em complemento a esse raciocínio, “A essência do pensamento computacional na IA” está na capacidade de reformular problemas aparentemente difíceis em problemas que sabemos resolver, seja por redução, incorporação, transformação ou simulação”<sup>57</sup>, pois “O pensamento computacional fornece a estrutura mental necessária para traduzir problemas do mundo real em representações que um sistema de IA pode processar e resolver eficientemente”.<sup>58</sup>

E para completar o entendimento sobre a paridade entre o pensamento humano, expresso através da linguagem, temos em sistemas inteligentes aqui estudados as LLMs (*Large Language Models*), com sua capacidade de compreender e gerar linguagem humana, têm um papel importante na explicação de decisões automatizadas de inteligências artificiais, tornando-as mais transparentes e compreensíveis para os usuários.

---

<sup>54</sup> WING, Jeannette M. Computational thinking and thinking about computing. **Philosophical Transactions of the Royal Society a Mathematical, Physical and Engineering Sciences**, v. 378, p. 3717-3725, 2023, p. 3718.

<sup>55</sup> GROVER, Shuchi; PEA, Roy. Computational Thinking in K–12: A Review of the State of the Field. **Educational Researcher**, v. 42, n. 1, p. 40, 2024.

<sup>56</sup> BRENNAN, Karen; RESNICK, Mitchel. **New Frameworks for Studying and Assessing the Development of Computational Thinking**. American Educational Research Association, 2024, p. 15. Available at: <https://scratched.gse.harvard.edu/ct/files/AERA2012.pdf>. Access in: 28 nov. 2024.

<sup>57</sup> PAPERT, Seymour; SOLOMON, Cynthia. Twenty Things to do with a Computer. **Artificial Intelligence Education**, v. 15, n. 3, p. 122, 2024.

<sup>58</sup> YADAV, Aman; STEPHENSON, Chris; HONG, Hai. Computational Thinking for Teacher Education. **Communications of the ACM**, v. 60, n. 4, p. 55-62, 1.º apr. 2017.

As *Large Language Models* são o “Estado da Arte” do Processamento de Linguagem Natural, como ensinam Russel e Norvig, para quem “O aprendizado profundo e a aprendizagem por transferência avançaram notavelmente o estado da arte para o PNL”.<sup>59</sup>

Utilizar LLMs para gerar explicações claras, concisas e personalizadas, podemos aumentar a transparência e a confiança nas decisões automatizadas, promovendo uma IA mais responsável e ética. É crucial integrar as LLMs de forma estratégica, considerando as necessidades dos usuários e as características específicas do sistema de IA em questão.

Evitando uma conceituação formal e estática, a Inteligência Artificial é o conjunto de disciplinas conjuntas (matemática, estatística, engenharia, computação, psicologia, Direito) que se conectam e interagem para a obtenção de resultados ou decisões que buscam proximidade com as decisões humanas.

O cumprimento de preceitos e normas legais na utilização de tecnologias existentes e futuras para que a ética e não discriminação ocorram, ou seja, se os dados utilizados na base de dados não são objeto de tomada de decisões automatizadas opacas ou não explicáveis, torna sua revisão por humanos prejudicada.

A reflexão colacionada tem correlação com o intercepto adicionado em uma equação linear, ou seja, é o parâmetro adicional na Rede Neural), usada para ajustar a saída de informações (decisões) junto da soma ponderada das entradas (*inputs*) para o neurônio (humano) e “neuron” (Inteligência Artificial), ou seja, o peso padrão que a Inteligência Artificial recebeu durante esse processo de aprendizado, cujos resultados decisórios não aferíveis ou passíveis de modificação (após aferição de opacidade, por exemplo), tornam a utilização da Inteligência Artificial um risco para a sociedade e para a economia.

É possível, dessa maneira, obter explicação de decisões automatizadas quando temos controles preventivos, boas práticas e implementação de programas (Frameworks), governança na utilização dessas ferramentas computacionais, iniciando tal tarefa quando o uso de dados coletados da vida privada, dados pessoais de identificação e dados corporativos que subsidiam as decisões não constituam máculas aos Direitos Fundamentais, Sociais e Econômicos, cada vez mais utilizados no perfilamento individual de grupos e de empresas, e que predeterminam os resultados exarados (decisões automatizadas, *Outputs*) pela Inteligência Artificial.

Ainda que oscilem entre visões otimistas e pessimistas sobre o impacto econômico, social, político e jurídico, diversos estudiosos convergem ao afirmar que a revolução homem-

---

<sup>59</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022, p. 793.



máquina mudará os rumos das relações humanas e tudo que dela decorre.<sup>60</sup>

Os meios e as características comportamentais utilizados pela Inteligência Artificial (por meio de dados obtidos com vigilância e invasão de privacidade), bem como características informadas em cadastros e tratamento de dados, ocasionam vieses discriminatórios, modulação de comportamento.

Conforme Chiarello e Garcia: “há que se entender a inteligência artificial minimamente para ser possível fazer análises mais profundas sobre seu futuro”.<sup>61</sup>

Em termos de mercado, Data Brokers e *Big Techs* podem manipular mercados e indivíduos ou grupos, ofertar produtos e serviços com preços dispares, segregando dessa forma indivíduos ou grupos de maneira a aplicar as tecnologias em favor da desigualdade humana e econômica.

Os sistemas de Inteligência Artificial possuem diferentes formas de atuação e expressão, ou seja, algumas aprendem, outras apenas auxiliam de forma a cumprir funções predeterminadas em seu desenvolvimento e programação.

A compreensão das diferentes espécies de Inteligência Artificial (IA) é crucial para o avanço tecnológico e a implementação ética destas tecnologias.

A Inteligência Artificial está em nosso dia a dia, e a pesquisa propõe um estudo de como humanos interagem com essa tecnologia em cada decisão humana, por meio de comportamentos, atitudes ou atividades exercidas na esfera particular considerando suas disponibilidades de informar tais decisões pessoais para aparelhos eletrônicos, aplicativos, organizações etc., transformando praticamente toda decisão humana em informação que é inserida em sistemas inteligentes.

Vainzof expõe nossa interação diária da seguinte forma:

A Inteligência Artificial (IA) já está incorporada em nosso cotidiano pessoal e profissional, com aplicações das mais diversas, como nos resultados de busca da Internet, no conteúdo que nos é apresentado na *timeline* das redes sociais, recomendações de filmes e séries nos serviços de streaming, assistentes virtuais, rotas de trânsito, ações no comércio eletrônico, como personalização, automação de atendimento e agilidade nas entregas, score de crédito e antifraude, classificação e seleção de candidatos a empregos, *chatbot* para atendimento de consumidores, medicina diagnóstica, procedimentos

---

<sup>60</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Desafios Internacionais da Aplicação da Inteligência Artificial no Direito. **Revista Justiça de Direito**, v. 35, n. 2, p. 7, mai./ago. 2021. Disponível em: <https://seer.upf.br/index.php/rjd/article/view/13040/114116181>. Acesso em: 24 abr. 2024.

<sup>61</sup> *Ibidem*, p. 9.

cirúrgicos assistidos por robô, carros e armas autônomas, apenas para citar exemplos.<sup>62</sup>

Importa considerar para a pesquisa a correlação da cognição humana com a Inteligência Artificial para o tema da formação e tomada de decisão, e como ela se compara à artificial.

Sobre a interação humano-máquina é verdade que o julgamento humano e a automação se entrelaçam de formas complexas do nosso cotidiano, que na Sociedade da Informação é constante.

A pesquisa sobre tomada de decisão procurou esclarecer a influência da forma como tomamos nossas decisões, e como tais decisões transformam-se em dados que uma vez inseridos em sistemas de Inteligência Artificial resultam em decisões automatizadas viciadas ou inconsistentes, comumente relatadas como visões algoritmos.

O artigo científico de junho de 2023 chamado “A Inteligência Artificial versus a Inteligência Humana”<sup>63</sup> propõe uma análise comparativa entre as duas formas de inteligência, destacando suas vantagens e limitações. O autor defende que a inteligência artificial não pode substituir a humana, mas sim complementá-la e ampliá-la, desde que respeitados os princípios éticos e legais que regem a sociedade. Apresenta também alguns exemplos de aplicações práticas da inteligência artificial como diagnóstico médico, reconhecimento facial e assistentes virtuais, bem como os desafios e riscos envolvidos, como vies, privacidade e responsabilidade, e conclui que a inteligência artificial é uma ferramenta poderosa para o desenvolvimento humano, mas que requer uma constante reflexão crítica.

Os métodos de aprendizado de máquina constituem uma subdivisão fundamental da inteligência artificial (IA), proporcionando a base para que as máquinas aprendam a partir de dados, identifiquem padrões e tomem decisões com mínima intervenção humana. Os tipos principais de aprendizado de máquina são classificados em três categorias: aprendizado supervisionado, não supervisionado e por reforço.

Para entendermos a dinâmica cognitiva da inteligência artificial foi preciso pesquisar e entender a cognição humana, tendo como base algumas inferências da Psicologia Cognitiva,

---

<sup>62</sup> VAINZOF, Rony; SERAFINO, Danielle; STEINWASCHER, Aline. Aplicações e Regulação da Inteligência Artificial. In: VAINZOF, Rony *et al.* **Legal Innovation: o Futuro do Direito e o Direito do Futuro**. São Paulo: Thomson Reuters, 2022, p. 243-263.

<sup>63</sup> MENEZES, Marcos Antonio de. A Inteligência Artificial versus a Inteligência Humana. **Saber Humano**, v. 13, n. 22, p. 220-239, jan./jun. 2023. Disponível em: <https://saberhumano.emnuvens.com.br/sh/article/view/576/590>. Acesso em: 29 abr. 2024.

“que é uma das disciplinas constituintes da Ciência Cognitiva”<sup>64</sup>, cuja conexão com a ciência da computação, uma vez que “foi influenciada pelo advento do computador digital”<sup>65</sup>, ambas remontam seu início em períodos aproximados da década de 1950 (mais precisamente 1956, quando se cunhou o termo Inteligência Artificial).

Neufeld, Brust e Stein informam que o prestigiado autor da primeira obra sobre Psicologia Cognitiva criou tal psicologia do ponto de vista da Teoria do Processamento da Informação, referindo-se a ela:

Como a psicologia que se refere a todos os processos pelos quais um *imput* (entrada) sensorial é transformado, reduzido, elaborado, armazenado, recuperado e usado. Neste livro, ele traz a noção de códigos cognitivos utilizados para possibilitar os processos mentais. Para ele os processos cognitivos criam códigos que são úteis para as pessoas, utilizáveis no seu dia a dia. Neisser também lança bases da noção de conhecimento, que seria desenvolvido individualmente por meio desses códigos cognitivos, tornando-os hábeis a trabalhar, tomar decisões, estudar, jogar futebol, etc.<sup>66</sup>

Causas e decisões que produzem efeitos e consequências são parte da Psicologia Cognitiva, e também:

Características como observação sistemática, ou seja, a busca por dados no empírico a mensuração ou quantificação dos resultados, e o controle e manipulação de variáveis com o intuito de estabelecer relações de causa e efeito são fundamentais na utilização de métodos experimentais até hoje na Psicologia Cognitiva. As representações proposicionais expressam ideias, um conteúdo abstrato, como por exemplo, a ideia de lar. As representações análogas expressam ideias concretas, análogas a como as percebemos os objetos na realidade, por exemplo, o tradicional desenho de uma casa. As representações análogas podem ser de origem visual, gustativa, olfativa, auditiva e proprioceptiva.<sup>67</sup>

As autoras do artigo científico citado acima afirmam ainda que a “Ciência Cognitiva é uma ciência multidisciplinar que busca compreender o processo de cognição, tanto de seres humanos quanto de máquinas pensantes”, demonstrando sua convergência e importância para essa pesquisa.<sup>68</sup>

---

<sup>64</sup> NEUFELD, Carmem Beatriz; BRUST, Priscila Goergewin; STEIN, Lilian Milnitsky. Bases Epistemológicas da Psicologia Cognitiva Experiental. **Psicologia: Teoria e Pesquisa**, v. 27, n. 1, p. 103, jan./mar. 2011. Disponível em: <https://www.scielo.br/j/ptp/a/DJVCxtgRTDRG6y45VJHbgzS/?format=pdf&lang=pt>. Acesso em: 29 abr. 2024.

<sup>65</sup> *Ibidem*, p. 103.

<sup>66</sup> *Ibidem*, p. 103.

<sup>67</sup> *Ibidem*, p. 103.

<sup>68</sup> *Ibidem*, p. 104.

É preciso cautela ao pesquisar neurociências, e esta pesquisa procura clareza ao invés de neuro mitos (mitos sobre neurociência) como o creditamento dos indivíduos que:

[...] acreditam que a explicação de um fenômeno psicológico envolve fatos que são relacionados com o cérebro do que quando envolve fatos que não o são, mesmo quando estas informações neurocientíficas são irrelevantes para a explicação do fenômeno em questão. [...] Somente se pode aferir que há evidências científicas para um fato após muitas replicações de um mesmo resultado por diferentes grupos e laboratórios ao redor do mundo, em pesquisas que investigam milhares de indivíduos, ou por meio de técnicas como meta-análises, isto é, estudos que analisam conjuntamente os achados de diversos experimentos publicados separadamente.<sup>69</sup>

Temas conectados como evidenciado porque IA e Cognição Humana possuem similaridades quanto aos *inputs* (entradas), a pesquisa verificou que mais presente essa conexão entre as inteligências (artificial e cognitiva humana).

Quanto ao “uso que as pessoas fazem de teorias (da Física) populares sobre o mundo físico, também se utilizam de teorias psicológicas populares para explicar o comportamento dos indivíduos, seus desejos e intenções”<sup>70</sup>, demonstrando a aquisição de conhecimento por meio de dados que coletamos diariamente.

O Capitalismo de Plataformas coleta e mensura as dinâmicas de interação realizadas *online* pelos usuários, consistentes:

Em processos algorítmicos de captura, análise e utilização de informações psíquicas e emocionais extraídas de dados e ações dos indivíduos em plataformas digitais (redes sociais, aplicativos, serviços de streaming, plataformas de compartilhamento e/ou consumo de conteúdo audiovisual etc.).<sup>71</sup>

O julgamento, descrito no livro “Ruído”<sup>72</sup>, explora um aspecto menos conhecido, mas igualmente importante dos erros de julgamento, qual seja, a variabilidade indesejada nas

<sup>69</sup> EKUNI, Roberta; POMPÉIA, Sabine. O impacto da divulgação científica na perpetuação de neuromitos na educação. **Revista de Biologia**, v. 15, n. 1, p. 22, 2016. Disponível em: <https://www.revistas.usp.br/revbiologia/article/view/114837/112576>. Acesso em: 24 abr. 2024.

<sup>70</sup> JOU, Graciela Inchausti de; SPERB, Tania Mara. Teoria da Mente: diferentes abordagens. **Psicologia: Reflexão e Crítica**, v. 12, n. 2, 1999. Disponível em: <https://www.scielo.br/j/prc/a/H7Bb5zCwRFqfLK8BNrf7ZGS/?lang=pt>. Acesso em: 24 abr. 2024.

<sup>71</sup> BARRETO JÚNIOR, Irineu Francisco; MOLINA, Fernanda Zampieri. Capitalismo de plataforma: a ameaça ao direito à autodeterminação informativa na Sociedade da Informação. **Revista Brasileira de Estudos Políticos**, Belo Horizonte, n. 125, p. 249, jul./dez. 2022. Disponível em: <https://pos.direito.ufmg.br/rbep/index.php/rbep/article/view/852/657>. Acesso em: 24 abr. 2024.

<sup>72</sup> KAHNEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R. **Ruído**. Tradução: Cássio de Arantes. 1.ª ed. Rio de Janeiro: Onketiva, 2021.

decisões que deveriam ser consistentes. Nela os autores argumentam que, além dos vieses cognitivos bem documentados, o “ruído” é uma fonte significativa de erros.

O “ruído” refere-se à variabilidade nas decisões que não se deve às informações novas ou às mudanças nas circunstâncias, mas sim aos fatores irrelevantes e imprevisíveis. Por exemplo, dois médicos podem dar diagnósticos diferentes para o mesmo paciente ou dois juízes podem dar sentenças diferentes para o mesmo crime, devido ao ruído em seu julgamento. Uma “auditoria de ruído”, que é um método para medir essa variabilidade e identificar onde o ruído está afetando as decisões.

Além disso, os autores fornecem estratégias práticas para reduzir o ruído, como técnicas de “higiene na tomada de decisões”, que podem ajudar a melhorar a qualidade e a consistência das decisões.

O humano colhe consequências e resultados de suas escolhas refletidas em decisões. É possível entender que as predições (humanas) para julgar são feitas – diversas vezes – pelos indivíduos mediante observação do comportamento de terceiros, formando a concepção dos autores sobre Teoria da Mente:

Ao dizer que um indivíduo tem uma teoria da mente, queremos significar que o indivíduo atribui estados mentais a si próprio e aos outros (seja da mesma espécie ou outra). Um sistema de inferência desse tipo é, apropriadamente, visto como uma teoria, primeiro, porque, primeiro, porque esses estudos não são observáveis diretamente, e, segundo, porque o sistema pode ser usado para fazer predições, especialmente, sobre o comportamento de outros organismos.<sup>73</sup>

As formas de decidir e entender como são feitas as tomadas decisórias em humanos, e como tais decisões são relevantes e conectam-se com as decisões automatizadas, que em breve serão objeto de análise específica desta pesquisa.

A pesquisa traz reflexão sobre a cultura da tecnologia:

Como um fator determinante ao invés de um dos componentes do mundo cultural e, assim, marcam uma determinada época histórica – a que alude à cibercultura, bem como à economia digital – parece interessante tentar observar os atributos pertinentes ao novo contexto produzido pela *Sociedade da Informação ou Sociedade em Rede*.<sup>74</sup>

---

<sup>73</sup> PREMACK, David; WOODRUFF, Guy. Does the chimpanzee have a theory of mind? **The Behavioural and Brain Science**, v. 1, n. 4, p. 515, 1978. Disponível em: <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1E96B02CD9850016B7C93BC6D2FEF1D0/S0140525X00076512a.pdf/does-the-chimpanzee-have-a-theory-of-mind.pdf>. Acesso em: 29 abr. 2024.

<sup>74</sup> BARRETO JÚNIOR, Irineu Francisco; UELZE, Hugo Barroso. Cibercultura e o Poder Informacional na Esfera Pública: Crise do Paradigma Positivista no Direito Tributário. **Revista Prima Facie**, Universidade Federal da

A citação acima nos faz aferir na pesquisa como a Sociedade Datificada, transformada em dados, reflete diretamente no aprendizado da inteligência artificial e conseqüentemente nas decisões decorrentes dos *Outputs* delas.

A exposição acima demonstra a relevância e a conexão que a pesquisa tem com o *profiling*, ou perfilamento, ou seja, os padrões de dados de personalidade e comportamento humanos inseridos em sistemas de Inteligência Artificial, fazendo-a aprender muito sobre nós e, por conseqüência, nos expondo a decisões automatizadas que poderão ser prejudiciais a nós mesmos e a sociedade, sendo melhores analisadas adiante.

As decisões que tomamos em nossas rotinas diárias são objeto de inserção nos sistemas computacionais de Inteligência Artificial, uma vez que decisões são expressas em atos que significam nossa personalidade, construindo nossa identidade.

Desse modo, as decisões que tomamos são quase que imediatamente transferidas como dados pessoais informados aos sistemas de telecomunicação e tecnologia, e cujas preferências, lugares, compras, vendas, escolhas, trânsito, observação, viagens, acesso aos sítios eletrônicos e aplicativos transformam-se em objeto de análise preditiva e perfilamento.

Nas palavras de Byung-Chul Han:

Chamamos regime de informação a forma de dominação na qual informações e seu processamento por algoritmos e inteligência artificial determinam decisivamente processos sociais, econômicos e políticos. Em oposição ao regime disciplinar, não são os corpos e energias que são explorados, mas as informações e dados. Não é, então, a posse de meios de produção que é decisiva para o ganho de poder, mas o acesso a dados utilizados para vigilância, controle e prognóstico de comportamento psicopolíticos. O regime de informação, que se desenvolve em capitalismo de vigilância e que degrada os seres humanos em gado, em *animais de consumo e dados*.<sup>75</sup>

A permissão para tomar decisões com base em diversos critérios como, por exemplo, nossas experiências, tempo (rápido ou devagar), custo (decisão financeira), crenças (aquilo que acreditamos como o melhor caminho com fundamento em valores imputados em nossa mente) positivas ou negativas, vivenciais, traumas, consciente e subconsciente, que compõem a forma como pensamos, e conseqüentemente como escolhemos e nos comportamos em relação a nós mesmos e em relação aos outros, principalmente quando estamos diante de uma questão que

---

Paraíba, João Pessoa, v. 19, n. 40, p. 5, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/primafacie/article/view/51047>. Acesso em: 24 abr. 2024.

<sup>75</sup> HAN, Byung-Chul. **Infocracia: digitalização e a crise da democracia**. Tradução: Gabriel S. Philipson. Petrópolis: Vozes, 2022, p. 7.

demanda mais de uma opção alternativa para chegarmos ao resultado ou consequência da decisão tomada.

De acordo com Luiz Flávio Autran Monteiro Gomes e Carlos Francisco Simões Gomes, decisões e consequências:

As decisões podem ser classificadas de várias formas, tais como: a) simples e complexas; b) específicas ou estratégicas etc. As consequências advindas das decisões podem apresentar-se da seguinte maneira: a) imediata; b) curto prazo; c) longo prazo; d) combinação das formas anteriores (impacto multidimensional).<sup>76</sup>

A decisão é inferida pelo ser-humano por meio de uma coleta de informações, cuja régua de importância é medida de acordo com suas habilidades, crenças, cognições, dados, tempo, local, sendo instado a analisar e escolher a melhor alternativa para solucionar um dado problema, cuja solução demandará um conjunto de conhecimento, seu estado de desenvolvimento humano, intelecto, etc.

Na obra “Rápido e Devagar: Duas Formas de Pensar”<sup>77</sup> são exploradas as velocidades das tomadas de decisão. O autor descreve essas duas formas como o “Sistema 1” e o “Sistema 2”.

No Sistema 1 o funcionamento é rápido, automático, intuitiva e emocional. Esse sistema é responsável por nossas reações imediatas e decisões rápidas sem reflexão consciente, como no caso de aquisição de produtos em oferta nas plataformas, microtransações e *lootboxes* dos games, ideias manifestadas em redes sociais que geram engajamento automático, principalmente em temas polêmicos. Kahneman usa esses conceitos para discutir como esses sistemas influenciam nosso pensamento e julgamento. Ele examina vários vieses cognitivos que resultam da interação desses dois sistemas, como o excesso de confiança, a aversão à perda e o efeito do ponto cego. Kahnemann também explora as implicações desses vieses e sistemas de pensamento em áreas como economia comportamental e tomada de decisão política e empresarial.

O autor conclui as duas maneiras distintas pelas quais o cérebro humano processa informação e toma decisões. A equação proposta por Kahneman entendo como sendo: “Sistema

---

<sup>76</sup> GOMES, Luiz Flávio Autran Monteiro; GOMES, Carlos Francisco Simões. **Princípios e Métodos da Tomada de Decisão**. 6.<sup>a</sup> ed. São Paulo: Atlas, 2019, p. 1.

<sup>77</sup> KAHNEMAN, Daniel. **Rápido e Devagar: duas formas de pensar**. Tradução: Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.

1: Informação – Complexidade + Velocidade + Pensar – Atenção = Decisão. Sistema 2: Informação + Complexidade – Velocidade + Pensar + Atenção = Decisão”.

Por ser rápido e basear-se em impressões e intuições, demonstrando o citado autor que o Sistema 1 pode ser suscetível a vieses e erros sistemáticos e, dessa forma, ao informamos como nos comportamos na rede mundial de computadores, por meio de plataformas e redes sociais, tais decisões rápidas podem ser utilizadas como meio de perfilamento para que sistemas de Inteligência Artificial ofertem produtos e serviços em momentos de vulnerabilidade emocional.

No Sistema 2 da obra de Kahnemann, temos um sistema de decisão que é mais lento, deliberativo, lógico e consciente. Este sistema é invocado para tarefas que requerem atenção como resolver problemas matemáticos difíceis ou tomar decisões importantes que necessitam de raciocínio detalhado. Com efeito, esse sistema requer esforço consciente e é mais confiável, mas é também mais lento e requer mais energia mental, demandando, portanto, maior segurança para o indivíduo.

As decisões automatizadas proferidas por Inteligência Artificial utilizam-se do aprendizado que a máquina teve com os padrões e análises preditivas, bem como do perfilamento de dados comportamentais e pessoais das pessoas conseguem decidir por meio de cálculos que tipos de decisões estamos tomando, se fundadas no sistema 1 ou no sistema 2, decidindo o momento exato de nossas vulnerabilidades quanto à tomada de decisões que precisam ser tomadas de forma rápida, sem critérios seguros de compreensão e tomada de decisão.

A teoria da decisão tem como pressuposto que os indivíduos são capazes de expressar suas preferências básicas, são racionais quando enfrentam decisões simples (Sistema 1 de Kahneman), mas sugere metodologias que foram desenvolvidas para auxiliar na resolução de problemas decisórios mais complexos (Sistema 2 proposto por Kahneman).

A tomada de decisão é uma atividade essencial na vida cotidiana, influenciando desde escolhas simples, como o que comer no café da manhã, até decisões complexas como investimentos financeiros ou escolhas de carreira. As neurociências do comportamento têm desempenhado um papel fundamental na compreensão dos processos subjacentes à tomada de decisão humana revelando os mecanismos neurais, os fatores influentes e as implicações práticas desse fenômeno complexo.

A relação entre decisões humanas e decisões automatizadas é complexa e interdependente, com cada uma influenciando a outra de maneiras significativas.



As decisões humanas desempenham um papel crucial na formação, no funcionamento e na evolução dos sistemas de decisão automatizada.

Os humanos são responsáveis por definir os objetivos e critérios que guiam as decisões automatizadas. Por exemplo, em um sistema de aprovação de crédito, os humanos determinam quais fatores (renda, histórico de crédito, etc.) são relevantes e como devem ser ponderados. Humanos projetam os algoritmos, selecionam os dados de treinamento e ajustam os parâmetros dos sistemas de decisão automatizada. Essas escolhas podem incorporar vieses humanos conscientes ou inconscientes, influenciando o comportamento do sistema.

Em muitos casos, humanos supervisionam as decisões automatizadas, podendo intervir para corrigir erros, ajustar parâmetros ou lidar com casos complexos que exigem julgamento humano.

As decisões humanas, na forma de feedback sobre o desempenho do sistema, são essenciais para o aprendizado e a adaptação dos algoritmos. O feedback humano permite que o sistema refine suas decisões e se torne mais preciso e eficaz ao longo do tempo.

A correlação estabelecida neste capítulo da pesquisa demonstra como nossas decisões são processadas e transformadas em dados estruturados para inserção em sistemas e softwares de Inteligência Artificial, que precisam ser explicados (escopo do Direito à Explicação) quando nos deparamos com decisões de máquinas que direta, ou indiretamente influenciam a sociedade.

## 1.2 DECISÕES E NEUROECONOMIA

A pesquisa analisou o campo da neurociência econômica e deparou-se com o *trade off* entre dados pessoais e serviços “gratuitos”, utilizado como modelo de coleta de dados pessoais para perfilamento e precificação, sendo um dos motores da economia digital que permitem que nossas decisões e direitos de personalidade sejam identificados, coletados e inseridos em sistemas que decidem com fundamento em padrões culturais, necessidades e vivências do ser humano.

Bagnoli aponta:

Diariamente, todos os indivíduos se deparam com situações em que devem fazer uma escolha. Toda escolha, em regra, é uma relação de perde e ganha, onde geralmente ao se optar por algo, assume-se um ônus ou risco do ônus pela escolha que foi feita na expectativa do bônus. No campo econômico é igual, e ao se optar por uma vantagem econômica, renuncia-se a outra ou assume-se o adimplemento da vantagem auferida. O que importa na verdade, é que se tenha a plenitude das informações para se fazer as melhores escolhas, pois inviavelmente, *trade-offs* são o dia a dia de todos os indivíduos. Nos

tempos de informação digital, uma quantidade enorme de serviços *online* contribui para o seu crescimento. Curiosamente, grande parte desses serviços são “gratuitos”, de modo que o indivíduo não precisa “pagar por eles”, ou seja, despende recursos financeiros para ter acesso a tais serviços. Mas na verdade, “não tem almoço grátis”, e o indivíduo está pagando um preço muito alto pelo serviço, e como não possui informação, deste trade-off muitas vezes o resultado é negativo. O pagamento que o indivíduo geralmente faz para ter os serviços gratuitos é entregar suas informações pessoais, expondo a sua privacidade. Essa é a “moeda de troca”, o “preço a ser pago”, pelo serviço “grátis”. A situação, entretanto, é mais delicada que se imagina.<sup>78</sup>

No capítulo anterior a pesquisa averigou como a complexidade da relação da Inteligência Artificial e da decisão humana, bem como elas possuem uma correlação de causalidade, vez que a proposta de Alan Turing conforme anteriormente citado – foi justamente idealizar uma máquina que pensasse, e tomasse decisões, da mesma forma que um humano.

As decisões, conforme César, Boggio e Campanhã:

Pessoas tomam decisões, todos os dias, sobre todas as coisas. Decidem sobre as atividades do dia, sobre o que deverão comer, sobre suas finanças pessoais, sobre a educação própria e a de seus filhos, entre outros aspectos. Todas essas decisões têm em comum: a existência de um problema (o que se deve fazer em uma dada situação), a existência de alternativas de resposta (os caminhos possíveis para a solução do problema em questão) e a aplicação de mecanismos (conscientes ou não) para escolher qual deverá ser a alternativa que represente a melhor solução para o problema inicial.<sup>79</sup>

César, Boggio e Campanhã, também resolveram discorrer em seu artigo científico sobre a relação e proximidade existente entre tecnologia, neurociência, psicologia e economia comportamental, surgindo a neuroeconomia, como sendo o estudo interdisciplinar que visa entender por meio de “pontos de partida” referenciais de diferentes áreas do conhecimento compreender a tomada de decisão por meio de modelos de processamento cognitivo:

Com o avanço tecnológico e a aproximação de áreas como a neurociência, psicologia, e a Economia Comportamental, importantes contribuições acerca da compreensão do funcionamento cerebral vêm sendo feitas, possibilitando a mensuração quantitativa e a correlação entre comportamento, sentimento e suas bases neurobiológicas subjacentes. Segundo Shiv *et al.* (2005) a neurociência pode ajudar, por meio de formas alternativas, a desvendar fenômenos contábeis, oferecendo aos estudos de tomada de decisão a possibilidade de localizar atividade neuronal associada a vários fenômenos,

<sup>78</sup> BAGNOLI, Vicente. **Direito econômico e concorrencial – O Poder Econômico e seus Limites Jurídicos**. 9.<sup>a</sup> ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2022, p. 523.

<sup>79</sup> CÉSAR, Ana Maria Roux Valentini; BOGGIO, Paulo; CAMPANHÃ, Camila. **Neuroeconomia: uma visão geral sobre o tema**, p. 2. Disponível em: <https://www.researchgate.net/publication/307575716>. Acesso em: 11 nov. 2024.

além de oferecer testes diretos para a investigação das teorias existentes e de novas teorias. A neuroeconomia surgiu como um esforço interdisciplinar para tentar preencher lacunas na compreensão do comportamento na tomada de decisão. A integração das diferentes disciplinas, suas diferentes abordagens teóricas e técnicas vêm auxiliando no conhecimento sobre a Tomada de Decisão Humana para a construção de modelos mais precisos (Sanfey et al, 2006). Um conhecimento mais aprofundado sobre a especialização funcional e sobre a contribuição de cada região cerebral em diferentes tarefas pode contribuir para a substituição de categorias comportamentais econômicas estabelecidas em função da compreensão do maior ou menor envolvimento das regiões cerebrais no processamento afetivo e cognitivo, ajudando a revelar e definir as possíveis causas dos desvios do comportamento da teoria normativa.<sup>80</sup>

Os autores supramencionados<sup>81</sup>, em sua pesquisa, que as decisões são tomadas mais com fundamento em conhecimento individual em situações econômicas do que fundamentadas em bases financeiras realmente aferíveis, e que em situações cujas informações são disponibilizadas sobre as regras foram a base a tomada de decisão.

Dessa forma, o estudo aponta que indivíduos decidem mesmo sem compreender o significado da decisão, e, portanto, como ela afeta o resultado, leia-se:

A revisão da literatura apontou cerca de 30 tarefas cognitivas (jogos ou não) que se repetem em estudos sobre decisão em ambientes econômicos e que são muito diferentes das situações de decisão enfrentadas por gestores, seja pelos pequenos montantes envolvidos, seja pelos estímulos (cenas do dia a dia, faces, figuras geométricas etc) utilizados; estudos com essas tarefas permitem que se conclua sobre muitos aspectos cognitivos, mas não sobre situações reais de decisão baseadas em informações contábeis. Finalizando, o Jogo de Metas® usa como estímulo um conjunto de informações contábeis apresentadas em forma de gráficos, que são combinações semânticas que em situações organizacionais funcionariam como uma linguagem que precisaria ser decifrada por aqueles que decidem. No Jogo de Metas® viu-se que pessoas aprendem regras de decisão com base em informações disponíveis, mesmo sem compreender o significado das mesmas. Outros estudos poderiam ser feitos explorando esses achados, tendo em vista que as organizações empenham muito esforço na construção de relatórios gerenciais para subsidiar decisões.<sup>82</sup>

A utilização constante de tecnologia, afeta a tomada de decisão uma vez que o indivíduo é levado a decidir por estímulos imediatos e com base em informações disponibilizadas quase de instantaneamente, o que faz com que os usuários não considerem a

---

<sup>80</sup> CÉSAR, Ana Maria Roux Valentini; BOGGIO, Paulo; CAMPANHÃ, Camila. **Neuroeconomia: uma visão geral sobre o tema**, p. 3. Disponível em: <https://www.researchgate.net/publication/307575716>. Acesso em: 11 nov. 2024.

<sup>81</sup> *Ibidem*, p. 11.

<sup>82</sup> *Ibidem*, p. 10.

causalidade entre a interação e a economia “datificada” como um produto infinito de dados que serão objeto de treinamento, aprendizado e decisão tecnológica da Inteligência Artificial.

A influência gerada por redes sociais e plataformas de conteúdo, com anúncios que provocam os usuários a adquirir produtos e serviços, e também, a influências exercidas pela prosperidade midiática provocam um estado de euforia consumerista, que é apoiada pelas decisões automatizadas proferidas por sistemas que induzem ao consumo.

Coutinho, contribuindo para o argumento acima, reflete:

No entanto, mesmo indo de encontro a alguns autores, a sociedade de consumo, sinônimo de pós-modernidade, apresenta-se bastante próxima do supérfluo e da abundância, ao mesmo diante das imposições e modulações de comportamento características da IA no consumo. [...] Assim, para que se conclui acerca da sociedade de consumo estar por vezes atrelada à ideia de ostentação nos dias atuais, ou seja, na sociedade em rede, não demonstra ser um dispare por basear-se na utilização da IA, mais especificamente, na internet das coisas como meio de ofertar de maneira contínua, por diversos meios produtos e serviços, o que por consequência coloca o consumo como meta.<sup>83</sup>

A neuroeconomia utiliza-se de tecnologia denominada *Data Mining* que:

É o processo de descobrir padrões interessantes por meio de grandes quantidades de dados. Pode ser realizado em qualquer tipo de dado, desde sejam dados significativos para uma aplicação-alvo, tais como informações de bancos de dados, utilizando técnicas de recuperação de informação, inteligência artificial, reconhecimento de padrões e de estatística para procurar correlações entre diferentes dados que permitam adquirir um conhecimento benéfico para uma empresa ou indivíduo. Embora o termo seja relativamente novo, a tecnologia não é. [...] A cadeia de dados torna possível analisar as transações de rotina e recolher uma quantidade significativa de informações sobre indivíduos hábitos de compra e preferências. Desta forma, certas aplicações de tecnologia de data mining que envolvem manipulação de dados pessoais levantam preocupações em relação a privacidade. Vale destacar ainda que para alguns o data mining ou mineração. De dados é censurável, por ser instrumento de discriminação que mascara a intenção de prejudicar certos membros de determinada classe. [...] Ou seja, padrões de gerais podem ser usados para adivinhar dados confidenciais, e, mais que isso, podem levar a estereótipos e preconceitos. Se os padrões são baseados em raça e gênero, essa questão é bastante controversa.<sup>84</sup>

---

<sup>83</sup> COUTINHO, Diego Araújo. **O consumidor e modulação algorítmica de comportamento: a influência da Inteligência Artificial por meio de algoritmos no poder decisório do consumidor**. São Paulo: Dialética, 2021, p. 92.

<sup>84</sup> LOURENÇO, Flávia de Faria. Data Mining, Big Data e o Direito à Privacidade. In: BLUM, Renato Opice; ABRUSIO, Juliana; WALDER, Priscila Ungaretti de Godoy. **Temas relevantes de direito digital**. São Paulo: Editora IASP, 2020, p. 97.

A pesquisa citou o perfilamento como uma das ferramentas que explicam as decisões automatizadas, e na questão do consumo, temos nas ferramentas de IA não somente um sistema que pretende melhorar a jornada do consumidor, mas que poderá sobremaneira complementá-lo.

E nesse sentido continua Coutinho:

O perfilamento de abordagem ao consumidor, como bem apontam os autores, é refinado. Porém o que parece ser uma qualidade no. Tratamento para com o usuário, traz em si uma característica bastante específica da sociedade de consumo que é o estímulo ininterrupto ao consumo, ou seja, consumismo.<sup>85</sup>

Algo que poderá causar agravamento na tomada de decisão, e que sempre acelerada e motivada por impulsos neurológicos rápidos (tomada de decisão rápida, como exposto, no sistema 1 observada por Daniel Kahnemann) da vida em ambiente digital da sociedade da informação, pois as interações com sites, plataformas, redes sociais, e-commerce e aplicativos, instalados em aparelhos eletrônicos que utilizamos no cotidiano são captados, e tornam-se dados de rastreamento de preferências e de personalidade, que são o alimento de conhecimento para treinamento e aprendizado da máquina como expomos linhas atrás.

As conexões virtuais em demasia, provocam a coleta e interoperabilidade de dados que “o ambiente virtual, sempre repleto de janelas, possibilidades, acúmulo de atividades e canais de acesso, afastam o consumidor do que é preciso ter: consciência.”<sup>86</sup>

A Inteligência Artificial requer de conhecimento sobre algo ou alguém, por meio de informações que são geradas e inseridas para análise, e aprenda e profira decisões automatizadas, como lê-se do entendimento abaixo exposto por Gabriel Filho:

As pesquisas em IA demonstram que a inteligência artificial requer conhecimento. Em consequência, somos forçados a concluir que uma técnica de IA, é um método que explora o conhecimento, o qual deve ser representado de tal forma que tenha a capacidade de: -Fazer inferências ou. Generalizações; isso significa que não é necessário representar todo o conhecimento (holística) sobre um determinado fenômeno. Basta representar uma parte do conhecimento (heurística), pois uma boa técnica de IA será capaz de “adivinhar” as outras situações não representadas, possibilitando assim a utilização de uma quantidade menor de memória; -Ser de fácil compreensão para os usuários; -Ser facilmente atualizado para absorver as mudanças non meio ambiente; -Ser usado em diversas situações, dentro do contexto de

---

<sup>85</sup> LOURENÇO, Flávia de Faria. Data Mining, Big Data e o Direito à Privacidade. In: BLUM, Renato Opice; ABRUSIO, Juliana; WALDER, Priscila Ungaretti de Godoy. **Temas relevantes de direito digital**. São Paulo: Editora IASP, 2020, p. 98.

<sup>86</sup> *Ibidem*, p. 111.

aplicação; -Fornecer o resultado no tempo requerido, apesar. De seu grande volume de dados.<sup>87</sup>

Prossegue o autor citado acima, a uma distinção quanto aos tipos de conhecimento que a Inteligência Artificial adquire com os dados que lhe são fornecidos nos inputs, que podem consistir em tipos e classificações como de: (i) conhecimento declarativo; (ii) conhecimento procedural; (iii) conhecimento semântico; e (iv) conhecimento episódico.<sup>88</sup>

A neuroeconomia e o *neuromarketing*, segundo Fisher, Chin e Klitzman, são uma realidade do ambiente digital e, portanto, da Inteligência Artificial, sendo:

O *neuromarketing* utiliza princípios neurocientíficos para compreender as respostas emocionais e cognitivas dos consumidores, aplicando esses insights no desenvolvimento de estratégias de marketing mais eficazes. Quando associado à Inteligência Artificial, o *neuromarketing* potencializa a personalização e a eficiência das interações com os consumidores.<sup>89</sup>

A integração do *neuromarketing* com algoritmos de IA para análise comportamental, cuja análise de dados de *neuromarketing* em conjunto com algoritmos de IA permite o desenvolvimento de perfis de consumidores e usuários para campanhas.

Morin manifestando seu entendimento sobre o assunto:

Enquanto a combinação de IA e *neuromarketing* oferece uma capacidade sem precedentes de adaptação de estratégias de venda, levanta-se uma preocupação

---

<sup>87</sup> GABRIEL FILHO, Oscar. **Inteligência Artificial e aprendizagem de máquina: aspectos teóricos e aplicações**. São Paulo: Blucher, 2023, p. 42-43.

<sup>88</sup> Conforme o autor: “Existem vários tipos de conhecimento. Alguns são conhecimentos especializados, como os que se encontram nas bulas dos remédios ou embutidos num jogo eletrônico; outros são conhecimentos pessoais que usamos quase que inconscientemente nas tarefas do dia a dia. Os conhecimentos são classificados em: **Conhecimento Declarativo** – Quando você vê um carro passando e o descreve a um amigo, está se baseando non conhecimento declarativo que você possui. É o conhecimento mais fácil de acumular, mas também é o mais difícil de utilizar nos computadores. **Conhecimento Procedural** – O conhecimento procedural especifica passo a passo o que fazer e quando fazer, podendo ser representado. Em programas de vários modos, sendo o mais comum na forma de código sobre como fazer alguma coisa. Boa parte do sucesso da IA Simbólica advém da utilização desse tipo de conhecimento. As ações para validar um cartão de ponto ou para instalar um programa de computador envolvem a execução de procedimentos. **Conhecimento Semântico** – O conhecimento semântico é o mais difícil de se capturar, uma vez que se trata de um aglomerado de conceitos e relações entrecruzadas (semântico do grego *semantikós* “que assinala, que indica”, adjetivo 1. Relativo à significação. Nesse caso, significação dos símbolos). Os profissionais da informação passam grande parte de seu tempo explorando a mente dos especialistas, para capturar o real significado de suas ideias. **Conhecimento Episódico** – Alguns dos conhecimentos sobre o mundo não foram obtidos com o estudo de fatos específicos. Pense como a sua mente divaga enquanto você dirige o carro. A atividade de dirigir está associada ao espaço e ao tempo. Dessa maneira, você consegue chegar até sua universidade não porque tenha raciocinado de forma procedural sobre o ato de dirigir seu carro, mas porque reproduziu um conhecimento episódico acumulado. O conhecimento episódico é difícil de reconhecer e capturar devido à sua natureza pessoal e à sua tendência de estar interconectado com o conhecimento semântico.” GABRIEL FILHO, Oscar. **Inteligência Artificial e aprendizagem de máquina: aspectos teóricos e aplicações**. São Paulo: Blucher, 2023, p. 43-44.

<sup>89</sup> FISHER, Carl Erik; CHIN, Lisa; KLITZMAN, Robert. Defining neuromarketing: Practices and professional challenges. **Nature Reviews Neuroscience**, v. 11, n. 4, p. 348, 2010.

ética significativa quanto à manipulação do consumidor e à proteção de dados sensíveis.<sup>90</sup>

Os conhecimentos das neurociências ao estudo do comportamento do consumidor e das estratégias de marketing<sup>91</sup> que engloba a Economia Comportamental que “é um campo de estudo relativamente novo, que aplica a compreensão da psicologia às escolhas econômicas”.<sup>92</sup>

O *neuromarketing* utiliza técnicas de imagem cerebral, como a ressonância magnética funcional (fMRI) ou o eletroencefalograma (EEG), para medir as reações neurais e psicofisiológicas dos consumidores diante de estímulos de marketing como anúncios, marcas, produtos ou preços.

O objetivo é entender melhor os processos cognitivos e emocionais que influenciam as decisões de compra e otimizar as campanhas de marketing de acordo com as preferências e motivações dos consumidores, o marketing altera a vontade livre e consciente do consumidor ao se utilizar de técnicas e mecanismos capazes de seduzir, criar sensações e despertar desejos de pertencimento social e consumo.

O consumo motivado por pertencimento e merecimento é uma das formas de utilizar o ambiente digital para promoção massificada de produtos como Rais e Barbosa:

Primeiro porque explora o sentimento íntimo de ser uma “pessoa de qualidade”, de ser diferente da massa: “L’Oréal, porque eu mereço”. E depois porque já não se apega tanto às informações e características literais do produto, mas busca se comunicar com os consumidores pelo humor, pela brincadeira, pelo exagero, visando muito mais vender um estilo de vida associado a marca do que o próprio produto. O desejo de consumo se democratizou ao passo que o mercado também passou a oferecer produtos e serviços baratos que permitissem que uma maior parcela da população tivesse acesso à experiência de compra: medicamentos genéricos, marcas de distribuidores, lojas de maxidesconto, produtos com preços mais baixos da linha, companhias *low cost*.<sup>93</sup>

<sup>90</sup> MORIN, Christophe. Neuromarketing: The new science of consumer behavior. **Society**, v. 48, n. 2, p. 131-135, 2011.

<sup>91</sup> SAJJACHOLAPUNT, Pitch; BALL, Linden J. The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness. **Frontiers in Psychology**, v. 5, article 166, p. 1-11, 3 mar. 2014. Disponível em: <https://www.researchgate.net/publication/260760911>. Acesso em: 27 abr. 2024.

<sup>92</sup> BRIDGER, Darren. **Neuromarketing: como a neurociência aliada ao design pode aumentar o engajamento e a influência sobre os consumidores**. Tradução: Afonso Celso da Cunha Serra. 1.ª ed., 2.ª reimp. São Paulo: Autêntica Business, 2019, p. 180.

<sup>93</sup> MOREIRA, Diogo Rais Rodrigues; BARBOSA, Nathalia Sartarello. O Reflexo da sociedade do hiperconsumo no Instagram e a Responsabilidade Civil dos Influenciadores Digitais. **Revista de Direitos Culturais**, Santo Ângelo, v. 13, n. 30, p. 73-88, mai./ago. 2018. Disponível em: [https://www.researchgate.net/publication/328376203\\_O\\_REFLEXO\\_DA\\_SOCIEDADE\\_DO\\_HIPERCONSUMO\\_NO\\_NO\\_INSTAGRAM\\_E\\_A\\_RESPONSABILIDADE\\_CIVIL\\_DOS\\_INFLUENCIADORES\\_DIGITAIS](https://www.researchgate.net/publication/328376203_O_REFLEXO_DA_SOCIEDADE_DO_HIPERCONSUMO_NO_INSTAGRAM_E_A_RESPONSABILIDADE_CIVIL_DOS_INFLUENCIADORES_DIGITAIS). Acesso em: 29 abr. 2024.

Nesse mesmo sentido temos nos *Nudges* (empurrãozinho comportamental), uma ferramenta à disposição no ambiente digital, e cuja aplicação encontra uma ligação quando os dados pessoais coletados em bases abertas ou por meio de cadastramento, além dos cookies já explicados, tornam a tomada de decisão modulada e inconsciente, uma vez que tomada sob o manto do Sistema 1, descrito por Kahneman linhas atrás.

O uso de técnicas de *neuromarketing* apoiadas por IA precisa ser abordada para proteger os direitos dos usuários, que primeiramente deixam coletar dados pessoais, que são objeto do perfilamento, e depois são inseridos em sistemas de Inteligência Artificial com o escopo de garantir que as decisões automatizadas possam conduzir nossas decisões ou interferir em nossas vidas.

As pessoas costumam tomar decisões boas, mas são frágeis aos impulsos do mercado, como frisa Richard Thaler e Cass Sunstein:

[...] voltaremos nossa atenção para questões relacionadas ao que certas consideram ser o poder mágico dos mercados. Veremos se e quando os livres mercados e a livre concorrência tendem a exacerbar, e não a abrandar, os efeitos da fragilidade humana. O importante a ter em conta aqui é que, apesar de todas as suas qualidades, muitas vezes os mercados incentivam as empresas a entender (e a lucrar com) as nossas fragilidades, em vez de tentar completá-las ou minimizar seus efeitos.<sup>94</sup>

Os autores acima mencionam algumas dessas fragilidades: (i) Distração; (ii) Benefícios agora, custos depois; (iii) Grau de fragilidade; (iv) Frequência; (v) Feedback; (vi) Saber do que você gosta.<sup>95</sup>

Atentamos ao último item “saber do que vocês gosta”, que está vinculado aos dados que você compartilha por meio de coleta de cadastros, alguns deles sem adequação às normas de proteção à privacidade, bem como o consumo digital alimentado pelo comodismo, velocidade e infinidade de ofertas, que aliado aos dados constantes do Big Data, e do compartilhamento entre agentes de mercado, compactuam com esse “saber do que você gosta”.

Decisões de máquina dependem da coleta massiva de dados pessoais, o que nos faz verificar significativas relações em termos de privacidade e a capacidade de mineração de dados pelo Estado e por corporações.

---

<sup>94</sup> THALER, Richard H.; SUNSTEIN, Cass R. **Nugde: Como tomar melhores decisões sobre saúde, dinheiro e felicidade**. Tradução: Ângelo Lessa. 1.ª ed. Rio de Janeiro: Companhia das Letras, 2023, p. 85.

<sup>95</sup> *Ibidem*, p. 93-102.



A intenção com a coleta de dados pessoais no *trade-off* de coleta de dados nos modelos de negócio movidos a dados é descrita por Bagnoli que entende que possuem uma ligação com a preocupação acima citada por Frank Pasquale, a seguir:

A relação entre *Big Data* e privacidade se intensifica, especialmente por dois aspectos: (i) dados pessoais capturados com intensidade cada vez maior e pelas mais variadas formas que o indivíduo muitas vezes sequer tem conhecimento e sequer sabe para que será utilizado; e (ii) a possibilidade de vazamento de dados pessoais capturados, ainda que autorizados pelo indivíduo, durante processos de armazenamento, transmissão e utilização.<sup>96</sup>

A correlação existente está na inserção de dados coletados por negócios “grátis” e que ficam disponíveis no Big Data, e sua utilização para treinamento e decisão dos sistemas de Inteligência Artificial, cujas decisões automatizadas podem conter vieses decorrentes da coleta, tratamento e armazenamento sem os cuidados necessários.

Como efeito da coleta massiva de dados, Barocas e Selbst:

Os vieses humanos são inadvertidamente incorporados aos modelos de machine learning durante o processo de seleção e rotulagem dos dados de treinamento, perpetuando discriminações históricas nas decisões automatizadas.<sup>97</sup>

Rudin discorrendo sobre decisões aponta que: “As decisões humanas na construção de modelos de IA, desde a escolha de variáveis até a definição de métricas de performance, moldam diretamente como estes sistemas irão se comportar em produção”.<sup>98</sup>

O comportamento humano e sua influência sobre os algoritmos é relacionado - e expresso - em informações (pessoais e corporativas como a jornada de consumo) inseridas para treinamento e aprendizado das máquinas, ensinando os algoritmos, e também definindo os limites éticos e morais dentro dos quais as decisões automatizadas proferidas.

Os neurônios espelho, descobertos inicialmente por Giacomo Rizzolatti e sua equipe na década de 1990, representam um marco fundamental na compreensão do comportamento humano e têm implicações significativas para o desenvolvimento da Inteligência Artificial (IA),

---

<sup>96</sup> BAGNOLI, Vicente. **Direito econômico e concorrencial – O Poder Econômico e seus Limites Jurídicos**. 9.<sup>a</sup> ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2022, p. 525.

<sup>97</sup> BAROCAS, Solon; SELBST, Andrew D. Big Data’s Disparate Impact. **California Law Review**, v. 104, n. 671, p. 685, 2016. Available at: <https://ssrn.com/abstract=2477899> or <http://dx.doi.org/10.2139/ssrn.2477899>. Access in: 29 abr. 2024.

<sup>98</sup> RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, v. 1, p. 208, 2019. Disponível em: <https://www.nature.com/articles/s42256-019-0048-x>. Acesso em: 24 abr. 2024.

e são neurônios especializados de um indivíduo para execução de uma ação quanto e quando observa outros realizando a mesma ação.

A economia comportamental, como visto neste capítulo da pesquisa, e como a tomada de decisão humana, a economia comportamental (neuro-economia) estão conectadas a Inteligência Artificial, formam dados que são inseridos e ensinam e influem nas decisões automatizadas, e podem, portanto, fazer parte da Explicação de como um sistema computacional inteligente expressa-se em seus resultados, entendem Quinelato e Faleiros Júnior:

Fala-se na inteligência artificial, mas é fato que o desenvolvimento algoritmo se entrelaça à Economia Comportamental, que, além de prever comportamentos, pode direcioná-los para onde deseja, gerando repercussões jurídicas de grande relevância.<sup>99</sup>

E continuam os autores sobre dados e algoritmos eficazes em mercados com muitos dados, e no mesmo artigo sobre a tomada de decisão:

Os algoritmos “mais eficazes” são aqueles que conseguem coletar mais dados, processá-los em menor tempo e o oferecer as mais rápidas respostas à finalidade operacional para a qual foram desenvolvidos, impondo novos desafios ao direito. É nesse ecossistema que são criados os “mercados ricos em dados” (*data-rich markets*). [...] Portanto, o *homo economicus*, racional e utilitário, passa a ser questionado em sua realidade prática, tendo em vista os vieses que os indivíduos podem incorrer em seus processos decisórios que se coadunam com a chamada *behavioral economics*, pela qual, a partir da limitação de racionalidade do agente, justifica-se a necessidade de regulação das atividades no mercado, o que, no ambiente virtual, dominado pelos algoritmos e pela coleta massiosa de dados, representa grandes riscos.<sup>100</sup>

Fica demonstrado que seres humanos tomam decisões influenciadas pela neuro economia, também chamada de economia comportamental, bem como nossas decisões nas relações com o ambiente digital nas mais diversas relações são transformadas em dados e informações que permitem-nos discutir benefícios, riscos, delegação e/ou compartilhamento da nossa vida e personalidade com máquinas que aprendem e tomam decisão:

Em resumo: i) por questões evolutivas seres humanos possuem emoções que se relacionam intimamente à alterações de estados fisiológicos; ii) quando estas emoções são despertadas, elas participam do processo de tomada de

<sup>99</sup> QUINELATO, Pietra Daneluzzi; FALEIROS JÚNIOR, José Luiz de Moura. Economia Comportamental e Inteligência Artificial na publicidade Veiculada em Mercados Ricos em Dados. In: PINHO, Anna Carolina *et al.* **Discussões sobre Direito na Era Digital**. 1.ª ed. Rio de Janeiro: GZ, 2021, p. 637-639.

<sup>100</sup> *Ibidem*, p. 639 e 643.

decisão do indivíduo; iii) lidando com estas emoções, o sujeito poderá lançar mão de mecanismos para dar unidade e coerência às suas crenças visando sentir-se bem consigo mesmo; e iv) o pertencimento grupal possui um papel importante nas avaliações, pois pode coibir vozes dissonantes por questões reputacionais ou ainda tornar mais extremos os comportamentos dos indivíduos.<sup>101</sup>

A pesquisa demonstra que é necessária a reflexão para conscientização de que as máquinas nos refletem muito em suas decisões automatizadas, os padrões decisórios e comportamentais de um povo, cultura, região etc., e também como podemos explicar as decisões automatizadas quando da personalização e modulação das decisões pessoais reforçadas pelo neuromarketing.

---

<sup>101</sup> KOGA, Bruno Yudi Soares. **Tecnologias persuasivas, decisões e sua liberdade: Como as tecnologias persuasivas afetam suas decisões e tolhem sua liberdade, formando um cabresto digital**. Londrina:Thoth, 2024, p. 120.

## 2 MECANISMOS DE FORMAÇÃO DA DECISÃO ALGORÍTMICA

### 2.1 PERSONALIZAÇÃO E ETICIDADE DA INTELIGÊNCIA ARTIFICIAL

Após entender o que são - e como são - treinados os sistemas de IA, e a influência dos indivíduos e grupos – a pesquisa passa a verificar como decisões pessoais e de mercado dependem de cultura, e de fatores endógenos e exógenos para suas interações digitais de consumo de conteúdo, rede social, mídia, produtos e serviços – esse capítulo procura esclarecer como a personalização, os vieses (nem sempre negativos) e a ética podem contribuir para o Direito à Explicação como uma séria de documentos, metodologias, amparadas por frameworks podem tornar a Inteligência Artificial uma ferramenta transparente, confiável e mais robusta quanto as condições de entendimento de uma decisão automatizada.

Colabora com o acima exposto, Ferrari e Becker, sobre o exercício do Direito à Explicação de decisões automatizadas, em virtude da hipossuficiência do usuário perante os sistemas de IA:

O objetivo dessa faculdade do usuário é dotar de transparência o processo automatizado de escolhas, recomendações. E decisões sobre ele. Com a previsão de um direito à explicação na legislação, o usuário é capaz de entender os critérios utilizados pela máquina. A ideia é garantir ao cidadão a possibilidade de ter um *feedback* científico da atuação do algoritmo. [...] Ademais, outro pilar do direito à explicação consiste na proteção ao hipossuficiente. [...] Em sentido amplo, com ou sem a solicitação do usuário, o direito à explicação exige que lhe seja informado, de forma automática ou personalizada, o motivo pelo qual aquela decisão foi adotada.<sup>102</sup>

A proteção de dados no Brasil, antes mesmo de tratarmos das decisões automatizadas, tem na lei duas especificidades atinentes à coleta dos dados, quais sejam: (i) consentimento e (ii) autodeterminação, que são considerados primordiais para a abordagem ética da IA.

A pesquisa demonstra a relação dos seres humanos, com suas decisões e comportamentos, ou seja, o exercício dos Direitos de Personalidade influencia sobremaneira os dados produzidos que são coletados e inseridos em sistemas de inteligência artificial.

A Lei Geral de Proteção de dados foi “a resposta do poder legislativo brasileiro formatada com o intuito de mitigar a *hiperexposição* dos usuários da internet”<sup>103</sup>, cujo *Hiper* da

<sup>102</sup> FERRARI, Isabela; BECKER, Daniel. Direito à Explicação e Decisões Automatizadas: Reflexões sobre o Princípio do Contraditório. In: NUNES, Dierle; LUCON, Paulo Henrique dos Santos; WOLKART, Erik Navarro. **Inteligência Artificial e Direito Processual: Os Impactos da Virada Tecnológica no Direito Processual**. Salvador: Juspodivm, 2021, p. 290.

<sup>103</sup> BARRETO JÚNIOR, Irineu Francisco; NASCIMENTO, Ariane Azevedo Carvalho do; FULLER, Greice Patrícia. Lei Geral de Proteção de Dados Pessoais: Efetividade Jurídica do Consentimento do Titular para

hiperatividade não é uma categoria imunológica, mas uma representatividade da massificação do positivo<sup>104</sup>, uma vez que dependemos (social e emocionalmente) de conexão, e com isso informações de nossa personalidade e vida que transferimos para o meio ambiente digital, local onde realizamos as mais:

Variadas atividades, tais como compras on-line, acesso às redes sociais, cadastro nos mais diversos sites, dentre outras. Ocorre que muitas dessas informações pessoais fornecidas na internet, por vezes, não são usadas de maneira correta pelas empresas. Isso acontece porque algumas empresas acabam extrapolando o poder que têm sobre o uso dos dados e os utilizando sem o consentimento dos usuários, podendo, em alguns casos, gerar violação aos direitos da intimidade e privacidade.<sup>105</sup>

A informação gerada pela oferta de ambientes, sites, aplicativos, a hiperconectividade e exposição da vida íntima e cotidiana decorre da expansão em escala mundial da internet e do advento das redes sociais fez com que a informação seja o centro do universo da era do dado, sendo seu valor econômico inegável:

Todavia, essa não foi e tampouco será a última das Revoluções, mas a Revolução Tecnológica a princípio, mostra-se como um mundo encantado e perfeito, não demonstrando que seu mau uso, pode gerar danos e inúmeras ameaças. Não valorizando a preciosidade que temos com nossos dados pessoais, pois se busca apenas a exposição do que se tem, ao invés da segurança do que se tem.<sup>106</sup>

Na esfera da privacidade e proteção dos dados pessoais, o consentimento se mostra importante para a coleta dos dados de um indivíduo (que são analisados e perfilados como adiante se demonstrará) – constante do art. 5º, inciso XII da Lei Geral de Proteção de Dados é considerado por alguns doutrinadores (cujo entendimento compartilho em meus estudos) como: “o mais importante requisito do tratamento de dados pessoais amealhados pela internet, ou seja,

---

Tratamento dos Registros. **Revista de Constitucionalização do Direito Brasileiro – RECONTO**, v. 3, n. 2, p. 2, jul./dez. 2020. Disponível em: <https://revistareconto.com.br/index.php/reconto/article/view/44/42>. Acesso em: 24 abr. 2024.

<sup>104</sup> HAN, Byung-Chul. **Sociedade do Cansaço**. Tradução: Ênio Paulo Giachini. 2.ª ed. Petrópolis: Vozes, 2017, p. 21.

<sup>105</sup> BARRETO JÚNIOR, Irineu Francisco; NASCIMENTO, Ariane Azevedo Carvalho do; FULLER, Greice Patrícia. Lei Geral de Proteção de Dados Pessoais: Efetividade Jurídica do Consentimento do Titular para Tratamento dos Registros. **Revista de Constitucionalização do Direito Brasileiro – RECONTO**, v. 3, n. 2, p. 3, jul./dez. 2020. Disponível em: <https://revistareconto.com.br/index.php/reconto/article/view/44/42>. Acesso em: 24 abr. 2024.

<sup>106</sup> BARRETO JÚNIOR, Irineu Francisco; COSTA, Bruna Marangoni Brancalione. Hiperexposição Pessoal nas Redes Sociais e seus Reflexos no Direito. **Revista de Direito, Governança e Novas Tecnologias**, Salvador, v. 4, n. 1, p. 94, jan./jun. 2018. Disponível em: <https://www.indexlaw.org/index.php/revistadgnt/article/view/4299/pdf>. Acesso em: 24 abr. 2024.

o usuário tem o direito de poder analisar como serão utilizados seus dados, com a opção de permitir ou não o tratamento desses dados”.<sup>107</sup>

O hiperconsumo que foi idealizado, ambientado:

Graças às compras *online*, o consumidor se liberta de todos os entraves espaço temporais, pois já não é mais necessário dirigir-se à loja física, como tampouco depende do horário certo para efetuar compras. Pode fazê-lo a qualquer momento e de qualquer lugar que tenha acesso à internet.<sup>108</sup>

A respeito da importância do princípio autodeterminação pode-se afirmar que é um dos mais importantes componentes do direito à privacidade, sob o espectro da Sociedade da Informação<sup>109</sup>, preponderantemente na era do Capitalismo de Plataforma, cujos dados pessoais e de personalidade são utilizados essencialmente em inteligência artificial, sendo de vital importância para o funcionamento desses sistemas computacionais.

A dinâmica econômica mercadológica pautada na exploração e tratamento (inclui a coleta indiscriminada também) no meio eletrônico estabeleceu uma economia da atenção, que obsta o poder de escolha de quais dados podem – ou não – serem coletados, e de que forma, gerando prejuízos ao direito de autodeterminação informativa, à proteção da intimidade, à vida privada e cotidiano dos usuários dos aplicativos, aos sites e às redes sociais.

A sociedade de plataforma é formada pelo tráfego social e econômico, que é cada vez mais capilarizado dentro de um ecossistema global de plataformas online, que por sua vez, é conduzido por algoritmos e alimentado por dados.<sup>110</sup>

Os dados inseridos sem prévia verificação quando a sua coleta, qualificação, finalidade, provocam problemas éticos no tratamento de dados, que são imputados na Inteligência Artificial resultando em impactos sociais e morais que podem ser prejudiciais.

---

<sup>107</sup> BARRETO JÚNIOR, Irineu Francisco; NASCIMENTO, Ariane Azevedo Carvalho do; FULLER, Greice Patrícia. Lei Geral de Proteção de Dados Pessoais: Efetividade Jurídica do Consentimento do Titular para Tratamento dos Registros. **Revista de Constitucionalização do Direito Brasileiro – RECONTO**, v. 3, n. 2, p. 3, jul./dez. 2020. Disponível em: <https://revistareconto.com.br/index.php/reconto/article/view/44/42>. Acesso em: 24 abr. 2024.

<sup>108</sup> MOREIRA, Diogo Rais Rodrigues; BARBOSA, Nathalia Sartarello. O Reflexo da sociedade do hiperconsumo no Instagram e a Responsabilidade Civil dos Influenciadores Digitais. **Revista de Direitos Culturais**, Santo Ângelo, v. 13, n. 30, p. 76, mai./ago. 2018. Disponível em: [https://www.researchgate.net/publication/328376203\\_O\\_REFLEXO\\_DA\\_SOCIEDADE\\_DO\\_HIPERCONSUMO\\_NO\\_INSTAGRAM\\_E\\_A\\_RESPONSABILIDADE\\_CIVIL\\_DOS\\_INFLUENCIADORES\\_DIGITAIS](https://www.researchgate.net/publication/328376203_O_REFLEXO_DA_SOCIEDADE_DO_HIPERCONSUMO_NO_INSTAGRAM_E_A_RESPONSABILIDADE_CIVIL_DOS_INFLUENCIADORES_DIGITAIS). Acesso em: 29 abr. 2024.

<sup>109</sup> BARRETO JÚNIOR, Irineu Francisco; MOLINA, Fernanda Zampieri. Capitalismo de plataforma: a ameaça ao direito à autodeterminação informativa na Sociedade da Informação. **Revista Brasileira de Estudos Políticos**, Belo Horizonte, n. 125, p. 245, jul./dez. 2022. Disponível em: <https://pos.direito.ufmg.br/rbep/index.php/rbep/article/view/852/657>. Acesso em: 24 abr. 2024.

<sup>110</sup> *Ibidem*, p. 247.

Tateoki discorre sobre o uso dos dados e o perfilamento para induzir usuários-consumidores a consumir:

Em um mundo onde a enxurrada de informações, a grande competitividade e variedade do mercado, para se destacar, se diferenciar ou ter atenção dos consumidores, é necessário compreender seus hábitos para oferecer produtos ou serviços mais personalizados. Desta forma, há a necessidade da coleta de dados para compreender os consumidores e prever o que eles querem. Assim, surge a prática do profiling, que é construir perfis das pessoas e enquadrá-las em grupos e padrões, que muitas vezes acabam estigmatizando as pessoas. [...] Tal visão, parece tentador, mas a longo prazo pode ocasionar inúmeros problemas. Nós humanos não somos racionais ao ponto de perceber que todas essas maravilhas podem estar afetando os nossos modos de vida e como pensamos. Não somos seres vivos somente racionais, como visto anteriormente, podemos ser influenciados e manipulados. Dessa forma, as plataformas digitais usem *nudges* para manter mais ainda atenção do usuário filtrando a informação. Cada vez mais nossa atenção está sendo redirecionada para as plataformas digitais, nosso Sistema Reflexivo está perdendo a capacidade de se concentrar, podendo, como visto anteriormente, acabar em cegueira e não perceber o que está acontecendo ao redor. Nosso Sistema Automático acaba se tornando o principal ao utilizar dessas plataformas. Acontece que ele possui heurísticas e vieses.<sup>111</sup>

No campo ético, é preciso inicialmente entender a distinção de ética filosófica e a ética prática, para sabermos como a Inteligência Artificial e seus resultados (*outputs*), e Santaella, afirma:

Quando se em ética, de saída, é preciso demarcar o campo. Uma vez que se trata de uma questão crucial que afeta nosso estar no mundo, nossas relações com o outro e nossos modos de agir, existe uma ética filosófica, reflexiva e uma ética prática. A primeira existe para iluminar a prática, quer dizer, fornecer princípios que a prática deveria ter como alvo colocar em ação. É providencial aqui a palavra “ação” pois a ética lida com o agir humano. Tanto é verdade que a ética depende de um campo de reflexão filosófica que o tema frequentou e até hoje continua frequentando o pensamento de muitos filósofos através dos séculos. Já em Aristóteles encontra-se uma filosofia da ética perfeitamente integrada ao todo do seu pensamento. Para ele, os seres, as coisas e os humanos especialmente tendem para o bem, uma tendência que conduz a felicidade. [...] A ação para esse fim é individual, mas ela ajuda a compor o *homo politicus*, habitante da polis, responsável pelo bem coletivo. Ao fim e ao cabo, com o perdão da simplificação, para Aristóteles, a ética é o caminho para o bem e este é dependente da ação humana, individual e coletiva.<sup>112</sup>

---

<sup>111</sup> TATEOKI, Victor Augusto. **O Uso dos Dados Pessoais como Mecanismo de Persuasão no Processo de Tomada de Decisão dos Usuários de Internet**. Rio de Janeiro: Lumen Iuris, 2021, p. 176-178.

<sup>112</sup> SANTAELLA, Lúcia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis Santiago Guerra Filho *et al.* **Direito e Inteligência Artificial: Fundamentos – Volume 1 - Inteligência Artificial, Ética e Direito**. Rio de Janeiro: Lumen Iuris, 2021, p. 112-113.

Dessa maneira, é possível concluir que para termos uma Inteligência Artificial dita “do bem” (ou “*for good*”) precisamos investir também no desenvolvimento desse citado “*homo policitus*” pela autora retromencionada, vez que o aprendizado que a máquina efetua, é proporcionalmente influenciado pelos comportamentos, culturas e ações que os seres humanos transferem em forma de dados e informações.

Nesse sentido, complementa e continua Santaella, ao pensar e escrever sobre princípios, diretrizes e categorias que podem auxiliar a IA a produzir resultados e efeitos mais voltados para o bem e a ética em todo o mundo, o que é um grande desafio para a IA, veja:

Da análise de documentos relativos às iniciativas tomadas para garantia de que o mal, a tão conhecida banalidade do mal, não se instale e encontre morada também na IA, foram extraídas aquilo que os autores chamam de dimensões e que podem também ser conhecidas como categorias, no sentido em que abrangem, em uma mesma classe, princípios comuns. As dimensões ou categorias que comparecem nos documentos, com mais ou menos ênfase são os seguintes: 1) Equidade (*Fairness*), 2) Confiabilidade e Segurança (*Reliability and Safety*), 3) Impacto Social (*Social Impact*), 4) Responsabilidade (*Accountability*), 5) Privacidade e Segurança (*Privacy and Security*) e, por fim, 6) Transparência (*Transparency*). Os documentos estão recheados de boas palavras e intenções as quais levantarei a seguir em uma enumeração relativamente caótica, mas não obstante, representativa. “Distribuição equitativa e justa de benefícios e custos”, isenção quanto a “preconceitos injustos, discriminação e estigmatização”, “evitar preconceitos não intencionais”, “respeito ao Estado de Direito”, os direitos humanos, os valores democráticos e a diversidade”, “promover o desenvolvimento sustentável da natureza e da sociedade”, “exercer níveis adequados de julgamento e permanecer responsáveis pelo desenvolvimento, implementação e uso da IA”, “manter a confiança dos usuários nos sistemas de IA”, mesmo diante daquilo que vem sendo chamado de algoritmos de Caixa Preta, e por aí vai. [...] É óbvio que princípios de proteção devem ser estabelecidos, não há dúvida quanto à implementação de exigentes padrões para a triagem dos dados para desenvolvimento de algoritmos, a dúvida que surge é até que ponto esses princípios podem ser efetivamente incrementados em uma realidade impregnada de contradições e paradoxos. Não são poucos os que colocam o dedo na ferida ao evidenciar que implantar os princípios. E guias para uma IA ética implica em uma luta incessante.<sup>113</sup>

Ética não só na coleta e inserção dos dados em um sistema inteligente artificialmente fundamental para mitigação coletiva de danos, mas também em observar seus resultados, uma vez que “A ética da inteligência artificial deve considerar não apenas a eficácia tecnológica, mas também os impactos sociais e morais”.<sup>114</sup>

<sup>113</sup> SANTAELLA, Lúcia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis Santiago Guerra Filho *et al.* **Direito e Inteligência Artificial: Fundamentos – Volume 1 - Inteligência Artificial, Ética e Direito**. Rio de Janeiro: Lumen Juris, 2021, p. 118-120.

<sup>114</sup> FLORIDI, Luciano. **Ethics of Artificial Intelligence**. Oxford: Oxford University Press, 2020, p. 45.



Desta forma, ética é um tema de extrema importância quando se trata de Inteligência Artificial, sendo importante:

Determinar o impacto ético potencial e real de um algoritmo é difícil por muitas razões. Identificar a influência da subjetividade humana no design e configuração de algoritmos frequentemente requer a investigação de processos de desenvolvimento de longo prazo e multiusuário. Mesmo com recursos suficientes, problemas e valores subjacentes muitas vezes não serão aparentes até que um caso de uso problemático surja. Os algoritmos de aprendizado, frequentemente citados como o ‘futuro’ dos algoritmos e análises (Tutt, 2016), introduzem incerteza sobre como e por que as decisões são tomadas devido à sua capacidade de ajustar parâmetros operacionais e regras de tomada de decisão ‘no mundo real’ (Burrell, 2016). Determinar se uma decisão problemática em particular é apenas um ‘bug’ pontual ou evidência de uma falha ou viés sistêmico pode ser impossível (ou pelo menos altamente difícil) com algoritmos de aprendizado mal interpretáveis e imprevisíveis. Tais desafios tendem a crescer, à medida que os algoritmos aumentam em complexidade e interagem com as saídas uns dos outros para tomar decisões (Tutt, 2016). A lacuna resultante entre o design e a operação dos algoritmos e nossa compreensão de suas implicações éticas pode ter consequências severas que afetam indivíduos, grupos e segmentos inteiros de uma sociedade.<sup>115</sup>

O princípio da autodeterminação e sua relação com a inteligência artificial trazidos a pesquisa reflete sobre como este princípio importante e fundamental – como um desdobramento da privacidade e do Direito de Personalidade, e sua importância está diretamente relacionada com os *inputs* nos sistemas de IA.

Com efeito, é imprescindível reforçar o pensamento de Pasquale, que destaca que no setor de saúde, setor cujas:

As decisões automatizadas podem ter consequências críticas, tanto positivas quanto negativas. A implementação de IA na medicina precisa de uma

---

<sup>115</sup> Texto Original: “Determining the potential and actual ethical impact of an algorithm is difficult for many reasons. Identifying the influence of human subjectivity in algorithm design and configuration often requires investigation of long-term, multi-user development processes. Even with sufficient resources, problems and underlying values will often not be apparent until a problematic use case arises. Learning algorithms, often quoted as the ‘future’ of algorithms and analytics (Tutt, 2016), introduce uncertainty over how and why decisions are made due to their capacity to tweak operational parameters and decision-making rules ‘in the wild’ (Burrell, 2016). Determining whether a particular problematic decision is merely a one-off ‘bug’ or evidence of a systemic failure or bias may be impossible (or at least highly difficult) with poorly interpretable and predictable learning algorithms. Such challenges are set to grow, as algorithms increase in complexity and interact with each other’s outputs to take decisions (Tutt, 2016). The resulting gap between the design and operation of algorithms and our understanding of their ethical implications can have severe consequences affecting individuals, groups and whole segments of a society.” MITTELSTADT, Brent *et al.* The ethics of algorithms: Mapping the debate. **Big Data & Society**, jul./dec. 2016. Available at: [https://www.researchgate.net/publication/309322060\\_The\\_Ethics\\_of\\_Algorithms\\_Mapping\\_the\\_Debate](https://www.researchgate.net/publication/309322060_The_Ethics_of_Algorithms_Mapping_the_Debate). Access in: 29 abr. 2024.

estrutura ética robusta para evitar que falhas algorítmicas prejudiquem a qualidade do cuidado ao paciente.<sup>116</sup>

A autodeterminação define-se como sendo:

A capacidade do indivíduo de autodeterminar seus dados pessoais seria parcela fundamental do seu direito de desenvolver sua personalidade com liberdade, condição mínima para o livre desenvolvimento da personalidade e a autonomia pessoal.<sup>117</sup>

O princípio vem disposto no art. 2º, inciso II da Lei Geral de Proteção de Dados do Brasil, claramente como um princípio basilar da privacidade, proteção e do desenvolvimento da personalidade do indivíduo tutelado pela legislação, condição essencial – e existencial – da natureza humana, como demonstram Chiarello e Oliveira:

Isso significa dizer que toda pessoa humana tem direitos da personalidade que devem ser zelados e respeitados. Dentro dessa categoria, existe o direito à imagem, que corresponde ao direito que toda pessoa tem sobre sua forma física com todos os componentes distintos que a individualizam (como boca, nariz, olhos e entre outras características físicas).<sup>118</sup>

Como acima exposto, personalidade é tanto a dinâmica emocional, como a dinâmica física do indivíduo, que ao tornarem-se dados digitais são fontes de aprendizado e resultados de decisões de máquinas que aprendem e decidem.

Personalidade, de acordo com o art. 11 do Código Civil, é um direito irrenunciável e intransmissível demandando, portanto, que o titular dele disponha por meio de consentimento para terceiros, devidamente informados e sob seu controle, ou seja, submetido à autodeterminação desse mesmo indivíduo.

O livre desenvolvimento da personalidade contém duas dimensões: a proteção da personalidade e a garantia da liberdade geral de ação.<sup>119</sup>

A identidade pessoal tem relação direta com a privacidade, como assevera Martins:

---

<sup>116</sup> PASQUALE, Frank. Algorithms, healthcare, and ethics: Understanding the implications of machine decision-making. **Health Policy Journal**, v. 15, n. 3, p. 87-104, 2019.

<sup>117</sup> *Ibidem*, p. 87-104.

<sup>118</sup> PINTO, Felipe Chiarello de Souza; OLIVEIRA, Gabriela Franklin de. Não acredite em tudo que vê: *deepfake pornography* e responsabilidade civil no ordenamento jurídico brasileiro. **Revista Direito & Política**, v. 18, n. 2, p. 437, 2023. Disponível em: <https://periodicos.univali.br/index.php/rdp/article/view/19869>. Acesso em: 24 abr. 2024.

<sup>119</sup> MARTINS, Pedro Bastos Lobo. **Profiling na Lei Geral de Proteção de Dados: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica**. 1.ª ed. Indaiatuba: Foco, 2022, p. 5. (Coleção Persona – coordenador: Brunello Souza Stancioli).

A construção da identidade pessoal envolve um constante processo de seleção e interpretação de informações pessoais, ensejando uma disputa entre diferentes narrativas das quais emerge a identidade. Por esse motivo, há uma evidente tensão entre a capacidade de controle dessas informações pessoais e o acesso a essas informações por terceiros.<sup>120</sup>

Ainda na mesma obra, o autor chama *Profiling*, o perfilamento de dados pessoais para sistemas computacionais e, portanto, Inteligência Artificial, como: “[...] uma atividade de tratamento automatizado específica, qual seja, *profiling*.”<sup>121</sup>

E sugere algumas categorias e completá-las: “por algumas premissas e características do *profiling*: (i) categorização; (ii) formação de perfis não distribuídos; (iii) modulação; (iv) possibilidade de cruzamento de dados e novas formas de discriminação, e finalmente (v) persistência dos dados.”<sup>122</sup>

A personalização, ou *Profiling*, na economia digital, possui e utiliza:

O número elevado de dados (volume) e sua variedade são as pedras fundamentais da análise de dados, pois é através desta quantidade e diversidade que se mostra possível estabelecer conexões (muitas delas não intuitivas) entre os dados obtidos, permitindo que sejam reveladas informações antes desconhecidas pelos processadores. Necessário pontuar que a qualidade dos dados será fator determinante na análise, pois as correlações extraídas do conjunto de dados terão a mesma natureza dos dados inseridos no sistema (*garbage in; garbage out; bias; in, bias out*). Como visto, dados imprecisos, desatualizados, inverídicos ou mesmo enviesados sofrerão análises marcadas pelos mesmos defeitos. Ou seja, não se deve supor que a simples quantidade de dados resolverá o problema de encontrar quais são os drivers necessários para realizar a perfilização. [...].<sup>123</sup>

Quinelato expõe que:

É a partir de dados tratados que empresas como Google, Amazon, Apple e Facebook (ou, recentemente, Meta) constroem seus bancos de dados, utilizando-os para aprimorar seus produtos e serviços, entender o perfil de seu público consumidor, vender espaços publicitários para, por exemplo, permitir que terceiros apresentem seus anúncios de uma forma direcionada, de acordo com as necessidades ou preferências já demonstradas por determinado indivíduo. Assim, tais dados coletados podem ser explorados como insumos para produzir e personalizar bens e serviços, auxiliando na criação de conhecimento e na automatização da tomada de decisões. [...] Portanto, a tecnologia possibilitou que, em troca do compartilhamento de dados, insumos

<sup>120</sup> MARTINS, Pedro Bastos Lobo. **Profiling na Lei Geral de Proteção de Dados: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica**. 1.<sup>a</sup> ed. Indaiatuba: Foco, 2022, p. 21. (Coleção Persona – coordenador: Brunello Souza Stancioli).

<sup>121</sup> *Ibidem*, p. 79.

<sup>122</sup> *Ibidem*, p. 80.

<sup>123</sup> KOGA, Bruno Yudi Soares. **Precificação Personalizada**. 1.<sup>a</sup> ed. São Paulo: Almedina, 2021, p. 93.

para empresas, fosse concedido acesso a algumas ferramentas que facilitam afazeres diários (como a busca de determinada informação no buscador Google ou o pedido de almoço por aplicativo) e permitem a interação entre indivíduos, empresas, governos e outras organizações (como as redes sociais e o aplicativo Whatsapp).<sup>124</sup>

E a autora discorre sobre a ligação de comportamento, tomada de decisões e decisões automatizadas, assim descreve:

A internet está intrinsecamente. Associada às atividades rotineiras da maioria dos indivíduos. Como consequência, uma maior quantidade de dados pessoais pode ser coletada e processada, auxiliando na personalização de produtos e serviços, como a publicidade direcionada comportamental e precificação personalizada. [...] Indivíduos não fazem escolhas racionais a todo momento. Essas decisões podem ter sido influenciadas por diversos fatores, principalmente em cenários de incerteza, sendo objeto de estudo da Economia Comportamental.<sup>125</sup>

A conexão é parte dessa pesquisa com o objetivo de esclarecer questões relacionadas com a Explicação proferida por instituições privadas e públicas quando provocadas à explicar as decisões automatizadas de seus sistemas de Inteligência Artificial, bem como para a própria Explicabilidade de Inteligência Artificial, cujos algoritmos e dados utilizados tem relação direta em decorrência da forma como o algoritmo aprende com os dados, seus padrões de *inputs* e pesos, que são determinantes como visto acima para o proferimento das decisões automatizadas.

O perfilamento, como categorização, poderá ocasionar discriminações.

Conforme aponta Adilson Moreira:

A palavra discriminação possui uma pluralidade de significados, embora tenha adquirido um sentido bem específico no mundo atual. Ela designa, por um lado, a ação de classificar objetos a partir de um determinado critério. Essa acepção genérica passou a segundo plano por causa da preponderância de sua dimensão moral e jurídica nos dias atuais. Hoje, o termo discriminar tem conotações claramente negativas, pois sugere que alguém foi tratado de forma arbitrária. Os dois sentidos dessa palavra estão presentes no vocabulário jurídico. Sabemos que instituições estatais classificam indivíduos a partir de uma série de critérios necessários para o alcance de algum interesse público. O vocábulo discriminar significa aqui categorizar pessoas ou situações a partir de uma característica para atribuir a elas algum tipo de consequência. Contudo, a palavra discriminação tem também outro significado no mundo jurídico: ela indica que uma pessoa impõe à outra um tratamento arbitrário a

---

<sup>124</sup> QUINELATO, Pietra Daneluzzi. **Preços Personalizados à luz da Lei Geral de Proteção de Dados**. 1.<sup>a</sup> ed. São Paulo: Foco, 2022, p. 4-5.

<sup>125</sup> *Ibidem*, p. 62-63.

partir de um julgamento moral negativo, o que pode contribuir para que a segunda esteja em uma situação de desvantagem.<sup>126</sup>

No capítulo anterior tratou-se sobre o “saber do que você gosta”, que reside na categorização e classificação dos indivíduos (n-gramas classificatórios para Processamento e Linguagem Natural).<sup>127</sup>

As máquinas têm sistemas treinados por dados que produzimos segundo a segundo, e a Lei Geral de Proteção de Dados prevê como marco regulatório que os titulares têm uma série de tutelas normativas enunciadas, mas:

Pelo grande estímulo e dependência das tecnologias, bem como rapidez evolutiva da IA e consequentemente de suas tecnologias, vê-se que não será tarefa fácil nem estatística a aplicação desta lei pelos seus desdobramentos. O uso de dados é uma realidade mundial, onde é notório que atualmente as maiores sociedades empresárias que comandam a economia mundial estão voltadas à coleta e tratamento de dados, onde seu impacto é bastante expressivo em todo o globo, resultado do processo de globalização, ou seja, interdependência das nações. [...] É perceptível que com a LGPD todos os setores sociais estarão sujeitos a severas mudanças de adequação às novas diretrizes de proteção de dados pessoais estabelecidas na lei. De igual forma, este necessário ajustamento alcança frontalmente as empresas e agências que trabalham e dependem da publicidade para convencer o consumidor, além de definir as estratégias de vigilância, modulação de comportamento e coleta e utilização de dados para dar complemento ao processo de perfilamento.<sup>128</sup>

Para entender os vieses da cognição humana e como se formam (como explicado anteriormente sobre a inserção de dados nos sistemas computacionais inteligentes) os vieses das decisões da Inteligência Artificial:

A atuação humana permite que erros e falhas ocorram na inteligência artificial, podendo culminar na inserção de vieses ediscriminações humanas e sociais, dentro de uma realidade computacional que não possui juízos de valor capazes de barrar decisões discriminatórias. [...] Algoritmos enviesados implicam a chamada discriminação algorítmica, que pode ensejar a tomada de decisões públicas e privadas que afetam direitos fundamentais de pessoas após tomadas de decisões, públicas e privadas, que se basearam nas respostas probabilísticas de uma inteligência artificial não isenta.<sup>129</sup>

<sup>126</sup> MOREIRA, Adilson José. **Tratado de Direito Antidiscriminatório**. 1.ª ed. São Paulo: Editora Contracorrente, 2020, p. 312.

<sup>127</sup> THALER, Richard H.; SUNSTEIN, Cass R. **Nudge: Como tomar melhores decisões sobre saúde, dinheiro e felicidade**. Tradução: Ângelo Lessa. 1.ª ed. Rio de Janeiro: Companhia das Letras, 2023, p. 282.

<sup>128</sup> COUTINHO, Diego Araújo. **O consumidor e modulação algorítmica de comportamento: a influência da Inteligência Artificial por meio de algoritmos no poder decisório do consumidor**. São Paulo: Dialética, 2021, p. 148.

<sup>129</sup> MARQUES, Elaine Cristina; SILVA, Victor Hugo Cunha. Capítulo 1 - Vieses Algorítmicos nas Aplicações de Inteligência Artificial: uma análise dos projetos de lei brasileiros. *In*: VIGLIAR, José Marcelo Menezes. **Inteligência artificial: aspectos jurídicos**. São Paulo: Almedina, 2023, p. 23.

Permite-se aqui dois apontamentos: o primeiro é sobre o uso dos termos “algoritmos enviesados” e “discriminação algorítmica”, posto que conforme foi anteriormente explicado, dificilmente o enviesamento e/ou discriminação é do algoritmo (séria de códigos que pretendem dar um resultado, como uma receita de bolo – que é um algoritmo), mas sim do Banco de Dados que, ao ser coletado e inserido em sistemas de Inteligência Artificial, provocam o enviesamento do aprendizado e das decisões/resultados da saída (*outputs*).

O algoritmo e seu aprendizado para posterior decisão baseia-se no cálculo computacional, nas probabilidades, nas decisões “ótimas” a partir do *input* de dados, e não simplesmente nos cálculos matemáticos (que não possuem padrões por si mesmos), conforme abaixo:

Algoritmos tendem a ser vulneráveis a características de seus dados de treinamento. Esta é um recurso destes algoritmos: a habilidade de se adaptar face a *inputs* cambiantes. Mas a adaptação algorítmica em resposta aos dados fornecidos também apresenta um vetor de ataque por usuários mal-intencionados. Esta vulnerabilidade da dieta de dados em algoritmos de aprendizado é um tema recorrente.<sup>130</sup>

Para podermos melhorar o *inputs* de dados pessoais precisamos observar diversos aspectos, ou seja, necessitamos de mais dados para transformar nossos conhecimentos em decisões melhores, pois nossas identidades comportamentais têm informado padrões de dados que são inseridos em Inteligência Artificial, resultando em um espelhamento da nossa cultura e sociedade.

O “Big Data público já é uma realidade, o controle de dados exercido pelo Poder Público passa a ostentar nova dimensão com a possibilidade de compartilhamento interorgânico”<sup>131</sup>, ou seja, a interoperabilidade de dados entre os cadastros é evidente e constituem uma base informacional relevante, sendo necessário entender também como a variável do algoritmo pode apresentar um viés.

Para tanto, Sicsu, Samartini e Barth expõem o dilema viés por variabilidade (*Bias – Variance Trade off*):

---

<sup>130</sup> OSOBA, Osonde; WELSER IV, William. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. **Rand Corporation**, 2017, p. 7. Disponível em: [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR1700/RR1744/RAND\\_RR1744.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf). Acesso em: 13 jun. 2024.

<sup>131</sup> FALEIROS JÚNIOR, José Luiz de Moura. **Administração Pública Digital: proposições para o aperfeiçoamento do Regime Jurídico Administrativo na sociedade da informação**. 2.<sup>a</sup> ed. Indaiatuba: Foco, 2024, p. 176/177.

Ao rodar um algoritmo para obter a relação entre variável alvo Y e as variáveis previsoras (x1, x2, ...) os valores ajustados de Y, em geral, não são iguais ou próximos dos valores observados. As diferenças entre esses valores são denominadas erros. O objetivo do analista é obter um modelo com erros pequenos, que possa ser generalizado, isto é, aplicado a outras amostras da população com bons resultados. Quando tentarmos representar uma relação complexa entre a variável alvo e as variáveis previsoras por meio de um modelo simples, este provavelmente não refletirá corretamente essa relação. Essa situação é conhecida como *underfitting* (subajuste). A diferença entre valores ajustados e os valores observados é grande. Diremos que o modelo apresenta um viés alto.<sup>132</sup>

Os dados que disponibilizamos para entradas (*Inputs*) em Inteligência Artificial refletem como recebemos as comunicações do ambiente (como Sistema) em que vivemos, como pensado por Niklas Luhmann, reconhecido por sua teoria dos sistemas sociais da Alemanha.

Cujo conceito de *sistema*:

[...] designa agora não mais um objeto, mas um conceito teórico diferencial, ou seja, *um sistema é a forma de uma diferença*, possuindo dois lados: o sistema (como interno da forma) e o ambiente (como o lado externo da forma). Somente ambos os lados constituem a diferenciação, a forma, o conceito. Como destaca Luhmann, o ambiente é tão importante e indispensável para esta forma, quanto o próprio sistema. Como diferenciação a forma é fechada, o que significa que tudo o que se pode observar e descrever com esta diferenciação pertence ou ao sistema ou ao ambiente. O limite entre sistema e ambiente marca a unidade da forma e por isso não deve ser concebido nem de um lado nem de outro. Mas o sistema auto-referencial ganha uma precisão inovadora com o conceito de *autopoiesis* que Luhmann transporta da designação dada pelo neurobiólogo chileno Humberto Maturana para os sistemas vivos. Com efeito, Maturana utiliza o termo grego “*poiésis*” (não “*68omple*”), para designar “produção” do próprio sistema e não de um resultado externo a ele (no caso da neurobiologia, a célula produz seus próprios elementos através da rede de seus próprios elementos). O conceito de *autopoiesis* veio a significar um avanço em relação ao conceito de “auto-organização” desenvolvido desde 1960.<sup>133</sup>

As decisões pessoais dos indivíduos, uma vez transmitidas aos sistemas computacionais com quem têm relações sociais, fazem a máquina aprender e decidir. A máquina não possui psique, crenças e outras propriedades do mundo do psicológico comportamental, mas como informam Russel e Norvig:

<sup>132</sup> SICSU, Abraham Laredo; SAMARTINI, André; BARTH, Nelson Lerner. **Técnicas de Machine Learning**. São Paulo: Blucher, 2023, p. 21.

<sup>133</sup> NEVES, Clarissa Eckert Baeta; SAMIOS, Eva Machado Barbosa. **Niklas Luhmann: a Nova Teoria dos Sistemas**. 1.ª ed. Porto Alegre: Editora da Universidade Federal do Rio Grande do Sul, Goethe-Institut/ICBBA, 1997. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/242209/000152200.pdf?sequence=1&isAllowed=y>. Acesso em: 24 abr. 2024.

Um agente aprenderá a melhorar seu desempenho após fazer observações sobre o mundo. A aprendizagem pode variar do corriqueiro, como anotar uma lista de compras, até o mais profundo, como mostrado por Albert Einstein, que deduziu uma nova teoria para o universo. Quando o agente é um computador, nós o chamamos aprendizado de máquina: um computador observa alguns dados, monta um modelo baseado nos dados e usa o modelo como uma hipótese sobre o mundo e um *software* que pode resolver problemas. [...] A tecnologia de aprendizado de máquina se tornou uma parte padrão da engenharia de *software*. Sempre que você estiver criando um sistema de *software*, mesmo que não pense nele como um agente de IA, os componentes do sistema podem ser potencialmente melhorados com o aprendizado de máquina.<sup>134</sup>

O usuário-indivíduo age perante a comunicação que recebe do Sistema Social, ele faz uso de seu conhecimento e age de forma livre sobre como se relacionará com a sociedade e o mercado, pois:

A liberdade de o homem escolher e agir sofre restrições de três tipos. Em primeiro lugar, estão as leis físicas a cujas inexoráveis determinações o homem tem que se submeter se quiser permanecer vivo. Em segundo lugar, estão as características e aptidões congênitas de cada indivíduo e sua inter-relação com o meio ambiente; tais circunstâncias, indubitavelmente, influenciam tanto a escolha dos fins e a dos meios, embora nosso conhecimento de como isso se processa seja bastante impreciso. Finalmente, existe a regularidade das relações de causa e efeito entre os meios utilizados e os fins alcançados; ou seja. As leis praxeológicas, que são distintas das leis físicas e fisiológicas.<sup>135</sup>

Os marcadores somáticos representam um conceito fundamental na neurociência cognitiva, introduzido inicialmente por António Damásio, que descreve como estados emocionais influenciam processos decisórios e comportamentais. Estes marcadores funcionam como sinais corporais que auxiliam na tomada de decisão, conectando experiências passadas a respostas emocionais e fisiológicas.<sup>136</sup>

Os marcadores somáticos atuam como mecanismos neurais que integram informações emocionais e cognitivas, permitindo decisões mais rápidas e adaptativas. Pesquisas utilizando neuroimagem funcional demonstram ativação significativa da amígdala e do córtex pré-frontal ventromedial durante processos decisórios influenciados por marcadores somáticos.<sup>137</sup>

<sup>134</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022, p. 590-591.

<sup>135</sup> MISES, Ludwig Von. **Ação Humana**. São Paulo: Instituto Ludwig Von Mises Brasil, 2023, p. 999.

<sup>136</sup> DAMÁSIO, Antonio. **A estranha ordem das coisas - As origens biológicas dos sentimentos e da cultura**. Tradução: Laura Teixeira Motta. São Paulo: Companhia das Letras, 2018, p. 20.

<sup>137</sup> BECHARA, Antoine; DAMASIO, Antonio R. The somatic marker hypothesis: Twenty years later. **Nature Reviews Neuroscience**, v. 22, n. 4, p. 273-289, 2021.



A modulação comportamental mediada por marcadores somáticos também se mostra relevante em contextos clínicos. Estudos recentes em pacientes com transtornos de ansiedade revelaram alterações na sinalização dos marcadores somáticos, sugerindo seu papel em quadros psicopatológicos.<sup>138</sup>

É preciso racionalidade para tomar decisão, e temos em Pinker a ideia de que:

Muitos racionalistas acreditam que a regra de Bayes está entre os modelos normativos que são desenhados com maior frequência no raciocínio humano cotidiano; e que, se fosse mais valorizada, poderia dar maior impulso à racionalidade pública. Em décadas recentes, o pensamento bayesiano apresentou uma ascensão vertiginosa, ganhando proeminência em todos os campos científicos. Embra poucos leigos saibam. Identificá-lo ou. Explica-lo, eles sentem sua influência no termo da mora “prior”, que se refere a uma das variáveis do teorema. [...] O grande *insight* do do reverendo Thomas Bayes (1701-1761) foi o de que o grau de crença numa hipótese pode ser quantificado como uma probabilidade.<sup>139</sup>

É incrível a convergência entre pensamento humano e pensamento da máquina, que se utiliza da matemática de padrão de Bayes como um de seus mecanismos de processamento dos dados e seus padrões, para tomar suas decisões.

Continua Pinker descrevendo que:

A racionalidade requer que distingamos o que é verdadeiro do que queremos que seja verdadeiro: o que não enterremos a cabeça na areia, construamos castelos no ar, nem decidamos que as uvas que estão fora do nosso alcance estejam verdes. As tentações do pensamento mágico e desejado estão sempre conosco porque nossa sorte depende do estado do mundo, que nunca podemos conhecer com certeza.<sup>140</sup>

A seleção e o tratamento dos *inputs* (inserções de dados) é fator importante para evitar problemas éticos e vieses negativos é selecionar e desenvolver o maior número de variáveis—no desenvolvimento e utilização da Inteligência Artificial, a seleção de variáveis, Sicsu, Samartini e Barth a seguinte exposição:

A seleção de variáveis predictoras (*inputs*) para treinar uma rede neural é, como em todos os tipos de modelagem preditiva, uma das etapas mais críticas do processo. A qualidade de uma previsão depende fundamentalmente das variáveis utilizadas para completá-la. A tendência usual é escolher um número

<sup>138</sup> OLIVEIRA, M. T.; SILVA, R. A. Marcadores somáticos em transtornos de ansiedade: uma revisão sistemática. *Revista Brasileira de Psiquiatria*, v. 46, n. 1, p. 45-58, 2024.

<sup>139</sup> PINKER, Steven. **Racionalidade: o que é, porque parece estar em falta, por que é importante**. Tradução: Waldéa Barcellos. 1.ª ed. Rio de Janeiro: Intrínseca, 2022, p. 168-169.

<sup>140</sup> *Ibidem*, p. 169.

muito grande de possíveis variáveis previsoras e depois verificar quais delas realmente terão impacto na acurácia da previsão. [...] Um ponto importante na consideração das variáveis quantitativas é que seus extremos englobem todos os casos possíveis na população onde o algoritmo será aplicado. A utilização de redes neurais em situações em que as variáveis podem assumir valores fora da faixa utilizada no treinamento (extrapolação) compromete a generalização dos resultados, ou seja, pode conduzir a previsões erradas.<sup>141</sup>

Desse modo, da leitura acima podemos entender que inputs sem seleção de tratamento correto de uma população favorecem (desde seu desenvolvimento até sua utilização institucional) eventuais ocorrências de vieses ou problemas éticos em suas decisões, uma vez que as diferenças culturais, linguísticas, comportamentais, geológicas e de personalidade de uma população são diferentes, ocasionando decisões automatizadas que podem ser alteradas ou prejudiciais, e também ocasionar problemas relacionados com a Explicação para um determinado fim solicitado.

Questão ética apontada por Bucci é a de que:

Em suma, não há nada de inaceitável no fato de expressivas parcelas da sociedade não entenderem nada de tecnologia, mas há algo de terrivelmente errado numa tecnologia que, sem entender nada de sociedade, age para formatar e programar a sociedade.<sup>142</sup>

A relação entre neurociências econômicas, comportamento e decisões automatizadas é um campo fascinante e em rápida expansão. As neurociências econômicas buscam entender como o cérebro humano toma decisões em contextos econômicos, utilizando técnicas de neuroimagem e psicologia cognitiva. Essa compreensão do comportamento humano, em particular dos vieses cognitivos e emocionais, é crucial para o desenvolvimento de sistemas de decisão automatizada mais eficazes, éticos e justos.

E para essa pesquisa, trago novamente Bucci discorrendo sobre os efeitos éticos das máquinas, a ideia ética relacionada as decisões das máquinas, uma vez que:

Estamos diante de um cenário em que as máquinas tomam decisões e fazem escolhas, sim. O direcionamento do tráfego urbano contém um encadeamento de decisões éticas. O monitoramento dos preços das ações nos mercados de capitais, igualmente. O controle do alastramento das pandemias, as sugestões de políticas públicas que daí decorrem quase automaticamente, também. Mesmo que se argumente que as decisões feitas por máquinas não passam por

<sup>141</sup> SICSU, Abraham Laredo; SAMARTINI, André; BARTH, Nelson Lerner. **Técnicas de Machine Learning**. São Paulo: Blucher, 2023, p. 21.

<sup>142</sup> BUCCI, Eugênio. Reflexões sobre ética & inteligência maquínica. In: SANTAELLA, Lúcia. **Simbioses do Humano & Tecnologias: Impasses, Dilemas, Desafios**. São Paulo: Editora Universidade de São Paulo/IEA-USP, 2022, p. 96.

um juízo de valor que ela tenha feito, não como negar que essas decisões têm efeitos éticos e, nesse sentido, é realmente espantoso que as máquinas possam decidir eticamente, mesmo em obediência a um ou outro humano. Como uma decisão que gera efeitos éticos pode ser deliberada por um agente inumano? A palavra “ética” vem do substantivo grego *ethos*. Esse termo, por sua vez, pode ser grafado com duas vogais diferentes, o etc e o épsilon. Numa ou noutra forma, os significados podem variar, mas em comum, *ethos*, em qualquer das grafias, quer dizer “costumes”, “hábito”, “tradição”. De maneira geral, *ethos* também pode significar “caráter” (de uma pessoa) ou “morada” (também de uma pessoa), e alude àquilo que é recorrente, seja como característica da personalidade (que tem a ver com temperamento de cada um), seja como lugar (a casa em que se mora). O que há de fecundo dessa palavra, “ética”, é o fato de ela nos trazer uma dimensão individual (o caráter de uma única pessoa) vinculada a uma dimensão coletiva e social (os costumes de uma cultura). A partir disso, para que possamos falar num plano. Ético da existência humana, é necessário que tenhamos em conta a tradição moral em que o sujeito se inscreve e, além disso, a capacidade individual de tomar decisões desse sujeito, em face da tradição moral em que ele se inscreve. A ideia de ética requer o solo comum dos costumes, assim como supões a incidência da escolha individual.<sup>143</sup>

A conclusão acima exposta guarda relação com a presente pesquisa, pois decisões automatizadas são influenciadas por decisões humanas de acordo com culturas, línguas, costumes, comportamentos, economia, região, população, emoções etc.

E prossegue Bucci:

O sujeito da ética, ou seja, qualquer pessoa, é definido assim como aquele capaz de tomar decisões que envolvem a escolha de valores articulada com o raciocínio prático. [...] A liberdade, a razão e a responsabilidade se impõem como pré-requisito em três faces para que possa existir a vida ética.<sup>144</sup>

O viés discriminatório, ou. Antiético, portanto, tem nas neurociências econômicas a revelação de que nossas decisões são frequentemente influenciadas e direcionadas antes mesmo das decisões automatizadas interferirem diariamente como fazem atualmente.

Desvios de cognitivos sobre nossa razão e liberdade, como a aversão à perda, o efeito de enquadramento e a heurística da disponibilidade, geram informações que tornam decisões automatizadas por vezes antiéticas.

Ao entender esses vieses, podemos projetar sistemas de decisão automatizada que os levem em consideração ou até mesmo os mitiguem. Por exemplo, um sistema de recomendação

---

<sup>143</sup> BUCCI, Eugênio. Reflexões sobre ética & inteligência maquina. In: SANTAELLA, Lúcia. **Simbioses do Humano & Tecnologias: Impasses, Dilemas, Desafios**. São Paulo: Editora Universidade de São Paulo/IEA-USP, 2022, p. 97.

<sup>144</sup> *Ibidem*, p. 98.

de investimentos pode ser projetado para apresentar as informações de forma neutra, evitando enquadramentos que induzam a decisões enviesadas.

Ao combinar dados comportamentais com modelos neuroeconômicos, podemos prever o comportamento humano com maior precisão. Isso é crucial para sistemas de decisão automatizada que interagem diretamente com pessoas, como *chatbots* de atendimento ao cliente ou sistemas de negociação automatizados, e desse modo conseguir uma previsibilidade do comportamento.

As neurociências econômicas podem ajudar a personalizar as decisões automatizadas de acordo com as preferências e características individuais. Ao analisar dados neurais e comportamentais, os sistemas podem aprender as preferências de cada usuário e adaptar suas decisões para melhor compreendê-las.

A pesquisa converge para o entendimento de que decisão humana, dados, perfilamento e sua consequente influência no aprendizado e resultado (*output*) da Inteligência Artificial, citando, inclusive, Doneda<sup>145</sup>, comprovando que a presente pesquisa e conexão sobre a formação da decisão automatizada tem como um de seus elementos a decisão humana, e como somos influenciados pelo Neuromarketing e pela Economia Digital, influenciadora de decisões diárias, que são inseridas e conectadas a Inteligência Artificial, que aprende e decide em manifestação similar (porém sem sentimentos e humanização) automatizada, veja-se:

Em geral, o perfil formado a partir do uso da tecnologia se torna uma representação virtual da pessoa e pode até mesmo ser confundido com ela. A utilização da técnica de *profiling*<sup>35</sup>, porém pode diminuir a liberdade das pessoas, por aqueles que se valem do perfil formado partem do pressuposto de que a pessoa tomará decisões com base em um padrão predefinido<sup>36</sup>. Por isso de grande aplicabilidade prática se mostra a previsão acerca da possibilidade de que o usuário não se sujeite a decisões tomadas exclusivamente de forma automatizada e também com base em *profiling*. Envolver uma pessoa nesses processos contribui para que cenários de discriminação possam ser identificados, evitando-se uma possível violação ao princípio da não discriminação, previsto no art. 6.º, inciso XI, da LGPD e na própria Constituição Federal (art. 3.º, inciso IV, e art. 5.º, inciso XLI).

Luciano Floridi estabelece quatro elementos éticos para a interação da Inteligência Artificial e a Sociedade, que são abaixo transcritos para encerramento deste capítulo para

---

<sup>145</sup> “Nela (técnica conhecida como *profiling*), os dados pessoais são tratados, com o auxílio de métodos estatísticos, técnicas de inteligência artificial e outras mais, com o fim de obter uma “metainformação”, que consistiria numa síntese dos hábitos, preferências pessoais e outros registros da vida desta pessoa. O resultado pode ser utilizado para traçar um quadro das tendências de futuras decisões, comportamentos e destinos de uma pessoa ou grupo.” DONEDA, Danilo Cesar Maganhoto. **Da Privacidade à Proteção de Dados Pessoais**. Rio de Janeiro: Renovar, 2006, p. 173-174.

demonstrar como a ética é um tema importante quando se trata de decisões e resultados proferidos por máquinas e sistemas computacionais inteligentes, abaixo traduzido por mim do italiano:

Como mencionado nos capítulos anteriores, não se trata mais de discutir se a IA terá ou não um impacto importante sobre indivíduos, sociedades e ambientes. O debate atual gira em torno de quão positivo ou negativo esse impacto será, para quem, de que maneira, em quais lugares e em qual horizonte temporal. Em outras palavras, no quarto capítulo, escrevi que as perguntas-chave agora são *quem, como, onde e quando* sentirá o impacto positivo ou negativo da IA. Para enquadrar essas perguntas de maneira mais substancial e prática, introduzo aqui aquelas que podem ser consideradas as quatro principais oportunidades que a IA oferece à sociedade. São quatro porque abordam os quatro pontos fundamentais da nossa antropologia filosófica, ou seja, da nossa compreensão da dignidade e do progresso da humanidade: a) A realização autônoma de nós mesmos, ou seja, quem podemos nos tornar; b) A ação humana, ou seja, o que podemos fazer; c) As capacidades individuais e sociais, ou seja, o que podemos alcançar; d) A coesão social, ou seja, como podemos interagir uns com os outros e com o mundo. Em cada caso, a IA pode ser utilizada para favorecer a natureza humana e suas potencialidades, gerando assim oportunidades, se subutilizada, criando assim custos de oportunidade; se superutilizada ou mal utilizada, criando assim riscos.<sup>146</sup>

## 2.2 RESULTADOS E DECISÕES ARTIFICIAIS

Os dados comunicados (inseridos por *inputs* humanos ou por meio de outros sistemas) do mundo exterior e dos comportamentos culturais da sociedade deve ser objeto de “rigoroso processo de coleta e seleção dos dados, configuração do algoritmo e avaliação de seus efeitos pessoais sensíveis, que poderão ser utilizados para controle de atuação do algoritmo e método de redução de impactos nocivos”.<sup>147</sup>

As atividades computacionais baseadas em dados determinam como os Algoritmos aprendem e decidem, determinando decisões, como leciona Abrusio, que assim discorre:

Do indivíduo ao sujeito digital, a vida atual lida com essa crescente demanda de experiências e encontros na Sociedade da informação, pautados por determinações algorítmicas. As características de uma determinada pessoa, hoje em dia, são mais detectáveis através da mineração e triangulação de múltiplas fontes vinculadas aos seus dados pessoais, com a afirmação de sua identidade por reconhecimento de grupos, redes sociais e outras presenças e digitais. Vê-se como a identidade, hoje, tem sido percebida por construções e padrões digitais. Nesse contexto, há uma forte influência da capacidade de persuasão dos processos algorítmicos, que buscam otimizar as interações

<sup>146</sup> FLORIDI, Luciano. **Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide**. Tradução: Massimo Durante. Milano – Itália: Raffaello Cortina Editore, 2022.

<sup>147</sup> KOGA, Bruno Yudi Soares. **Precificação Personalizada**. 1.ª ed. São Paulo: Almedina, 2021, p. 93.

humanas em um número crescente de setores. Os algoritmos, portanto, integram a percepção da construção da própria identidade. [...] Os indivíduos são comumente confrontados com uma assimetria informacional sobre quais tipos e quantos dados seus dispositivos, redes e plataformas geram, processam ou compartilham. Os diferentes tipos de aplicações existentes, associados ao uso de inteligência artificial podem afetar o direito à privacidade e proteção de dados pessoais em diferentes formas.<sup>148</sup>

O raciocínio acima coloca a inteligência artificial no rol de sistemas que produzem efeitos sociais, fazendo com que o indivíduo interage com o sistema.

A pesquisa investiga se decisões automatizadas (descritas abaixo), principalmente em aprendizado profundo e redes neurais complexas (*Deep Learning*), viabilizam a ocorrência de vieses, discriminações e problemas éticos, prejudicando seres humanos – como indivíduos ou grupos.

O processo de tomada de decisão automatizada geralmente envolve as seguintes etapas, como: (i) Coleta de dados: os sistemas coletam dados relevantes para a decisão, que podem incluir informações pessoais, histórico de transações, dados de sensores, entre outros; (ii) Processamento de dados: os dados coletados são processados e analisados por algoritmos, que identificam padrões e tendências, (iii) Geração de modelo: com base nos dados processados, o sistema gera um modelo que representa a relação entre os dados e a decisão a ser tomada. Esse modelo pode ser uma árvore de decisão, uma rede neural, ou outro tipo de algoritmo; (iv) Tomada de decisão: o sistema utiliza o modelo gerado para tomar a decisão automaticamente, com base nos dados de entrada; (v) Avaliação e monitoramento: o desempenho do sistema é avaliado e monitorado para garantir que as decisões tomadas sejam precisas e justas.

Cita a pesquisa exemplos de decisões automatizadas: (i) Análise de crédito: bancos e instituições financeiras utilizam sistemas automatizados para analisar o risco de crédito de clientes e decidir sobre a concessão de empréstimos, (ii) Recomendação de produtos: plataformas de e-commerce e streaming utilizam algoritmos para recomendar produtos e conteúdos aos usuários, com base em seus históricos de compras e preferências, (iii) Detecção de fraudes: sistemas automatizados são utilizados para detectar fraudes em transações financeiras e outros tipos de atividades, (iv) Diagnóstico médico: algoritmos de inteligência artificial podem auxiliar médicos no diagnóstico de doenças, com base em imagens e dados de pacientes.

---

<sup>148</sup> ABRUSIO, Juliana. **Proteção de Dados na Cultura do Algoritmo**. 1.<sup>a</sup> ed. São Paulo: D'Plácido, 2020, p. 208. (Coleção Direito privado, v. 1).

As decisões automatizadas trazem consigo desafios éticos e legais, como a necessidade de garantir a transparência, a responsabilidade e a justiça no processo de tomada de decisão. É importante que os sistemas sejam desenvolvidos e utilizados de forma responsável, para evitar a discriminação e o enviesamento algorítmico.

Quanto ao cenário Europeu, o Regulamento Geral de Proteção de Dados da União Europeia (RGPD), cuja sigla em língua inglesa ficou conhecida como G.D.P.R – *General Data Protection Regulation*, normatizando sobre o tratamento e perfilamento, a saber:

Artigo 22. Decisões individuais automatizadas, incluindo definição de perfis: 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.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 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. 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. 2 não se baseiam nas categorias especiais de dados pessoais a que se refere o artigo 9º, n. 1, a não ser que o n. 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.<sup>149</sup>

O artigo da lei europeia acima mencionado refere-se às decisões automatizadas que são tomadas pelos sistemas de Inteligência Artificial levando-se em conta padrões de personalidade compostas por dados pessoais obtidos por meio de coleta espontânea do indivíduo que preenche cadastros, e também por dados pessoais coletados de aparelhos que enviam dados pessoais de caráter privado para a base de dados do *Big Data* por meio de aparelhos eletrônicos conectados à Internet, como as:

Incluídas as situações em que há definição de perfis (*profiling*). As organizações em geral obtêm informações pessoais oriundas de uma ampla gama de fontes. Apenas para dar alguns exemplos de fontes relativas à obtenção de informações pessoais mencionam-se as que decorrem de buscas na Internet, hábitos de compras, estilo de vida e comportamento colhidos da utilização *smarthphones*, além de utilização de redes sociais, capturas de

---

<sup>149</sup> EUROPA. **General Data Protection Regulation**. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504>. Acesso em: 14 nov. 2022.

imagens de vídeo e internet das coisas (IoT). Tais informações coletadas são analisadas para a classificação de pessoas em diferentes grupos ou setores, o que se dá por meio da utilização de algoritmos e de *machine learning*. Com base nessas análises, são criados perfis que revelam preferências individuais e que também predizem comportamentos e tomam decisões acerca deles. Essa espécie de decisão automatizada mostra-se eficiente para alguns específicos setores empresariais, mas por outro lado, representam potenciais riscos aos titulares de dados, de modo que o GDPR busca minorá-los de forma a proteger os indivíduos.<sup>150</sup>

Tateoki, sobre dados e o perfilamento, indica a indução dos usuários-consumidores:

Em um mundo onde a enxurrada de informações, a grande competitividade e variedade do mercado, para se destacar, se diferenciar ou ter atenção dos consumidores, é necessário compreender seus hábitos para oferecer produtos ou serviços mais personalizados. Desta forma, há a necessidade da coleta de dados para compreender os consumidores e predizer o que eles querem. Assim, surge a prática do *profiling*, que é construir perfis das pessoas e enquadrá-las em grupos e padrões, que muitas vezes acabam estigmatizando as pessoas. [...] Tal visão, parece tentador, mas al longo prazo pode ocasionar inúmeros problemas. Nós humanos não somos racionais ao ponto de perceber que todas essas maravilhas podem estar afetando os nossos modos de vida e como pensamos. Não somos seres vivos somente racionais, como visto anteriormente, podemos ser influenciados e manipulados. Dessa forma, as plataformas digitais usem *nudges* para manter mais ainda atenção do usuário filtrando a informação. Cada vez mais nossa atenção está sendo redirecionada para as plataformas digitais, nosso Sistema Reflexivo está perdendo a capacidade de se concentrar, podendo, como visto anteriormente, acabar em cegueira e não perceber o que está acontecendo ao redor. Nosso Sistema Automático acaba se tornando o principal ao utilizar dessas plataformas. Acontece que ele possui heurísticas e vieses.<sup>151</sup>

Alinhar a Inteligência Artificial quando tratamos de valores éticos é fundamental para o desenvolvimento de decisões, pois segundo Tegmark:

Quanto mais inteligentes e poderosas as máquinas se tornam, mais importante é que seus objetivos estejam alinhados com os nossos. Enquanto construirmos apenas máquinas relativamente estúpidas, a questão não é se os objetivos humanos prevalecerão no final, mas apenas quantos problemas essas máquinas podem causar à humanidade antes de descobrirmos como resolver o problema de alinhamento de objetivos. Se uma superinteligência for desencadeada, no entanto, será o contrário: como a inteligência é a capacidade de atingir objetivos, uma IA superinteligente é, por definição, muito melhor em realizar seus objetivos do que nós, humano, em realizar os nossos e, portanto, prevalecerá. [...] Em outras palavras, *o risco real da IAG não é a maldade, mas a competência*. Uma IA superinteligente será extremamente boa

<sup>150</sup> BLUM, Renato Opice; MALDONADO, Viviane Nóbrega. **Comentários a GDPR, Regulamento Geral de Proteção de Dados da União Europeia**. São Paulo: Thomson Reuters, 2018, p. 106.

<sup>151</sup> TATEOKI, Victor Augusto. **O Uso dos Dados Pessoais como Mecanismo de Persuasão no Processo de Tomada de Decisão dos Usuários de Internet**. Rio de Janeiro: Lumen Iuris, 2021, p. 176-178.



em atingir seus objetivos e, se esses objetivos não estiverem alinhados com os nossos, teremos problemas. [...] Descobrir como alinhar os objetivos da IA superinteligente com os nossos não é apenas importante, é também difícil. De fato, é atualmente. Um problema não resolvido. Ele se divide em três subproblemas difíceis, e cada um deles é objeto de pesquisa ativa para cientistas da computação e outros pensadores: 1. Fazer a IA *aprender* nossos objetivos; 2. Fazer a IA *adotar* nossos objetivos; 3. Fazer a IA *reter* nossos objetivos.<sup>152</sup>

A racionalidade e a ética estão unidas para estabelecer uma relação com o universo digital, que transforma nossas decisões em dados, e que influenciam e transformam as decisões automatizadas para um sentido ético quando os usuários, no dizer de Pinker, indica que:

Deveríamos nos importar com a virtude das pessoas quando as vemos como amigas, mas não quando consideramos as ideias que transmitem. As ideias são verdadeiras ou falsas, coerentes ou contraditórias, promotoras do bem-estar ou não, não importa quem as tenha dito. [...] O poder da racionalidade para conduzir o progresso moral é parte integrante de seu poder para conduzir o progresso material e escolhas prudentes em nossa vida. Nossa capacidade para conquistar com esforço incrementos de bem-estar a partir de um cosmo impiedoso e para ser generosos com os outros apesar da nossa natureza falha depende da compreensão de princípios imparciais que transcendem nossa experiência restrita. Somos uma espécie equipada com uma faculdade elementar de raciocínio que descobriu fórmulas e instituições que amplificam seu alcance. Elas nos despertam para ideias e nos expõe a realidades que confundem nossas intuições, mas que mesmo assim são verdadeiras.<sup>153</sup>

A Inteligência Artificial deve ser utilizada para beneficiar o humano em suas mais variadas categorias e funções, Russel esclarece sobre evolução da IA, e como é preciso auxiliar os indivíduos, usuários diretos ou indiretos, a tomarem decisões pautadas em mais informações para evitar o mau uso da tecnologia de IA:

Não é de se surpreender, portanto, que agências de inteligência artificial já tenham percebido o potencial de utilização da IA em seu trabalho. Durante anos, elas vêm aplicando formas simples de tecnologia de IA incluindo reconhecimento de fala e identificação de palavras-chave tanto na fala como no texto. Os sistemas de IA são cada vez mais capazes de *compreender o conteúdo* do que as pessoas dizem ou fazem, seja na vigilância de fala, de texto ou de vídeo. [...] Uma vez que as aptidões de vigilância estejam prontas, o passo seguinte é modificar nosso comportamento para a conveniência daqueles que estiverem empregando essa tecnologia. Um método muito tosco é a chantagem automatizada, personalizada: um sistema que entenda o que você está fazendo – seja por escuta, leitura ou observação – pode facilmente

<sup>152</sup> TEGMARK, Max. **Vida 3.0: o ser humano na era da inteligência artificial**. Tradução: Petê Rissati. São Paulo: Benvirá, 2020, p. 266-267.

<sup>153</sup> PINKER, Steven. **Racionalidade: o que é, porque parece estar em falta, por que é importante**. Tradução: Waldéa Barcellos. 1.ª ed. Rio de Janeiro: Intrínseca, 2022, 360-361.

identificar coisas que você não deveria estar fazendo. Tendo descoberto alguma coisa, ele começa a se corresponder com você para arrancar a maior quantidade possível de dinheiro (ou para coagir seu comportamento, se o objetivo for controle político ou espionagem). A extorsão de dinheiro funciona como perfeito sinal de recompensa para um algoritmo de aprendizado por reforço, portanto, podemos esperar um progresso rápido nos sistemas de IA em sua capacidade de identificar o mau comportamento e lucrar com ele. [...] Uma maneira mais sutil de mudar o comportamento das pessoas é modificar seu ambiente de informações, de maneira que passem a acreditar em coisas diferentes e a tomar decisões diferentes.<sup>154</sup>

O conhecimento das habilidades cognitivas, de escolhas e decisões humanas é campo de pesquisa e estudo da teoria da mente, capacitando agentes inteligentes com tais dados sobre crenças, programação neurolinguística, estado emocional, cultural etc:

Uma das habilidades humanas mais importantes é a teoria da mente. Indivíduos com teoria da mente entendem crenças, as emoções e os estados mentais dos outros. Assim, eles têm consciência de que a perspectiva ou as crenças de outra pessoa muitas vezes diferem das suas. A teoria da mente é de enorme valor na interação social e na comunicação.<sup>155</sup>

Indivíduos que aprendem sobre outros indivíduos podem, por conclusão, serem máquinas. É possível assim direcionar as interações de indivíduos ou grupos para que interajam, adquiram produtos e serviços por meio de aplicativos e plataformas, que ao coletarem e tratarem tais dados diretamente em sistemas de inteligência artificial permitem que decisões automatizadas possam ser direcionadas.

Este capítulo demonstra que sistemas computacionais inteligentes aprendem e decidem influenciados por uma série de informações e dados pessoais, que são parte do Direito à Explicação quando um usuário-cidadão necessita de respostas sobre as decisões automatizadas.

---

<sup>154</sup> RUSSEL, Stuart. **Inteligência Artificial a nosso favor: Como manter o controle sobre a tecnologia**. Tradução: Berilo Vargas. 1.ª ed. São Paulo: Companhia das Letras, 2021, p. 104-105.

<sup>155</sup> EYSENCK, Michel W.; EYSENCK, Christine. **Inteligência Artificial x Humanos: o que a ciência cognitiva nos ensina ao colocar frente a frente à mente humana e a IA**. Tradução: Gisele Klein. Revisão Técnica: Vitor Geraldini Haase. Porto Alegre: Artmed, 2023, p. 83.

### 3 INSTRUMENTOS DE REGULAÇÃO CONTEMPORÂNEA DA INTELIGÊNCIA ARTIFICIAL

#### 3.1 COMPLIANCE DIGITAL

Verifica-se no capítulo anterior a importância da adoção de diretrizes e formatações para que tenhamos legislações estatais que dialogam com regulações *ex ante*, e normas de integridade e *Compliance*, como adiante será exposto.

A presente pesquisa partiu da premissa demonstrar de que existem condições de estabelecer boas práticas e ações positivas para melhor utilizar e compreender os dados e informações (pessoais e corporativos) e sua relação com a Inteligência Artificial.

Boas práticas diminuem riscos e fragilidades existentes em organizações que contratam, licenciam e utilizam softwares e tecnologias com (e de) Inteligência Artificial, e que podem melhorar e tornar tais sistemas inteligentes mais fortes e confiáveis em suas decisões (e consequências) proferidas pelos computadores inteligentes.

Analisando meios e formas mitigar riscos, melhorar a qualidade dos dados para inputs, bem como tornar o desenvolvimento e utilização da Inteligência Artificial mais transparente e confiável, pesquisei boas práticas para a utilização de tecnologia.

O Direito tem exercido seu papel, seja por meio de normas enunciadas publicamente (leis), seja por meio de estruturas para incentivo de cumprimento de normas privadas de conduta, ética e governança, apoiadas por políticas de Estado, incentivos, ações afirmativas, princípios de Direito aplicáveis e pela legislação, como aponta Santana: Por fim, pode-se constatar que o Direito tem relevante papel a exercer no atual mundo do algoritmo, principalmente na normatização, estabelecendo princípios e regras a serem seguidas.<sup>156</sup>

Os sistemas computacionais inteligentes são ferramentas pensadas e desenvolvidas para ajudar os seres humanos, o Estado e a sociedade, cumprindo seu papel inovador e promissor de evolução e revolução industrial.

Todavia, como visto linhas atrás, precisamos de diretrizes e formatações para cumprimento de garantias e Direitos dos indivíduos e de toda a coletiva sujeita a decisões automatizadas.

---

<sup>156</sup> SANTANA, Paulo Campanha. O papel do Direito no Mundo do Algoritmo. In: AVANCI, Thiago Felipe S. **O Futuro do Direito: o que esperar do Direito e das Tecnologias, Estudo em homenagem ao professor Angelo Vighianisi Ferraro**. Londrina: Thoth, 2022, p. 51.

Atualmente vivemos em um mundo digitalizado e em claro processo de digitalização constante, com muito acesso a aparelhos chamados de “inteligentes”, justamente em virtude do processamento e troca de informações, como expõe Bagnoli:

Desde o início do séc. XXI diversas inovações tecnológicas e o maior acesso aos aparelhos digitais viabilizaram a Revolução Industrial de Dados (*Industrial Revolution of Data*) caracterizada pela enorme difusão em quantidade e diversidade de dados digitais em tempo real. A tecnologia digital em quantidade e diversidade de dados digitais em tempo real. A tecnologia digital cada vez mais presente na vida das pessoas permite um aprendizado sem precedentes do comportamento humano e incluiu a internet na forma como as pessoas se comunicam, relacionam-se, divertem-se e trabalham. Contudo, engana-se quem pensa que o Big Data seja apenas umas “dessas revoluções tecnológicas” que cada quatro ou cinco anos surgem e são incorporadas pelas empresas pelas empresas como a solução de todos os problemas para a sua consolidação nos mercados.<sup>157</sup>

A economia digital do presente e do futuro precisa de Boas Práticas de governança como ponto imprescindível para maior controle sobre a produção de dados que são objeto de coleta, tratamento e armazenamento das instituições privadas e públicas, em todos os setores e áreas que utilizam softwares conexos com sistemas inteligentes ou mesmo que utilizem Inteligência Artificial direta.

As boas práticas de *compliance* digital visam não somente mitigar riscos decorrentes das tecnologias da informação, mas prever e dar resposta a possíveis incidentes e danos, tornando o ambiente digital mais seguro para os usuários, sejam eles cidadãos ou instituições, tornando a inovação e o desenvolvimento mais sustentável, tanto do indivíduo quanto da sociedade.

A descrição sobre a economia digital e as plataformas é assim descrita por Bagnoli:

A economia digital se caracteriza por sua competição dinâmica essencial em ciclos contínuos de inovação, desenvolvimento e rupturas. Baseada em dados (DDI; *data-driven-innovation*, em inglês) a inovação se traduz na melhoria significativa de um produto, processo, método organizacional ou mercado, ou o seu desenvolvimento, realizado pelo uso do Big Data. As plataformas digitais, ou mercados de dois lados e até mesmo multimercados têm aumentado significativamente a relevância para a interação entre diferentes grupos de usuários, tais como: consumidores ou usuários dos serviços prestados pela plataforma; empresas que exibem anúncios aos consumidores na plataforma, e criadores de conteúdo, como notícias. [...] O que torna as plataformas tão únicas é o Big Data, cujo uso promove um ciclo de retroalimentação (*feedback loop*, em inglês) permitindo que as empresas

---

<sup>157</sup> BAGNOLI, Vicente. **Direito econômico e concorrencial – O Poder Econômico e seus Limites Jurídicos**. 9.ª ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2022, p. 505-506.

criem produtos, ofereçam serviços e atraiam cada vez mais usuários. Isso assegura economias substanciais de escala e impulsionando o mercado em favor das plataformas já dominantes, que não só participam de uma miríade de negócios como smartphones, e-commerce, pagamentos digitais, mas também são se envolvem no desrespeito à privacidade de dados, desinformação, questões trabalhistas e interferência eleitoral, inclusive abalando princípios da democracia.<sup>158</sup>

Para maior clareza sobre essa importante questão relacionada aos dados aplicados aos algoritmos, pois não temos os mesmos resultados de Inteligência Artificial, ou seja, não são os algoritmos necessariamente os culpados por vieses ou questões éticas de forma exclusiva, o conjunto de dados tem significativa influência em um modelo, como abaixo descrito por Sicsu, Amartini e Barth:

Modelo é o produto que resulta quando aplicamos um algoritmo a uma base de dados. Em um sentido mais amplo, o modelo deve incluir não só valores fornecidos pelo algoritmo, mas também, caso necessário, regras a serem seguidas para utilizá-lo nas previsões ou classificações dos dados da população. Um mesmo algoritmo, quando aplicado a diferentes conjuntos de dados, fornece diferentes modelos.<sup>159</sup>

O cumprimento de normas é imprescindível para as sociedades modernas e que estão em interação constante com Inteligência Artificial como formas de mitigar riscos e melhorar a resposta pós inserção de dados e uso pelos algoritmos, que aprendem conforme processam e analisam padrões, para posteriormente emitirem decisões, sendo razoável considerar que o *Compliance* auxilia em qualquer setor ou negócio como instrumento garantido, conforme lecionam Pinto, Azzari e Silva:

A amplitude setorial da aplicação do *compliance* faz com que sejam preservadas as bases, princípios éticos e normas de conduta que regem a sua formação, sendo sua aplicação alterada de acordo com o segmento de atuação, segundo os seus objetivos e as suas operações, de modo a garantir as boas práticas de conduta em seu respectivo eixo.<sup>160</sup>

E sobre o tema os autores supra complementam que é justamente:

<sup>158</sup> BAGNOLI, Vicente. **Direito econômico e concorrencial – O Poder Econômico e seus Limites Jurídicos**. 9.<sup>a</sup> ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2022, p. 508.

<sup>159</sup> SICSU, Abraham Laredo; SAMARTINI, André; BARTH, Nelson Lerner. **Técnicas de Machine Learning**. São Paulo: Blucher, 2023, p. 17.

<sup>160</sup> PINTO, Felipe Chiarello de Souza; AZZARI, Bruna; SILVA, Amanda Scalisse. Ética e Integridade nas Instituições de Ensino Superior: A importância da Implementação de Programas de *Compliance* nas Universidades. **Revista de Direito Brasileira**, Florianópolis, v. 26, n. 10, p. 464, mai./ago. 2020. Disponível em: <https://www.indexlaw.org/index.php/rdb/article/view/6252/5122>. Acesso em: 24 abr. 2024.

A partir da análise dos riscos, é criado o programa de *compliance*, estabelecendo a criação de procedimentos, controles e mecanismos de gestão, com o objetivo de minimizar os riscos. Destaca-se a definição de áreas de *compliance*, com gestores responsáveis por cada setor, assegurando que todos os cenários de risco estejam protegidos, podendo ser diminuídos ou eliminados, resguardando a integridade da instituição.<sup>161</sup>

Para implementar instrumentos de *Compliance* específicos para desenvolvimento, fornecimento, aquisição, interação com softwares (APIs), utilização de Inteligência Artificial, o *Compliance* também é forma de mitigar riscos e melhorar as “saídas” após trabalho interno do sistema de lógica algorítmica.

Na obra supre mencionada, Bagnoli leciona sobre a causalidade existente entre o uso dos algoritmos (que funcionam a base de informações produzidas por pessoas, empresas e setor público), reforçando a necessidade de disseminação da aplicação de instrumentos de *compliance* digital e ações afirmativas:

O uso de algoritmos pode causar danos diretos aos consumidores, por exemplo os utilizados para personalizar preços e manipular os consumidores em suas escolhas. O uso de algoritmos também pode incorrer em práticas exclusionárias, ao excluir concorrentes, consciente ou inconscientemente.<sup>162</sup>

A pesquisa propõe um olhar do *Compliance* Digital como pressuposto diminuir fragilidades e incidentes com utilização de tecnologias, como softwares, blockchain, criptoativos, Inteligência Artificial etc., tornando a utilização dos dados e dos algoritmos menos prejudiciais.

Como início de preparação e manutenção da utilização de tecnologias por instituições e pessoas, temos a Segurança da Informação como um fator basilar de proteção dos usuários (jurídicos e físicos).

A Segurança da Informação é parte extremamente importante desse processo *ex ante* de *Compliance Digital* para proteção de um ativo essencial para negócios e relações jurídicas de uma organização.

---

<sup>161</sup> PINTO, Felipe Chiarello de Souza; AZZARI, Bruna; SILVA, Amanda Scalisse. Ética e Integridade nas Instituições de Ensino Superior: A importância da Implementação de Programas de *Compliance* nas Universidades. **Revista de Direito Brasileira**, Florianópolis, v. 26, n. 10, p. 465, mai./ago. 2020. Disponível em: <https://www.indexlaw.org/index.php/rdb/article/view/6252/5122>. Acesso em: 24 abr. 2024.

<sup>162</sup> BAGNOLI, Vicente. **Direito econômico e concorrencial – O Poder Econômico e seus Limites Jurídicos**. 9.<sup>a</sup> ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2022, p. 521.

A Segurança da Informação tem como norma internacional de proteção do conhecimento da organização (sobre alguém ou algo), cujos parâmetros devem ser observados como aponta Teixeira:

Para a NBR 27002:2005, a segurança da informação consiste na **preservação dos princípios básicos da confidencialidade, da integridade e da disponibilidade** da informação: **Confidencialidade:** o acesso à informação deve ser obtido apenas por pessoas autorizadas; **Integridade:** as informações em trânsito ou em sistema de computador somente podem ser modificadas pelas partes autorizadas; **Disponibilidade:** as informações podem ser acessadas pelas pessoas autorizadas sempre que for necessário. Adicionalmente, outras propriedades podem estar envolvidas como: **Autenticidade:** a origem da informação deve ser identificada e o seu remetente deve ser realmente a pessoa indicada na própria mensagem; **Responsabilidade:** capacidade de se responsabilizar um usuário pelos seus atos, no tratamento de informações; **Irretratabilidade ou não repúdio:** quem enviou uma. Informação não poderá negar que a enviou; **Confiabilidade:** garantia de tolerância a falhas de um sistema de informação.<sup>163</sup> (grifo nosso).

Assim, Segurança da Informação por Bastos e Caubit:

Para melhor definir o conceito de Segurança da Informação será necessário citar três destaques: 1) Preservação da confidencialidade, integridade e disponibilidade da informação; adicionalmente, outras propriedades, tais como autenticidade, responsabilidade, não repúdio e confiabilidade, podem também estar envolvidas. (ABNT NBR SAI/IEC 27002: 2005); 2) Podemos definir a Segurança da Informação como uma área do conhecimento dedicada à proteção de ativos da informação contra acessos não autorizados, alterações indevidas ou sua indisponibilidade. (Sêmola, 2003, p. 43); 3) A segurança de informação é caracterizada pela aplicação adequada de dispositivos de proteção sobre um ativo ou um conjunto de ativos visando preservar o valor que este possui para as organizações. A aplicação destas proteções busca preservar a confidencialidade, a integridade e a disponibilidade (CID), não estando restritos somente a sistemas ou aplicativos, mas também informações armazenadas ou veiculadas em diversos meios além do eletrônico ou em papel.<sup>164</sup>

Os destaques acima não esgotam os princípios quando tratamos de melhorar a conformidade da segurança informacional, considerando a soma de atributos proposto por Donn B. Parker, conhecido por Hexagrama Parkeriano, explicado abaixo conforme Hintzbergen:

O hexagrama Parkeriannon, ou Parkeriann hexad, é o conjunto de seis elementos da segurança da informação proposto por Donn B. Parker. O termo

<sup>163</sup> TEIXEIRA FILHO, Sócrates Arantes. **Segurança da Informação Descomplicada**. 1.<sup>a</sup> ed. Brasília-DF: 2015, p. 1.

<sup>164</sup> BASTOS Alberto; CAUBIT, Rosângela. **Gestão de Segurança da Informação. SAI 27001 e 27002 - Uma Visão Prática**. Rio Grande do Sul: Zouk, 2009, p. 17.

foi cunhado por M.E. Kabay. O hexagrama Parkeriano soma três atributos clássicos de segurança do triângulo CIA (confidencialidade, integridade, disponibilidade – ou confidentiality, integrity, availability). Em segurança da informação, um backup ou processo de fazer backup se refere a fazer cópias dos dados de forma que essas cópias adicionais possam ser usadas para restaurar o original.<sup>165</sup>

O backup e o hexagrama Parkeriano são aliados da Inteligência Artificial, quando pensamos em análise, auditoria da governança de dados, resposta a questionamentos, e sem dúvida, auxiliar na proteção dos indivíduos e organizações, e certamente como parte dos documentos que poderão fornecer subsídios para o Direito à Explicação.

A segurança da informação é fundamental para o *Compliance* Digital da organização, pois propõe dentro do cumprimento de normas SAI e princípios acima descritos, o gerenciamento de riscos, possibilitando que a gestão da instituição possa diminuir e organizar seus processos e ecossistema, de forma ostensiva, gerenciando, portanto, as formas que a organização promoverá a proteção das informações, documentos, dados, ou sejam, seus ativos informacionais.

Hintzbergen também aponta a importância gerenciamento de riscos de informação:

Gerenciamento de riscos é o processo de planejar, organizar, conduzir e controlar as atividades de uma organização visando minimizar os efeitos do risco sobre o capital e o lucro de uma organização. [...]. Diversos padrões de gerenciamento de riscos foram desenvolvidos, incluindo os do Project Management Institute (PMI), National Institute of Science and Technology (NIST). Métodos, definições e objetivos variam muito – por exemplo se o método de gerência de riscos se encontra no contexto da gerência de projetos, segurança, engenharia, processos industriais, carteiras financeiras, avaliações atuariais ou segurança e saúde pública. A estratégia de riscos pode incluir transferir o risco para outra parte, evitar o risco, reduzir o efeito negativo do risco e aceitar algumas ou todas as consequências de um risco em particular. Gerenciamento de riscos é um processo contínuo que se aplica a todos os aspectos dos processos operacionais. Em grandes organizações, a tarefa de monitorar esse processo é conduzida por um especialista em segurança da informação, tal como um Encarregado de Segurança da Informação (Information Security Officer – SAI) ou Chefe de Segurança da Informação (Chief Information Security Officer – CISO), que é designado especialmente para essa função e responde pelo mais alto nível de gestão.<sup>166</sup>

A gestão da segurança da informação, como parte do *Compliance* Digital é a continuidade do negócio e as consequências e impactos que podem afetar a instituição em caso

---

<sup>165</sup> HINTEZBERGEN, Jule *et al.* **Fundamentos da Segurança da Informação: com base na ISO 27001 e na ISO 27002**. Tradução: Alan de Sá. Rio de Janeiro: Brasport, 2018, p. 27.

<sup>166</sup> *Ibidem*, p. 30.



de incidente de segurança da informação, e como Política e cumprimento dos processos de Segurança da informação podem ajudar a responder e auxiliar uma organização na manutenção de suas atividades, comprovando que conformidade e integridade dos processos de segurança é fundamental, o autor acima citado aponta que é necessário:

Pensar de forma antecipada sobre a continuidade dos processos do trabalho é essencial para uma organização. [...] O propósito da gestão de continuidade dos negócios (Business Continuity Management – BCM) é prevenir que as atividades da empresa sejam interrompidas, proteger processos críticos das consequências de grandes perturbações nos sistemas de informação e permitir uma rápida recuperação. [...] As consequências dos desastres e dos incidentes de segurança e as falhas dos serviços são avaliadas em uma Análise de Impacto no Negócio, ou Business Impact Analysis (BIA). O plano de continuidade descreve como a informação requerida por processos críticos de negócio pode ser rapidamente disponibilizada. Na segurança da informação, a gestão da continuidade é normalmente dividida em dois componentes separados, mas intimamente relacionados: Planejamento de Continuidade do Negócio, ou Business Continuity Planing (BCP), onde a continuidade do processo do negócio é garantida; Planejamento de Recuperação de Desastres, ou Disaster Recovery Planing (DRP), onde é organizada a recuperação após o desastre. A gestão de continuidade dos negócios é descrita na SAI 27031:2012. Embora a ISSO/IEC 27002:2013 inclua algumas medidas de BCM, elas visam principalmente a informação, enquanto a ISSO 27031 deve ser aplicada integralmente em toda a organização.<sup>167</sup>

A conformidade com políticas e padrões de segurança da informação é parte do *Compliance* Digital para utilização positiva da Inteligência Artificial, devendo a instituição estabelecer qual (ou quais) são os melhores frameworks se ajustam a seus interesses e ativos informacionais, pois “Existem muitas organizações e padrões sobre segurança da informação. Padrões importantes são desenvolvidos pela SAI, NIST E ANSI”.<sup>168</sup>

O cumprimento das normas regulatórias conjuntas para proteção dos indivíduos hipossuficientes em relações digitais, e deve ser objeto de desenho específico para cada instituição, posto que cada organização, negócio ou instituição – pública ou privada – e mesmo pessoa física que colete dados para fins econômicos (como um advogado, médico ou odontólogo que atue solitariamente em sua profissão), deve necessariamente adequar seus processos empresariais e administrativos aos ditames regulatórios próprios.

---

<sup>167</sup> HINTEZBERGEN, Jule *et al.* **Fundamentos da Segurança da Informação: com base na ISSO 27001 e na ISSO 27002**. Tradução: Alan de Sá. Rio de Janeiro: Brasport, 2018, p. 164-165.

<sup>168</sup> *Ibidem*, p. 178.

A ideia de proteção da privacidade vem desde a publicação do artigo “The Right to Privacy”.<sup>169</sup>

A proteção de dados pessoais – desde a Proposta de Emenda Constitucional 115/2022 – advém do paradigma central da regulação de dados pessoais que aqui a pesquisa colaciona:

[...] A proteção de dados pessoais é fruto da evolução histórica da própria sociedade internacional: diversos são os Países que adotaram leis e regras sobre privacidade e proteção de dados. Isso porque o assunto, cada vez mais, na Era informacional, representa riscos às liberdades e garantias individuais do cidadão. O avanço da tecnologia, por um lado, oportuniza racionalização de negócios e da própria atividade econômica: pode gerar empregabilidade, prosperidade e maior qualidade de vida. Por outro lado, se mal utilizada ou se utilizada sem um filtro prévio moral e ético, pode causar prejuízos incomensuráveis aos cidadãos e à própria sociedade, dando margem, inclusive, à concentração de mercados. Por isso, países de todo o planeta já visualizaram a importância e imprescindibilidade de se regular juridicamente o tratamento de dados dos cidadãos.<sup>170</sup>

Os dados, as informações de privacidade e as informações pessoais, exclusivas dos indivíduos na sociedade da informação, formam nossa personalidade que precisa da promoção das normas de proteção de privacidade e proteção de dados, pois as informações pessoais provenientes da esfera de personalidade, cujas técnicas de análise de dados e *Business Intelligence*, aliadas ao processamento automatizado por computadores possuem grande capacidade de coleta, tratamento e armazenamento, uma vez que:

A captura de dados se dá basicamente de duas formas: 1) mediante o fornecimento pelo usuário de dados, quando do preenchimento de formulários em sites; 2) por meio da captação dos cookies, [...]. Eles, os cookies, são instalados, pelos sites visitados, dentro do computador do usuário, no disco rígido, com informações que o próprio usuário forneceu. Exemplos: nome, senha, e-mail, preferências (de compras, notícias etc) endereço, nacionalidade, profissão, estado civil, data de nascimento ou casamento, idade dos filhos, se porta doenças, dados comerciais em geral, entre outras. Eles têm a função de reconhecer o usuário, o que pode ser positivo. No entanto, tornam-se preocupantes na medida em que são utilizados indiscriminadamente e sem o devido esclarecimento aos usuários.<sup>171</sup>

<sup>169</sup> WARREN, Samuel D.; BRANDEIS, Louis D. The Right to Privacy. *Harvard Law Review*, v. 4, n. 5, p. 193-220, 15 dec. 1890. Disponível em: <https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf>. Acesso em: 28 abr. 2024.

<sup>170</sup> BRASIL. Câmara dos Deputados. Comissão de Constituição e Justiça e de Cidadania. **Parecer acerca da Proposta de Emenda à Constituição nº 17/2019**, p. 1. Disponível em: [https://www.camara.leg.br/proposicoesWeb/prop\\_mostrarintegra;jsessionid=node01vqwnuiffbkdp1v8m3yk5aotpe37875732.node0?codteor=1791032&filename=Tramitacao-PEC+17/2019](https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=node01vqwnuiffbkdp1v8m3yk5aotpe37875732.node0?codteor=1791032&filename=Tramitacao-PEC+17/2019). Acesso em: 7 dez. 2024.

<sup>171</sup> TEIXEIRA, Tarcísio. **Direito Digital e Processo Eletrônico**. 8.ª ed. São Paulo: SaraivaJur, 2024, p. 20.

O Princípio da Autodeterminação Informativa não é suficiente por si só para garantir proteção aos dados e a personalidade, uma vez que estamos na era dos dados e dos sistemas de Inteligência Artificial.

O consentimento deve, de acordo com a enunciação da legislação de proteção de dados e normas de diálogo ou complementação, ser devidamente esclarecido e livre por parte do agente ou instituição no ato da coleta para que o titular de dados possa exercer o denominado princípio da autodeterminação e escolha, sob pena de serem consideradas nulas porque são abusivas ou enganosas ao consumidor ou no caso de instituições e agentes de mercado que não estejam em conformidade com o art. 9.º, § 1.º da Lei Geral de Proteção de Dados Pessoais.

A aplicabilidade do Código de Defesa do Consumidor descreve como dever do fornecedor, conforme artigo 4.º, apresentar sumário do contrato antes da contratação, o que se aplica aos Termos de Uso e Contratação de empresas digitais.

Como os dados são o principal ativo necessário para aprendizado e decisão das máquinas, não seria possível tratar do tema do Direito à Explicação sem que se apliquem as normas enunciadas na Lei Geral de Proteção de Dados (Lei n.º 13.709/2018) para prevenção e mitigação de riscos no desenvolvimento e uso da IA, mas também para as demais tecnologias, uma vez que o framework quando implementado, monitorado e atualizado tem como escopo tudo ao que se refere a inovação e tecnologias, vez que a IA é de uso geral e se conecta com todas as demais tecnologias e hardwares.

O framework de proteção de dados pessoais contribui para proteção de dados da organização, e para ambientes mais confiáveis e responsivos, portanto, conectados a positivamente com o Direito à Explicação e a Explicabilidade.

É um dever da organização, cumprir a regulação estatal (lei) e as normas enunciadas em documentos de instituições especializadas com o objetivo de auxiliar organizações na busca por conformidade, integridade e melhoria do ambiente digital, que trazem benefícios para o indivíduo-usuário, para o coletivo como Sociedade, sendo de interesse público e privado.

A organização tem o dever de cumprir a legislação, na qualidade de Agentes de Tratamento de Dados conceituados na LGPD, posto ser um dever e uma obrigação instituída por lei.

O dever de cumprir, de diligência e cuidado na coleta, tratamento e armazenamento (incluindo transferência) dos dados pessoais, segue a manifestação de Facchini Neto e Soares:

O art. 6.º. Da LGPD indica os princípios a serem observados em qualquer atividade de tratamento de dados pessoais, indicando, dentre eles, o da

responsabilização e prestação de contas (inc. X): “demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, de eficácia dessas medidas” (g.n.). Tal redação parece apontar para a ideia de que não é suficiente a adoção de medidas abstratamente capazes de demonstrar o emprego de tais normas protetivas, mas é necessário que tais medidas sejam concretamente eficazes, pois caso contrário não faria sentido a repetição do substantivo “eficácia” após ter usado o adjetivo “eficaz”. Ou seja, tratar-se-ia aparentemente de uma obrigação de resultado.<sup>172</sup>

Dessa maneira, e compactuando no sentido de mitigar danos, riscos e outros questionamentos advindos da aplicação da norma protetiva do direito fundamental à proteção de dados, somente poderemos pensar em meios de explicar como os dados se relacionam com as decisões automatizadas quando temos gestão dos dados, ou seja, governança de dados.

Os dados pessoais, e sua proteção – constituem elementos centrais, e são na sociedade da informação elevados a categoria de direito fundamental nas palavras de Doneda:

Os dados pessoais acabam por identificar ou mesmo representar a pessoa em uma série de circunstâncias, nas quais a sua presença física não é possível ou conveniente. São elementos centrais, portanto, da proteção da personalidade e da construção da identidade em nossa sociedade. O tratamento de dados pessoais, em particular por processos automatizados, é ao mesmo tempo, uma atividade que apresenta riscos cada vez mais claros. Risco que se concretiza na possibilidade de exposição e utilização indevida ou abusiva de dados pessoais; na eventualidade de esses dados não serem corretos e representarem erroneamente seu titular; na sua utilização por terceiros sem o conhecimento ou autorização de seu titular; na eventualidade de serem utilizados para fins discriminatórios, somente para citar algumas hipóteses concretas. Saí a necessidade de mecanismos que possibilitem à pessoa deter conhecimento e controle sobre seus próprios dados – que são, no fundo, expressão direta de sua própria personalidade.<sup>173</sup>

Aplicar as suas atividades institucionais a conformidade e consonância com normas cogentes como a Lei Geral de Proteção de Dados (Lei n.º 13.709/2018) e outras regulações da Sociedade da Informação para proporcionar segurança digital, mecanismos de proteção, meios de resposta de incidentes eficiente e, ainda, documentos e protocolos de respostas para terceiros, explicando como a atividade organizacional propõe medidas transparentes e responsivas para suas relações digitais.

<sup>172</sup> FACCHINI NETO, Eugênio; SOARES, Flaviana Rampazzo. **Responsabilidade Civil pela Violação ao Dever de Proteção de Dados na LGPD**. In: PINHO, Anna Carolina *et al.* **Discussões sobre direito na era digital**. 1.ª ed. Rio de Janeiro: GZ, 2021, p. 251.

<sup>173</sup> DONEDA, Danilo Cesar Maganhoto. O direito fundamental à proteção de dados pessoais. In: SOUZA, Allan Rocha de *et al.* **Direito Digital: Direito Privado e Internet**. 4.ª ed. Indaiatuba: Foco, 2021, p. 33.

A Lei n.º 13.709/2018, Marco Geral de Proteção de Dados Pessoais estabelece que controladores e operadores (agentes de tratamento de dados) devem implementar políticas de acesso, uso, segurança e armazenamento dos dados coletados (estabelece, inclusive, critérios para a coleta, evitando abusos e exageros).

A história da Lei de Proteção de Dados brasileira, e a observância da boa-fé na coleta, tratamento e armazenamento são citadas abaixo, apontam Momo e Bone:

A LGPD, que entrou em vigor em agosto de 2020, surge como complemento ao Marco Civil da Internet e incluiu o Brasil no mapa dos países que proporcionam um nível adequado de proteção de dados pessoais. A nova Lei de Proteção de Ados Pessoais, inspirada no modelo regulatório europeu da General Data Protection Regulation, dispõe sobre o uso, tratamento e armazenamento de dados pessoais, inclusive nos meios digitais, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural (art. 1º.), estimulando práticas transparentes e seguras na utilização de dados pessoais. Dentre os dispositivos, a lei brasileira estabelece que as atividades de tratamento de dados pessoais deverão observar a boa-fé e serão submetidas, por exemplo a princípios relativos à finalidade, adequação e necessidade, garantido a realização do tratamento dos dados para propósitos legítimos, específicos e previamente informados ao titular, sem a possibilidade de tratamento de forma incompatível com essas finalidades (art. 6º, I) ou com as finalidades informadas ao titular (art. 6º, II), devendo-se limitar, esse tratamento ao mínimo necessário para a realização das finalidades pretendidas (art. 6º, III). De igual forma, assim como garante aos titulares dos dados pessoais o livre acesso e o direito a correção ou atualização de seus dados (art. 18º., incisos II e III), permite-se revogar o consentimento prestado (art. 8º., parágrafo 5º e 18º, IX) eliminar os dados armazenados em bancos de dados (arts. 5º, XIV e 18º, incisos IV e VI), bem como solicitar a portabilidade de seus dados pessoais para outro responsável (art. 18º, V).<sup>174</sup>

É possível compreender pela pesquisa como a Lei de Proteção de Dados proporciona a segurança jurídica necessária para que os dados sejam coletados, tratados e armazenados da melhor maneira possível, e dessa forma seja utilizado em inputs para que Inteligências Artificiais treinem, aprendam e decidam de forma mais transparente e ética, evitando assim vieses e problemas de decisão proferida por máquina não afrontem Direitos Fundamentais e de personalidade dos indivíduos.

---

<sup>174</sup> MOMO, Maria Vitória Galvan; BONE, Leonardo Castro de. A Proteção de Dados Pessoais Enquanto Direito Fundamental: um Olhar à Luz do Direito Comparado. *In*: COMÉRIO, Murilo Siqueira; JUNQUILHO, Tainá Aguiar. **Direito e Tecnologia: Um debate multidisciplinar**. Rio de Janeiro: Lumen Iuris, 2021, p. 39-40.

Frazão e Mulholland explicam que: “Do ponto de vista econômico, dados importam na medida em que podem ser convertidos em informações necessárias ou úteis para a atividade econômica.”<sup>175</sup>

Ou seja, ao inserir dados para treinamento e aprendizados de máquinas que irão decidir nas mais variadas formas, lugares, modos e tempo na vida das organizações e indivíduos, os dados (pessoais e corporativos) tornaram-se não só importantes, mas fundamentais por se tratar do principal ativo (ou combustível) para a Inteligência Artificial.

A lição de Frazão<sup>176</sup> continua sua análise sustentando e citando o pensamento de Frank Pasquale<sup>177</sup>:

Como bem descreve Frank Pasquale, os dados pessoais têm sido utilizados por governos e grandes players econômicos para a criação do que chama de one-way mirror, possibilitando que tais agentes saibam tudo sobre cidadãos, enquanto estes nada sabem dos primeiros. E tudo isso acontece por meio de um monitoramento e vigília constantes sobre cada passo da vida das pessoas, o que leva a um verdadeiro capitalismo de vigilância, cuja principal consequência é a constituição de uma sociedade também de vigilância.

Não somente os grandes players coletam e utilizam dados pessoais para tomada de decisão automatizada diária, com fins econômicos, tornando a Governança de Dados e a Ciência de Dados fundamental para tornar a qualidade, e a possível explicação clara e transparente para os indivíduos, pois a precisão das informações influencia na decisão automatizada das Inteligências Artificiais, porque:

Se os dados são insumos e os *inputs* da economia digital, os algoritmos são os instrumentos por meio dos quais os dados são processados e podem ser revertidos em resultados (*outputs*) a serem utilizados para as mais diversas finalidades. Muito além de aperfeiçoar estratégias econômicas já existentes, como seriam os casos de marketing personalizado (target marketing) e das classificações ou perfilizações (profiling), tais aplicações podem levar à total modificação do cenário econômico, político e social. Com efeito, entre as principais tarefas da chamada Data Science, estruturada a partir dos algoritmos, estão (i) a agregação ou segmentação de informações (clustering ou segmentation); (ii) a identificação de fraudes ou anomalias; (iii) a busca por associações e complementariedades (association-rule mining e cross-selling; e (iv) as predições. A partir dessas funcionalidades, os algoritmos estão hoje sendo programados para a extração de padrões e inferências a partir dos quais serão tomadas, de forma automatizada, decisões sobre questões objetivas, mas que estão atreladas a importantes dados sensíveis, assim como

<sup>175</sup> FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito – Ética, Regulação e Responsabilidade**. São Paulo: Revista dos Tribunais, 2019, p. 26.

<sup>176</sup> *Ibidem*, p. 27.

<sup>177</sup> PASQUALE, Frank. **The Black Box Society. The Secret Algorithms That Control Money and Information**. Cambridge-MA: Harvard University Press, 2015, p. 9.

decisões sobre questões subjetivas e que envolvem complexos juízos de valor, tais como. (i) avaliar as características, a personalidade, as inclinações e as propensões de uma pessoa, inclusive no que diz respeito à sua orientação sexual; (ii) analisar o estado de ânimo ou de atenção de uma pessoa; (iii) identificar estados emocionais, pensamentos, intenções e mesmo mentiras; (iv) detectar a capacidade e a habilidade para determinados empregos ou funções; (v) analisar a propensão a criminalidade; (vi) antever sinais de doenças, inclusive depressão, episódios de mania e outros distúrbios, mesmo antes da manifestação de qualquer sintoma.<sup>178</sup>

As decisões que tomamos, da forma como tomamos, e que foram objeto dos capítulos antecedentes, repercutem em forma de informações (dados) para Inteligência Artificial, e se não coletadas conforme a Lei Geral de Proteção de Dados pode causar vieses, discriminações, ou questões éticas das decisões automatizadas, e que podem prejudicar a explicação decisória da máquina utilizada por parte das organizações.

Frazão expõe sua concordância com o parágrafo acima da pesquisa, veja:

Ora, para que se pudesse ter o mínimo de confiança e tranquilidade em relação a tais processos, seria necessário haver algum tipo de controle tanto sobre (i) a qualidade dos dados, a fim de se saber se atendem aos requisitos da veracidade, exatidão, precisão, acurácia e sobretudo adequação e pertinência diante dos fins que justificam a sua utilização quanto sobre (ii) a qualidade do processamento de dados, a fim de saber se, mesmo a partir de dados de qualidade, a programação utilizada para seu tratamento é idônea para assegurar resultados confiáveis. Basta a descrição do problema para se observar que são muitas as variáveis sobre as quais se precisará ter um mínimo de supervisão e *accountability* para assegurar que *outputs* sejam fidedignos e confiáveis.<sup>179</sup>

A pesquisa aponta que o framework de proteção de dados pessoais pretende tornar mais ético, confiável e transparente a coleta e tratamento dos dados que são a base da decisão automatizada.

O cumprimento da norma enunciada na Lei n.º 13.709/2018 – por meio do Framework de Implementação, manutenção e atualização – é quem garantirá (em diálogo com normas regulatórias específicas de cada organização, setor ou modelo de negócio) a qualidade dos dados e a tutela da privacidade (como Direito de Personalidade) para inserção na IA, e que poderão tornar as decisões automatizadas mais transparentes, éticas e responsáveis.

---

<sup>178</sup> FRAZÃO, Ana. **Fundamentos da proteção de dados pessoais – Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados**. In: FRAZÃO, Ana. TEPEDINO, Gustavo; OLIVA, Milena Donato. **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. 1.ª ed. São Paulo: Thomson Reuters Brasil, 2019, p. 26-38.

<sup>179</sup> *Ibidem*, p. 38.

Importa esclarecer que não é somente implementar de acordo com a Lei Geral de Proteção de Dados, o *Compliance* Digital possuem um aspecto mais abrangente, ou seja, atua para proteção de dados corporativos, segredos estratégicos, dados financeiros, proteção de informações e utilização de tecnologias de maneira mais segura, não somente com políticas escritas, mas também com treinamentos constantes, monitoramento, análise e cumprimento de contratos de licenciamento de software (e suas integrações com Inteligência Artificial a depender do caso), segurança da informação, treinamentos constantes, monitoramento e atualização do quanto foi implementado, comitê de privacidade. E também com a verificação das normas do Direito Digital aplicáveis, bem como para todo o ecossistema (stakeholders) institucional, com o objetivo de conferir maior confiabilidade e um sistema de respostas em caso de incidentes ou solicitação de Explicação.

O *Compliance* Digital deve ser aplicado por desenho institucional (by default e by design) e poderá ser medida positiva para concretização do Direito à Explicação, vez que instituições que possuem a implementação com as diretrizes acima indicadas, que não são taxativas e não pretendem esgotar o tema, poderão contribuir para o desenvolvimento, o aprendizado de máquina e possibilitarão, portanto, conferir maior assertividade, segurança, transparência, confiabilidade e robustez dos sistemas de Inteligência Artificial.

De forma preventiva, e em caso de provocação por titular de dados quanto ao exercício da faculdade em exercer o Direito à Explicação, existirão documentos e mecanismos de aferição e resposta para explicar a decisão automatizada.

No sentido de implementação do programa de *Compliance* by default e by design – no caso específico para proteção de dados, Frazão e Cueva<sup>180</sup> sustentam suas reflexões citando Lawrence Lessig:

Nesse sentido, Lawrence Lessig sai destaca, há bastante tempo, a importância da tecnologia como importante regulador de comportamentos, de forma que, ao lado do controle hierárquico (lei), do controle baseado na competição (mercado) e do controle baseado na comunidade (normas sociais), haveria também o controle baseado no design ou na arquitetura (tecnologia). É importante advertir que, ao assim propugnar, o autor não adere a uma concepção estritamente determinista, até por entender que a arquitetura da rede não é definida apenas por questões técnicas, mas sobretudo políticas e valorativas. [...] É diante desse contexto que crescem as perspectivas da implementação do que se chama *privacy by design* e *by default*.<sup>181</sup>

---

<sup>180</sup> FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas Cueva. *Compliance e Políticas de Proteção de Dados*. São Paulo: Revista dos Tribunais, 2021, p. 47-50.

<sup>181</sup> LESSIG, Lawrence. *Code and Other Laws of Cyberspace*. New York: Basic Books, 1999.



O quanto descrito acima vem de encontro com o escopo supra sugerido, que proponho para conferir maior robustez e confiança no uso das tecnologias de Inteligência Artificial. Corroborando com a pesquisa Palhares, Prado e Vidigal:

Muito embora a Lei Geral de Proteção de Dados Pessoais seja o assunto da vez quando se fala em *compliance* digital, seguido de aspectos relacionados ao Marco Civil da Internet e à regulação do comércio eletrônico, as mudanças pelas quais o Direito passa inevitavelmente trarão outras preocupações para organizações de todos os portes quando se fala em conformidade com normas.<sup>182</sup>

A citação acima reforça a pesquisa no sentido de afirmar o *Compliance* Digital não somente nos termos da proteção de dados, mas como cumprimento de normas tecnológicas, normais sociais, culturais, éticas para conferir legitimidade ao Direito de Explicar as decisões automatizadas, mesmo sendo a privacidade e a proteção de dados são fundamentais para o titular de dados (usuário).

Colabora para a pesquisa a Declaração de Osaka dos Líderes do G20, que indica princípios e diretrizes para sistemas de Inteligência Artificial que coadunam com o que foi explanado em termos de *Compliance* Digital, item 12:

12. A fim de promover ainda mais a inovação na economia digital, apoiamos o intercâmbio de boas práticas sobre políticas eficientes e enfoques e marcos regulatórios que sejam inovadores, ágeis, flexíveis e adaptadas à era digital, inclusive por meio de regras experimentais (“regulatory sandboxes”). O desenvolvimento e o uso responsável da Inteligência Artificial (IA) podem ser uma força motora para ajudar a promover os ODS e a criar uma sociedade sustentável e inclusiva. Para promover a confiança do público em tecnologias de IA e realizar plenamente seu potencial, comprometemo-nos com um enfoque de IA centrado no ser humano e saudamos os Princípios sobre IA do G20, extraídos da Recomendação da OCDE sobre IA. Adicionalmente, reconhecemos a crescente importância de promover a segurança na economia digital e tratar de lacunas e vulnerabilidades. Afirmamos a importância da proteção da propriedade intelectual. Juntamente com a rápida expansão de tecnologias emergentes, inclusive a “Internet das Coisas (IoT, em inglês), cresce o valor de uma permanente discussão sobre a segurança na economia digital. Nós, como membros do G20, afirmamos a necessidade de continuar tratando desses desafios urgentes. Reafirmamos a importância de superar o hiato digital e promover a adoção da digitalização pelas pequenas e médias empresas (PMEs) e por todos os indivíduos, particularmente os grupos

---

<sup>182</sup> PALHARES, Felipe; PRADO, Luis Fernando; VIDIGAL, Paulo. *Compliance Digital e LGPD*. São Paulo: Thomson Reuters, 2021, p. 330.

vulneráveis, e também encorajamos a comunicação e a troca de experiências entre cidades, para o desenvolvimento de cidades inteligentes.<sup>183</sup>

A privacidade e a proteção de dados pessoais, claro, é fundamental para conferir transparência, não maleficência, prestação de contas (*accountability*) e ética aos sistemas de Inteligência Artificial, como adverte Doneda:

A privacidade deve ser vista não apenas como um direito individual, mas como um direito fundamental que assegura a liberdade e a autonomia do indivíduo na sociedade. A proteção de dados pessoais requer um arcabouço normativo robusto que promova os princípios da finalidade, necessidade e transparência, além de responsabilização e prestação de contas por parte dos controladores de dados.<sup>184</sup>

Os dados (pessoais e corporativos) são parte substancial da existência de um sistema de Inteligência Artificial, sendo aliada à Governança de Dados, cujo objetivo é promover a intersecção da proteção à privacidade e da proteção de dados como uma das formas possíveis de mitigação de riscos no uso dos sistemas de Inteligência Artificial, e como uma das ferramentas que possibilitam explicar as decisões automatizadas que serão tratadas mais adiante.

A Lei Geral de Proteção de Dados (Lei: regulou conceitos como dados pessoais, dados sensíveis, agentes de tratamento, encarregado de dados, dentre outras questões anteriormente não enunciadas em normas que com ela convergem ou dialogam, como no caso do Direito Fundamental à Proteção de Dados estabelecido na Constituição Federal do Brasil, no Código de Defesa do Consumidor, bem como Lei de Sigilo Bancário e Sigilo Tributário.

“Um dos problemas mais comuns quando se preparar uma base de dados para análise é a presença de valores faltantes, ou *missing values*.”<sup>185</sup>, sendo possível mitigar essa fragilidade que provoca problemas no aprendizado da Inteligência Artificial, e, portanto, na Decisão Automatizada, se os processos de coleta foram adequados as finalidades a que se destinam o cumprimento legal ou a execução contratual por exemplo.

Desse modo, ao cumprir normas para o *compliance* digital, é possível – evitar ou corrigir tais valores faltantes.

---

<sup>183</sup> BRASIL. Ministério das Relações Exteriores. **Declaração de Osaka dos Líderes do G20**, 7 jan. 2021. Disponível em: [https://www.gov.br/mre/pt-br/canais\\_atendimento/imprensa/notas-a-imprensa/2019/declaracao-de-osaka-dos-lideres-do-g20](https://www.gov.br/mre/pt-br/canais_atendimento/imprensa/notas-a-imprensa/2019/declaracao-de-osaka-dos-lideres-do-g20). Acesso em: 10 nov. 2024.

<sup>184</sup> DONEDA, Danilo Cesar Maganhoto. **Da Privacidade à Proteção de Dados Pessoais: elementos da formação da nova cultura jurídica**. São Paulo: Revista dos Tribunais, 2017, p. 35.

<sup>185</sup> SICSU, Abraham Laredo; SAMARTINI, André; BARTH, Nelson Lerner. **Técnicas de Machine Learning**. São Paulo: Blucher, 2023, p. 69.

A máquinas aprendem com a inserção de diversas informações que lhe são disponibilizadas – pessoais e institucionais – sendo necessárias “primordialmente, a organização dos dados, que significa o uso para estruturação; o auxílio a tomada de decisão, que indica o uso de processamento e a automação da decisão, que demonstra a decisão em si”.<sup>186</sup>

O *framework* de Governança de Dados e *Compliance* Digital permite averiguar, no campo da coleta, tratamento e armazenamento, estrutura de coleta, tratamento e armazenagem dos dados, bem como sua estruturação necessita do desenho institucional (*by design*) e do padrão mínimo (*by default*); comprovação de aplicação de Boas Práticas para prestação de contas (*Accountability*); Segurança da Informação; Transparência e Explicabilidade; Não Discriminação, Supervisão Humana, Responsabilidade como elencado pelos professores Felipe Chiarello e Lara Rocha Garcia no artigo intitulado “Desafios internacionais da aplicação da inteligência artificial no Direito” de 2021.

A denominada Governança de Dados, também conhecida como Framework de Proteção de Dados, não protege somente os Dados Pessoais, como é de se pensar, mas protege e melhora os processos de gestão e tratamento de Dados Corporativos, uma vez que não delimita o Framework apenas para Dados Pessoais, sendo a Governança de Dados uma implementação de múltiplos fatores protegem os Dados Corporativos conjuntamente.

A Segurança da Informação é parte integrante da Governança de Dados, como não poderia deixar de ser, uma vez que a segurança é parte integrante da gestão e da governabilidade, fazendo parte dos princípios dispostos na Lei Geral de Proteção de Dados e do Regulamento Europeu de Proteção de Dados.

A Segurança da Informações como parte integrante da Governança dos Dados propõe cinco macrofases, a saber: “(i) planejar; (ii) avaliar; (iii) executar; (iv) apresentar resultados e (v) revalidar”.<sup>187</sup>

Organizações coletam, tratam, compreendem e utilizam dados para embasar tomada de decisões (humanas de seus gestores e automatizadas quando são imputadas em sistemas computacionais inteligentes), denominando-se tais análises de Inteligência de Negócios, também comumente nominada de *Business Intelligence*. Dessa forma, necessitam da implementação cuidadosa e correta da Governança de Dados e da Segurança, como expõe

---

<sup>186</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Desafios Internacionais da Aplicação da Inteligência Artificial no Direito. **Revista Justiça de Direito**, v. 35, n. 2, p. 10, mai./ago. 2021. Disponível em: <https://seer.upf.br/index.php/rjd/article/view/13040/114116181>. Acesso em: 24 abr. 2024.

<sup>187</sup> MANOEL, Sérgio Silva. **Governança de Segurança da Informação: como criar oportunidades para o seu negócio**. 1.ª ed. Rio de Janeiro: Brasport, 2014, p. 75.

Sharda e Ramesh<sup>188</sup> ao descreverem a importância da segurança e proteção de privacidade. Esta é uma questão extremamente importante no desenvolvimento de um sistema computadorizado, sobretudo em BI que contém dados que podem possuir valor estratégico. Além disso, a privacidade de funcionários e clientes precisa ser protegida.

A proteção dos dados pessoais deve ser fundamentada na legislação devendo o agente de tratamento observar as bases legais, cujas hipóteses de tratamento são limitadas e taxativas, e dependem de consentimento.

Não se lê da lista de exceções a coleta e tratamento de dados para utilização em sistemas de Inteligência Artificial, portanto, devendo ser os dados pessoais objeto de proteção e governança pelas organizações como a pesquisa observou.

Dados inconsistentes que estejam mal alocados e que, portanto, são mal imputados em sistemas de Inteligência Artificial darão resultados equivocados ou alterados, ou seja, dados não válidos e que são objeto de inputs em sistemas inteligentes certamente proferirão decisões automatizadas que não terão condições de validação e verificação, prejudicando o Direito à Explicação.

O relatório de Impacto à proteção produzido e disponibilizado após implementação do framework de governança de dados se faz necessário a proteção do agente institucional envolvido na utilização de sistemas de inteligência artificial, sendo um dos documentos que compõe a manutenção e possibilitam a atualização da governança, e tem o escopo de documentar a gestão do tratamento de dados, permitindo assim fazer parte dos documentos que compõe o manejo para resposta do Direito de Explicação, bem como conexão com o tratamento e armazenamento de dados que são inseridos para aprendizado de máquina e decisões automatizadas.

Como a pesquisa demonstrou, o framework de proteção de dados e o relatório de impacto à proteção de dados da qual é parte, é fundamental para as decisões automatizadas e formação de perfis que serão utilizados para ensinar Inteligência Artificial.

Quinelato reforça a pesquisa ao afirmar que:

Assim, situações de risco em que possa ocorrer algum risco ou possível dano ao titular, recomenda-se a elaboração, redação, manutenção e armazenamento do relatório. Portanto, é consenso a recomendação de sua realização quando existe tratamento de dados para formação de perfis e uma avaliação sistemática e aprofundada de dados pessoais com base em sistema automatizado, cujas decisões que produzem efeitos podem afetar o titular.

---

<sup>188</sup> SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio**. Tradução: Ronald Saraiva de Menezes. Revisão: Ângela Brodbeck. 4.<sup>a</sup> ed. Porto Alegre: Bookman, 2019, p. 23.

Dessa forma, entende-se que a plataforma digital que. Realizar tratamento de dados para precificação personalizada deve elaborar um relatório de impacto.<sup>189</sup>

Isto reforça o entendimento que a Governança dos Dados e do *Compliance* Digital no fluxo de utilidade e ciclo de vida da Inteligência Artificial, para que os sistemas de aprendizado de máquina possam ter dados imputados em conformidade com as normas, e para exercício do Direito à Explicação das decisões automatizadas e da confiabilidade dos sistemas de Inteligência Artificial.

### 3.2 MECANISMOS POSITIVOS

A transparência, confiança e robustez da Inteligência Artificial é tema importante para essa pesquisa, pois precisamos de padrões rigorosos de transparência e *accountability* para sistemas automatizados de tomada de decisão para garantir menor potencial de risco possível.

A pesquisa analisou regulação *ex ante* como medidas protetivas, boas práticas e ações afirmativas para desenvolvimento e utilização da Inteligência Artificial.

De acordo com Floridi<sup>190</sup>, a mitigação de viés em sistemas de IA requer políticas que favoreçam a equidade e a inclusão desde o início do desenvolvimento dos algoritmos.

O ciclo de vida da inteligência artificial (IA) é o conjunto de etapas e processos que envolvem a criação, implementação, monitoramento e avaliação de sistemas de IA. O ciclo de vida da IA pode variar de acordo com o tipo, a complexidade e o propósito do sistema, mas existem alguns elementos comuns que devem ser considerados em qualquer projeto de IA.

A Governança da Inteligência Artificial tem como função, pressuposto criar um desenvolvimento amparado por diretrizes e elementos. Um desses elementos é a transparência. Sua concepção – como um dos maiores motivadores de implementação de princípios e documentos de governança da IA – em relações as suas decisões.

A pesquisa traz opinião de Santaella:

O conceito de transparência, que também faz parte integrante do dicionário da IA, é majoritariamente utilizado no contexto da política e do governo, mas é também utilizado pela ciência, na engenharia, nos negócios, inclusive na humanidade. Compreende-se por transparência a abertura e acesso sem entraves a informação. Por isso, comparece nos estudos sobre conflitos e

<sup>189</sup> QUINELATO, Pietra Daneluzzi. **Preços Personalizados à luz da Lei Geral de Proteção de Dados**. 1.<sup>a</sup> ed. São Paulo: Foco, 2022, p. 163.

<sup>190</sup> FLORIDI, Luciano. Addressing Bias in Artificial Intelligence: A Holistic Ethical Framework. **Ethics and Information Technology**, v. 23, n. 1, p. 15-27, 2021.

organizações internacionais, política monetária, comércio, corrupção, teoria democrática, gestão pública, contabilidade e finanças etc. Nesses contextos, o conceito atingiu uma significância “quase religiosa”. Ademais, a transparência é apontada como condição fundamental para a efetivação da responsabilidade no discurso público sobre a governança.<sup>191</sup>

Os *frameworks* de inteligência artificial da OCDE e do NIST fornecem orientações e recomendações para a gestão do ciclo de vida da IA, baseadas em valores e princípios comuns, tais como: benefício humano, inovação, confiança, equidade, explicabilidade, responsabilidade, privacidade, segurança e sustentabilidade.

Esses *frameworks* visam promover o desenvolvimento e o uso de sistemas de IA que sejam alinhados com os valores democráticos, que respeitem os direitos humanos e que contribuam para o bem-estar social e econômico. A expansão acelerada da Inteligência Artificial (IA) trouxe consigo a necessidade imperativa de estabelecer estruturas de governança robustas para garantir seu desenvolvimento e implementação éticos.

Os frameworks desenvolvidos por organizações como NIST<sup>192</sup>, IBM<sup>193</sup> e OCDE<sup>194</sup> emergem como pilares fundamentais para a pesquisa e como exemplos de regulação contemporânea para Inteligência Artificial. O cenário contemporâneo da IA apresenta desafios significativos relacionados a vieses algorítmicos, discriminação automatizada e questões éticas. O National Institute of Standards and Technology (NIST) desenvolveu um framework que enfatiza: (i) Transparência algorítmica; (ii) *Accountability*; (iii) Gerenciamento de riscos. E (iv) Validação de modelos

O AI Act Europeu<sup>195</sup> primeira norma legal abrangente tem como objetivo promover uma IA confiável na Europa e também garantir que os sistemas de IA respeitem os direitos fundamentais, a segurança e os princípios éticos e abordando os riscos de modelos de IA muito poderosos e impactantes. propõe: (i) Classificação de risco para sistemas de IA; (ii) Requisitos obrigatórios para IA de alto risco; (iii) Proibições específicas para práticas prejudiciais.

<sup>191</sup> SANTAELLA, Lúcia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis Santiago Guerra Filho *et al.* **Direito e Inteligência Artificial: Fundamentos – Volume 1 - Inteligência Artificial, Ética e Direito**. Rio de Janeiro: Lumen Juris, 2021, p. 127-128.

<sup>192</sup> UNITED STATES. Department of Commerce. **AI Risk Management Framework**. Gaithersburg, MD: National Institute of Standards and Technology, 26 jul. 2024. Disponível em: <https://www.nist.gov/itl/ai-risk-management-framework>. Acesso em: 7 dez. 2024.

<sup>193</sup> IBM. **Introduction to the AI Framework for z/OS**, 27 feb. 2025. Disponível em: <https://www.ibm.com/docs/en/zos/3.1.0?topic=zos-introduction-ai-framework>. Acesso em: 13 dez. 2024.

<sup>194</sup> ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). OECD Framework for the Classification of AI systems. **OECD Digital Economy Papers**, 22 feb. 2022. Available at: [https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems\\_cb6d9eca-en.html](https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html). Access in: 13 dec. 2024.

<sup>195</sup> EUROPEAN ARTIFICIAL INTELLIGENCE ACT. The AI Act Explorer. **Official Journal**, 13 jun. 2024. Available at: <https://artificialintelligenceact.eu/ai-act-explorer/>. Access in: 13 dec. 2024.

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) estabelece: (i) Crescimento inclusivo; (ii) Desenvolvimento sustentável e (iii) Bem-estar social; A implementação efetiva destes frameworks requer: (i) Avaliação de riscos; (ii) Estabelecimento de políticas; (iii) Treinamento de equipes; (iii) Monitoramento contínuo.

A adoção destes frameworks tem demonstrado resultados positivos na mitigação de problemas éticos e vieses, bem como e certamente, serão meios para cumprimento e viabilidade de obtermos meios para responder aos usuários como um sistema de Inteligência Artificial teve uma base de dados de qualidade disponível, quais foram as metodologias de inserção (inputs) desses dados em seu desenvolvimento e em sua utilização, como o modelo proferido pelo algoritmo, e da rede neural foram desenvolvidas, monitoradas, atualizadas e documentadas para atender ao Direito à Explicação.

A governança e a regulação (interna e externa) são mitigadores (e ferramentas) de averiguação sobre o uso dos dados e do respeito aos Direitos Fundamentais para as tomadas e meios de decisão automatizada em Inteligência Artificial. As decisões de Inteligência Artificial poderão, ainda, pelo uso indevido de informações dos indivíduos e não cumprimento de normas de Direitos Fundamentais, afetar a sociedade e os indivíduos, sendo utilizadas para abuso de Poder Econômico e concentração de Mercado por determinados agentes que detêm essa tecnologia.

Adicionada à Governança de Dados, cujo objetivo é coletar, tratar e armazenar dados com qualidade não somente para a privacidade e da proteção de dados, é claramente um instrumento de mitigação de riscos de emissão de decisões automatizadas prejudiciais, de difícil ou impossível Explicação quando solicitado.

Estruturalmente, algoritmos têm nos dados e informações inseridas seu principal alicerce, pois tais sistemas dependem de dados para aprendizado e proferimento de resultados.

A questão é como os algoritmos, fechados em suas estruturas e definições, podem limitar oportunidades, comprometer a liberdade individual e colocar em risco a proteção de dados pessoais.<sup>196</sup>

A Governança dos Dados e da Inteligência Artificial não se confundem com Governança Corporativa, sendo esta última um modelo de gestão (formalmente estruturado ou não) que as empresas possuem como formatação de suas lideranças e técnicas de liderar, administrar, monitorar o fluxo de processos e pessoas, bem como incentivar o cumprimento de

---

<sup>196</sup> ABRUSIO, Juliana. O princípio da Inteligência Artificial Explicável e a Proteção de Dados. In: RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 169.

seus objetivos de negócios para o propósito de constituir valores (incluindo-se os financeiros), cultura, regras (tácitas ou formais) e princípios da administração da empresa.

O papel da Governança da Inteligência Artificial é mais específico, constituído de metodologias de implementação, atualização, monitoramento e evolução da Inteligência Artificial.

A pesquisa pretende apresentar, sem esgotamento do tema, propostas emitidas como sugestão de Governança da Inteligência Artificial visando proteger os seres humanos, o mercado, o Estado e a Sociedade, mantendo dentro da empresa, até que seja solicitada em eventual fiscalização ou para compor Direito à Explicação de uma decisão automatizada, como sugere Frank Pasquale: “Pode manter suas análises em sigilo, compartilhá-las com outros reguladores ou torná-las públicas, dentro de certos limites legais e constitucionais.”<sup>197</sup>

A Comissão Européia emitiu o *Ethics Guideline for Trustworthy Artificial Intelligence* e o *The Assessment List for Trustworthy Artificial Intelligence* (A.L.T.A.I., na sigla em inglês), a Organização para Cooperação e Desenvolvimento Econômico (O.E.C.D, na sigla em inglês) publicou o *OECD Framework for the Classification of Artificial Intelligence Systems* e o *Recommendation of the Council on Artificial Intelligence*, pela *National Institute of Standards and Technology* (N.I.S.T, na sigla em inglês), na Inglaterra a *Information Commissioner's Office* (I.C.O., na sigla em Inglês) expediu o *Guidance on The A.I. Auditing Framework*, ISO/IEC 42001.<sup>198</sup>

A ISO/IEC 42001 é voltada para a governança e gestão de sistemas de IA fornecendo uma estrutura para garantir que o desenvolvimento, a implantação e a manutenção de IA sejam conduzidos de forma segura, ética e alinhada aos objetivos estratégicos da organização.

Os *frameworks* para Governança de Inteligência Artificial mencionados não esgotam os documentos e sugestões que enunciam normas para melhorar a aplicabilidade dos sistemas e formas de Inteligência Artificial. A pesquisa aponta também ferramentas como o *Ethics & Algorithms Toolkit*<sup>199</sup>, como instrumento de ética e ferramentas para IA responsável.

---

<sup>197</sup> Texto Original: “It can keep its analyses private, share them with other regulators, or open them to the public, within certain statutory and constitutional limits.” PASQUALE, Frank. **The Black Box Society. The Secret Algorithms That Control Money and Information.** Cambridge-MA: Harvard University Press, 2015, p. 169.

<sup>198</sup> ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 42001:2024 - Tecnologia da informação — Inteligência artificial — Sistema de gestão**, 18 abr. 2024. Disponível em: <https://www.abntcatalogo.com.br/pnm.aspx?Q=eXkvMUMzazlVVWZzRjN1eEJxQVpoOU05WkVaamF6ZDIBVnRIem5zSDBkRT0=>. Acesso em: 7 dez. 2024.

<sup>199</sup> ANDERSON, David *et al.* **Ethics & Algorithms Toolkit.** Available at: <https://ethicstoolkit.ai/>. Access in: 7 dez. 2024.



O Brasil ainda não possui um sistema ou documento que enuncie um procedimento de orientação para desenvolvimento, fornecimento, uso e aplicabilidade da Inteligência Artificial, uma vez que não temos ainda um marco legal específico.

O fato de não possuir uma lei geral de Inteligência Artificial não impede ou repele quaisquer aplicações dos documentos internacionais, que certamente podem ser utilizados como parâmetro e complemento de normas autoaplicáveis para implementação, monitoramento, atualização, de fatores éticos e de respostas das decisões e impactos da tecnologia.

Afinal, existem uma série de legislações que dialogam e compõe a sistemática de governança da *Artificial Intelligence*, como Lei Geral de Proteção de Dados, Código Civil, Código do Consumidor, Lei da Concorrência, Marco Civil da Internet, Constituição Federal do Brasil, Tratado de Direitos Humanos, normas de Segurança da Informação, etc, apenas como exemplos de enunciações que também compõe e interagem para a ética, universalização de princípios e fundamentos dos sistemas artificialmente inteligentes.

Expõe-se, ainda, que existem *frameworks* customizados, ou seja, desenhados de acordo com a cultura e o propósito do agente de mercado. Cito como exemplos os frameworks da empresa automobilística Rolls Royce<sup>200</sup> e da Mercedes Bens<sup>201</sup>.

Tais metodologias foram discutidas e propostas pelas organizações internacionais como forma de melhorar os resultados e decisões automatizadas em processos de gestão e produção, visam a precaução e prevenção para tornar os sistemas de Inteligência Artificial mais governáveis, possibilitando assim diminuição de riscos aos destinatários, sejam eles cidadãos, agentes de mercado, instituições ou empresas públicas, e sociedade.

Tais *frameworks* possibilitam maior robustez, transparência, prestação de contas (*accountability*), responsabilidade, interoperabilidade de dados, ética e não confronto com Direitos Fundamentais e Humanos no uso da Inteligência Artificial.

Isso comprova que, para maior segurança informacional, tecnológica e jurídica, faz-se necessário desenhar especificamente a Governança institucional, melhorando as respostas e suas explicações junto ao Estado, ao mercado e ao cidadão, e tais frameworks, no entendimento da pesquisa, indicam que são necessárias análises específicas para o design da governança da Inteligência Artificial, concluindo-se que as normas descritas nos frameworks gerais são

---

<sup>200</sup> ROLLS ROYCE. **O Marco Aletheia – Ajudando a construir confiança na inteligência artificial**. Disponível em: <https://www.rolls-royce.com/innovation/the-aletheia-framework.aspx>. Acesso em: 27 abr. 2024.

<sup>201</sup> MERCEDES-BENZ GROUP. **Como a Mercedes-Benz usa a Inteligência Artificial (IA) – Duas Letras e Quatro Princípios**. Disponível em: <https://group.mercedes-benz.com/responsibility/compliance/digital/ki-guidelines.html>. Acesso em: 28 abr. 2024.

fundamentais para a execução real e prática que servem de mitigação de riscos e danos para todo o ecossistema da sociedade.

É importante entender que a sociedade é complexa, e que os aspectos culturais e linguísticos devem ser levados em conta para a Governança específica da Inteligência Artificial em conexão com valores, princípios e fundamentos universais como Direitos Humanos e Fundamentais.

Um conhecido modelo de governança de IA é o *Business Software Alliance (BSA)*<sup>202</sup>, que congrega as maiores empresas do mundo, chamada de *Confronting Bias: BSA's framework to build in IA*.

A proposta de efetiva aplicação e implementação da Governança da Inteligência Artificial é importante, uma vez que não será possível dar respostas e explicações das decisões automatizadas e resultados proferidos por tal tecnologia por estar o Framework intimamente conectado ao entendimento de como desenvolver, fornecer, adquirir e utilizar tais sistemas tecnológicos nas mais variadas aplicações e usabilidades (incluindo certamente a interoperabilidade de dados), sem que se compreenda como a tecnologia relaciona-se interna e externamente em diversos contextos institucionais.

Ponto de encontro é a desconfiança nos sistemas e nas empresas que desenvolvem, utilizam os sistemas inteligentes, uma vez que se trata de confiar sobre de que forma e com que segurança são utilizadas as Inteligências Artificiais.

Confiança e reputação dos agentes de mercado e Estado somente poderão ser comprovados e explicados por meio do diferencial que a publicização, exposição quanto à aplicação e à execução de tais princípios, normas e regras de governar a Inteligência Artificial sejam efetivamente de conhecimento da sociedade, cujos fundamentos serão tratados a seguir.

Certamente a Governança de Inteligência Artificial apoiará a tomada de decisão da gestão de um determinado negócio ou instituição. O documento e orientação importante para a relação jurídica entre usuário é a Política de Inteligência Artificial e o Treinamento para utilização das ferramentas.

O comitê de ética para a governança de inteligência artificial é um órgão colegiado que tem por objetivo orientar, monitorar e avaliar as atividades de pesquisa, desenvolvimento, implementação e uso de sistemas de inteligência artificial de acordo com princípios, normas e regras éticas reconhecidas internacionalmente.

---

<sup>202</sup> BUSINESS SOFTWARE ALLIANCE. **Business Software Alliance Transition Priorities**. Available at: <https://www.bsa.org/pt>. Acesso em: 7 dec. 2024.

O comitê de ética deve seguir as recomendações da Organização para a Cooperação e o Desenvolvimento Econômico (OCDE)<sup>203</sup>, do Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (NIST)<sup>204</sup> e do Comissariado de Informação do Reino Unido (ICO-UK) sobre a governança responsável e transparente da inteligência artificial.

Algumas dessas recomendações constantes dos instrumentos acima são: (i) Respeitar os direitos humanos, a dignidade humana, a democracia, o estado de direito e a diversidade cultural; (ii) Promover a inclusão, a equidade, a justiça, a solidariedade e o bem-estar social; (iii) Garantir a segurança, a confiabilidade, a robustez, a precisão e a qualidade dos sistemas de inteligência artificial; (iv) Assegurar a explicabilidade, a interpretabilidade, a prestação de contas e a auditabilidade dos processos e resultados da inteligência artificial; (iv) Proteger a privacidade, a confidencialidade, a integridade e a segurança dos dados pessoais e não pessoais; (v) Prevenir e mitigar os riscos, os danos e os impactos negativos da inteligência artificial para as pessoas, a sociedade e o meio ambiente. (vi) Fomentar a inovação, a criatividade, a diversidade, a colaboração e o aprendizado contínuo na pesquisa e no desenvolvimento da inteligência artificial; (vii) Estimular o engajamento, a participação, a consulta e a deliberação de todas as partes interessadas na governança da inteligência artificial.

Outros exemplos de Governança de Inteligência Artificial são o proposto pelo Governo de Cingapura, o *Model AI Governance* desenvolvido pelo governo como uma iniciativa desenvolvida pela *Infocomm Media Development Authority* (IMDA) e pelo *Personal Data Protection Commission* (PDPC).<sup>205</sup>

A pesquisa verificou como uma das ações positivas de implementação *ex ante* o Comitê de ética, que deve adotar metodologias e ferramentas adequadas para realizar suas funções, também constantes dos documentos acima mencionados, tais como:

a) Definir um código de conduta e uma política de ética para a inteligência artificial, que estabeleçam os princípios, as normas e as regras que devem orientar as decisões e as ações dos envolvidos na pesquisa, no desenvolvimento, na implementação e no uso de sistemas de inteligência artificial.

---

<sup>203</sup> ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *Governing with Artificial Intelligence. OECD Artificial Intelligence Papers*, 13 jun. 2024. Available at: [https://www.oecd.org/en/publications/governing-with-artificial-intelligence\\_26324bc2-en.html](https://www.oecd.org/en/publications/governing-with-artificial-intelligence_26324bc2-en.html). Access in: 7 dec. 2024.

<sup>204</sup> UNITED STATES. Department of Commerce. *AI Risk Management Framework*. Gaithersburg, MD: National Institute of Standards and Technology, 26 jul. 2024. Disponível em: <https://www.nist.gov/itl/ai-risk-management-framework>. Acesso em: 7 dez. 2024.

<sup>205</sup> PERSONAL DATA PROTECTION COMMISSION. *Singapore's Approach to AI Governance*. Available at: <http://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework>. Access in: 7 dec. 2024.

b) Realizar uma avaliação de impacto algorítmico, que consiste em uma análise sistemática dos potenciais benefícios e riscos da inteligência artificial para as pessoas, a sociedade e o meio ambiente, bem como das medidas de prevenção, mitigação e remediação desses riscos.

c) Implementar uma auditoria de inteligência artificial, que consiste em um processo independente e periódico de verificação e validação do cumprimento dos requisitos éticos, legais e técnicos pelos sistemas de inteligência artificial, bem como da correção e melhoria das eventuais falhas ou deficiências encontradas.

d) Estabelecer canais de comunicação, informação, educação e sensibilização sobre a inteligência artificial, que visem aumentar o conhecimento, a compreensão, a confiança e a responsabilidade dos pesquisadores, desenvolvedores, usuários, consumidores e da sociedade em geral sobre a inteligência artificial, seus benefícios, desafios e limitações.

e) Criar mecanismos de escuta, feedback, reclamação e recurso sobre a inteligência artificial, que permitam receber, avaliar e responder às manifestações, às dúvidas, às críticas, às sugestões, às denúncias e aos pedidos de revisão ou reparação das partes interessadas ou afetadas pela inteligência artificial.

Portanto, atuar de forma transparente, independente, imparcial e responsável, respeitando os valores, os direitos e os deveres éticos, legais e profissionais envolvidos na governança da inteligência artificial.

A Governança Algorítmica é o conjunto proposto de boas práticas somadas à governança de dados, à segurança da informação e às atividades complementares aplicáveis na utilização dos sistemas de Inteligências Artificiais, podendo contribuir e facilitar a execução do Direito à Explicação das decisões judiciais automatizadas.

Sugere-se, também, o Relatório de Impacto Algorítmico, *Algorithmic Impact Assessment (AIA)*<sup>206</sup>, cujo conceito é o mesmo do Relatório de Impacto de Dados. A conformidade dos processos juntamente com as pessoas que compõe o Judiciário governando a inteligência artificial pautam um ecossistema de responsabilidade social e ambiental, cuja premissa é a transparência e a segurança (institucional, informacional e jurídica).

Sobre a avaliação (denominada entre alguns de relatório) de Impacto Algorítmico, Vainzof:

---

<sup>206</sup> ABRUSIO, Juliana. **Proteção de Dados na Cultura do Algoritmo**. 1.ª ed. São Paulo: D'Plácido, 2020, p. 334. (Coleção Direito privado, v. 1).

Uma das formas de *accountability* é avaliar impactos e mitigar riscos, por meio de instrumentos que geram evidências das diligências, como no. Caso do Relatório de Impacto de Proteção de Dados (RIPD), previsto na LGPD. Para aplicações de IA, o Artificial Intelligence Assessment (AIA) é metodologia que ajuda as organizações a mapear os benefícios de sistemas de IA, inclusive sociais, analisar sua confiabilidade, segurança e transparência, identificar seus valores e objetivos, entender e limitar seus riscos, e refletir se as opções foram feitas com ponderações éticas. Referida avaliação torna-se preponderante especialmente quando a IA possa afetar direitos e garantias fundamentais, a elegibilidade de um indivíduo e o acesso dele a serviços essenciais (saúde, educação, financeiro, energia, por exemplo) a admissão de uma pessoa em um país ou sua cidadania, a admissão de indivíduos em escolas, universidades e empregos, entre outras possibilidades. Ou seja, quanto maior o risco dos sistemas de IA, mais forte e sofisticada deve ser a implementação de medidas e controles não apenas para avaliar seus riscos, mas também definir, aplicar e verificar a eficácia dos controles e medidas de mitigação. De acordo com o uso da IA, as organizações conseguem sopesar os riscos de forma distinta e decidir sobre as salvaguardas entre os diferentes resultados, medidas mitigadoras e frequência de suas revisões. Assim, a Avaliação de Impacto de Inteligência Artificial, enquanto não regulada, torna-se uma boa prática empresarial para organizações que desejam usar I.A. e analisar suas consequências jurídicas e éticas, tanto na fase de design quanto durante o uso dos sistemas, oportunidade em que podem avaliar as consequências das aplicações.<sup>207</sup>

As preocupações e riscos que envolvem a IA se alteram, à medida em que este modelo de regulação não se dedica a criar regras extensivas e detalhadas, mas sim a mapear pontos de preocupação e geri-los conforme o nível de risco.

O Projeto de Lei n.º 2.338 de 2023<sup>208</sup> prevê em seu Capítulo III, abaixo transcrito, uma avaliação preliminar de risco, cuja aferição somente será possível mediante a aplicação de

---

<sup>207</sup> VAINZOF, Rony; SERAFINO, Danielle; STEINWASCHER, Aline. Aplicações e Regulação da Inteligência Artificial. In: VAINZOF, Rony *et al.* **Legal Innovation: o Futuro do Direito e o Direito do Futuro**. São Paulo: Thomson Reuters, 2022, p. 256-257.

<sup>208</sup> *In verbis*: “CAPÍTULO III - DA CATEGORIZAÇÃO DOS RISCOS. Seção I - Avaliação preliminar. Art. 12. Antes de sua introdução e circulação no mercado, emprego ou utilização, o agente de IA poderá realizar avaliação preliminar para determinar o grau de risco do sistema, baseando-se nos critérios previstos neste Capítulo, de acordo com o estado da arte e do desenvolvimento tecnológico. § 1º A realização da avaliação preliminar será considerada como medida de boa prática e poderá resultar em benefícios para o agente de IA para fins do disposto no art. 50, § 1º, podendo, inclusive, receber tratamento prioritário em procedimentos para avaliação de conformidade, nos termos do art. 34, desta lei. § 2º Caberá às autoridades setoriais definir as hipóteses em que a avaliação preliminar será simplificada ou dispensada, observadas as normas gerais da autoridade competente. § 3º O agente poderá requerer junto aos demais agentes dos sistemas de inteligência artificial informações que o capacitem a efetuar avaliação preliminar, nos termos da presente Lei, respeitados os segredos comercial e industrial. § 4º Garantido o contraditório e a ampla defesa, a autoridade competente poderá, em colaboração com as autoridades setoriais do SAI, determinar a reclassificação do sistema de IA, mediante notificação prévia, bem como determinar, de forma fundamentada, a realização de avaliação de impacto algorítmico. § 5º O resultado da avaliação preliminar poderá ser utilizado pelo agente de IA para demonstrar conformidade com os requisitos de segurança, transparência e ética previstos nesta lei. § 6º A autoridade competente e as autoridades setoriais, quando houver, poderão requerer a realização ou o acesso à avaliação preliminar do sistema de IA para fins de avaliação de risco do sistema, respeitados os segredos comerciais ou industriais.” BRASIL. Câmara dos Deputados. **Projeto de Lei n.º 2.338 de 2023**. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco. Brasília-DF: Senado

estruturas e condutas que produzam informações para relatórios e aplicação dos instrumentos *ex ante* de *Compliance* e Governança de toda tecnologia da organização).

A Comissão Temporária Interna sobre Inteligência Artificial no BRASIL<sup>209</sup>, resolveu em suas considerações já registradas no relatório apresentado em 18 de junho de 2024 e na complementação de voto de 4 de julho de 2024, e as alterações decorrentes do acatamento das emendas acima indicadas, apresenta-se substitutivo com os seguintes aprimoramentos principais, sendo: “X. avaliação preliminar passa a ser boa prática e medida de accountability para fins de identificação e gerenciamento de riscos de um sistema de IA”.

A avaliação de Impacto Algorítmico é prevista no artigo 27<sup>210</sup> da norma Europeia, prevendo aplicação para Inteligência Artificial de alto risco.

---

Federal, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1742240889313&disposition=inline>. Acesso em: 27 nov. 2023.

<sup>209</sup> Voto Complementar. BRASIL. Senado Federal. **Comissão Temporária Interna sobre Inteligência Artificial no Brasil**, p. 1-73. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9852013&ts=1732829652162>. Acesso em: 13 dez. 2024.

<sup>210</sup> Vide: “Artigo 27. O Avaliação de impacto dos sistemas de IA de risco elevado sobre os direitos fundamentais. VII. Antes de implementarem um sistema de IA de risco elevado a que se refere o artigo 6. O , n. o 2, à exceção dos sistemas de IA de risco elevado destinados a ser utilizados nos domínios enumerados no anexo III, ponto 2, os responsáveis pela implantação que sejam organismos de direito público, ou entidades privadas que prestam serviços públicos e responsáveis pela implantação de sistemas de IA de risco elevado a que se refere o anexo III, ponto 5, alíneas b) e c), devem executar uma avaliação do impacto que a utilização desse sistema possa ter nos direitos fundamentais. Para o efeito, os responsáveis pela implantação executam uma avaliação que inclua: a) Uma descrição dos processos do responsável pela implantação em que o sistema de IA de risco elevado seja utilizado de acordo com a sua finalidade prevista; b) Uma descrição do período em que o sistema de IA de risco elevado se destina a ser utilizado e com que frequência; c) As categorias de pessoas singulares e grupos suscetíveis de serem afetados no contexto específico de utilização do sistema; o d) Os riscos específicos de danos suscetíveis de terem impacto nas categorias de pessoas singulares ou grupos de pessoas identificadas nos termos da alínea c) do presente número, tendo em conta as informações facultadas pelo prestador nos termos do artigo 13. ; e) Uma descrição da aplicação das medidas de supervisão humana de acordo com as instruções de utilização; f) As medidas a tomar caso esses riscos se materializem, incluindo as disposições relativas à governação interna e aos mecanismos de apresentação de queixas. A obrigação estabelecida no n. JO L de 12.7.2024 2. O 1 aplica-se à primeira utilização do sistema de IA de risco elevado. O responsável pela implantação pode, em casos semelhantes, basear-se em avaliações de impacto sobre os direitos fundamentais efetuadas anteriormente ou em avaliações de impacto existentes realizadas pelo prestador. Se, durante a utilização do sistema de IA de risco elevado, o responsável pela implantação considerar que algum dos elementos enumerados no n. o deixou de estar atualizado, deve tomar as medidas necessárias para atualizar as informações. 3. Uma vez realizada a avaliação de impacto a que se refere o n.º 1 se alterou ou 1 do presente artigo, o responsável pela implantação deve notificar a autoridade de fiscalização do mercado dos resultados da avaliação, apresentando o modelo preenchido a que se refere o n. 5 do presente artigo como parte da notificação. No caso referido no artigo 46. O, n. pela implantação podem ser dispensados desta obrigação de notificação. 4. O 1, os responsáveis se alguma das obrigações previstas no presente artigo já tiver sido cumprida através da avaliação de impacto sobre a proteção de dados realizada nos termos do artigo 35. O do Regulamento (EU) 2016/679 ou do artigo 27. 2016/680, a avaliação de impacto sobre os direitos fundamentais a que se refere o n. o da Diretiva (EU) 1 do presente artigo deve complementar essa avaliação de impacto sobre a proteção de dados. 5. O Serviço para a IA deve desenvolver um modelo para um questionário, nomeadamente através de um sistema automatizado, a fim de facilitar aos responsáveis pela implantação o cumprimento simplificado das obrigações do presente artigo 6.” JORNAL OFICIAL DA UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828**, p. 1-144. Disponível em: [https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689). Acesso em: 29 nov. 2024.

As organizações podem tê-la como um dos instrumentos que permitam compor um protocolo de monitoramento, atualização e Explicação da decisão automatizada.

O projeto brasileiro de regulação da inteligência artificial (Projeto de Lei n.º 2.338/2023)<sup>211</sup>, prevê no artigo 22 a 26 também recomenda em seu texto a elaboração da Avaliação de Impacto Algorítmico, demonstrando como são importantes instrumentos de governança que apoiem o desenvolvimento e uso governado dos sistemas de Inteligência Artificial.

---

<sup>211</sup> “Seção III Avaliação de Impacto Algorítmico. Art. 22. A avaliação de impacto algorítmico de sistemas de inteligência artificial é obrigação dos agentes de inteligência artificial, sempre que o sistema for considerado como de alto risco pela avaliação preliminar. Parágrafo único. A autoridade competente será notificada sobre o sistema de alto risco, mediante o compartilhamento das avaliações preliminar e de impacto algorítmico. Art. 23. A avaliação de impacto algorítmico será realizada por profissional ou equipe de profissionais com conhecimentos técnicos, científicos e jurídicos necessários para realização do relatório e com independência funcional. Parágrafo único. Caberá à autoridade competente regulamentar os casos em que a realização ou auditoria da avaliação de impacto será necessariamente conduzida por profissional ou equipe de profissionais externos ao fornecedor. Art. 24. A metodologia da avaliação de impacto conterà, ao menos, as seguintes etapas: I – preparação; II – cognição do risco; III – mitigação dos riscos encontrados; IV – monitoramento. § 1º A avaliação de impacto considerará e registrará, ao menos: a) riscos conhecidos e previsíveis associados ao sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar; b) benefícios associados ao sistema de inteligência artificial; c) probabilidade de consequências adversas, incluindo o número de pessoas potencialmente impactadas; d) gravidade das consequências adversas, incluindo o esforço necessário para 109omple-las; e) lógica de funcionamento do sistema de inteligência artificial; f) processo e resultado de testes e avaliações e medidas de mitigação realizadas para verificação de possíveis impactos a direitos, com especial destaque para potenciais impactos discriminatórios; g) treinamento e ações de conscientização dos riscos associados ao sistema de inteligência artificial; h) medidas de mitigação e indicação e justificação do risco residual do sistema de inteligência artificial, acompanhado de testes de controle de qualidade frequentes; e i) medidas de transparência ao público, especialmente aos potenciais usuários do sistema, a respeito dos riscos residuais, principalmente quando envolver alto grau de nocividade ou periculosidade à saúde ou segurança dos usuários, nos termos dos artigos 9º e 10 da Lei nº 8.078, de 11 de setembro de 1990 (Código de Defesa do Consumidor). § 2º Em atenção ao princípio da precaução, quando da utilização de sistemas de inteligência artificial que possam gerar impactos irreversíveis ou de difícil reversão, a avaliação de impacto algorítmico levará em consideração também as evidências incipientes, incompletas ou especulativas. § 3º A autoridade competente poderá estabelecer outros critérios e elementos para a elaboração de avaliação de impacto, incluindo a participação dos diferentes segmentos sociais afetados, conforme risco e porte econômico da organização. § 4º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto, considerando o ciclo de vida dos sistemas de inteligência artificial de alto risco e os campos de aplicação, podendo incorporar melhores práticas setoriais. § 5º Os agentes de inteligência artificial que, posteriormente à sua introdução no mercado ou utilização em serviço, tiverem conhecimento de risco inesperado que apresentem a direitos de pessoas naturais, comunicará o fato imediatamente às autoridades competente e às pessoas afetadas pelo sistema de inteligência artificial. Art. 25. A avaliação de impacto algorítmico consistirá em processo iterativo contínuo, executado ao longo de todo o ciclo de vida dos sistemas de inteligência artificial de alto risco, requeridas atualizações periódicas. § 1º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto. § 2º A atualização da avaliação de impacto algorítmico contará também com participação pública, a partir de procedimento de consulta a partes interessadas, ainda que de maneira simplificada. Art. 26. Garantidos os segredos industrial e comercial, as conclusões da avaliação de impacto serão públicas, contendo ao menos as seguintes informações: I – descrição da finalidade pretendida para a qual o sistema será utilizado, assim como de seu contexto de uso e escopo territorial e temporal; II – medidas de mitigação dos riscos, bem como o seu patamar residual, uma vez implementada tais medidas; e III – descrição da participação de diferentes segmentos afetados, caso tenha ocorrido, nos termos do § 3º do art. 24 desta Lei.” BRASIL. Câmara dos Deputados. **Projeto de Lei n.º 2.338 de 2023**. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco. Brasília-DF: Senado Federal, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1742240889313&disposition=inline>. Acesso em: 27 nov. 2023.

A Avaliação de Impacto Algoritmo é documento regulatório que compõe e comprova ao usuário o nível de impacto em sua relação jurídica com a decisão automatizada, bem como poderá informar se uma decisão computacional tem nível de risco alto ou não sobre aquela relação jurídica estabelecida.

O artigo 17 do projeto de Lei n.º 2.338 de 2.023, trata da governança da IA, com a proposição em seu Capítulo IV, detalha a governança de sistemas de IA destacando a importância da segurança, dos direitos da pessoa e grupos afetados pela IA:

Art. 17 - Os agentes de IA estabelecerão processos internos aptos a garantir a segurança dos sistemas e o atendimento dos direitos de pessoa ou grupos afetados, nos termos previstos no Capítulo II desta Lei e da legislação pertinente, observados os segredos comerciais e industriais, nos termos do regulamento, incluindo, ao menos, medidas de transparência quanto ao emprego e à governança de sistemas de IA em especial para mitigação e prevenção de potenciais vieses discriminatórios.<sup>212</sup>

Além da avaliação de impacto algorítmico, temos na intervenção humana um fator que se necessário para o monitoramento e controle das saídas (outpus) das IA, pois essa observação, monitoramento e atualização dos Frameworks de Governança de I.A.

Kleinberg, Lakkaraju, Leskovec, Ludwig e Mullainathan<sup>213</sup>, complementam a pesquisa conceituando ao afirmar que “O valor do human-in-the-loop está na capacidade humana de identificar contextos excepcionais e considerar fatores que os sistemas automatizados não podem adequadamente processar.”

Veale, Van Kleek e Binns citando aspectos práticos para a pesquisa:

Para ser efetivo, o human-in-the-loop requer:  
 Tempo adequado para análise  
 Acesso a informações relevantes  
 Autoridade real para modificar decisões  
 Proteção contra pressões de automação.<sup>214</sup>

<sup>212</sup> BRASIL. Câmara dos Deputados. **Projeto de Lei n.º 2.338 de 2023**. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco. Brasília-DF: Senado Federal, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1742240889313&disposition=inline>. Acesso em: 27 nov. 2023.

<sup>213</sup> KLEINBERG, Jon *et al.* Human Decisions and Machine Predictions. **The Quarterly Journal of Economics**, v. 133, n. 1, p. 240, 2018.

<sup>214</sup> VEALE, Michael; VAN KLEEK, Max; BINNS, Reuben. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. **CHI Conference on Human Factors in Computing Systems**, n. 440, p. 7, 21 apr. 2018.



A efetividade do conceito de “human-in-the-loop” é manifestada por Crawford e Calo<sup>215</sup> apontando que: “A efetividade do human-in-the-loop depende crucialmente da qualidade da interface humano-máquina e da real capacidade de intervenção humana significativa”.

A efetividade do “human-in-the-loop” é conferir o desenvolvimento e utilização da Inteligência Artificial ética, transparente, robusta e que permite explicações e explicabilidade de suas decisões e resultados, que se tornam mais garantidos quanto: (i) Autonomia decisória; (ii) Proteção contra pressões; (iii) Recursos adequados; (iv) Suporte institucional.

Schwede expõe como a regulação da inteligência artificial, pode auxiliar para decisões automatizadas mais éticas e úteis para a sociedade, uma vez que enunciaram a proteção de valores e princípios de Direitos Fundamentais e éticos das decisões de máquinas para a sociedade, menciona:

Aliás, ainda que se discuta qual tipo de ética seria mais adequada a seguir, sabe-se que é extremamente urgente tomar medidas para que regulem a criação e uso dos sistemas inteligentes, dado os impactos negativos que às vezes as tecnologias podem causar. Por isso, é necessário pensar em formas para que sistemas autônomos não diferenciem injustamente as pessoas e não amplifiquem preconceitos sociais. [...] Como já observado, dados não são neutros, assim como os próprios sistemas inteligentes, pois aprendem com esses dados e replicam os vieses neles presentes. As decisões, muitas vezes, também são consideradas neutras, trazendo uma falsa noção de que seriam imparciais por serem operadas por métricas de acurácia e precisão. A partir desse ponto surge a necessidade de que os desenvolvedores de IA entendam e reflitam sobre o seu papel e sua responsabilidade no desenvolvimento de sistemas que respeitem padrões éticos, para não amplificarem na sociedade os vieses que os conjuntos de dados contêm. É necessário fazer com que os sistemas operem para identificar grupos que possam ser prejudicados com os *outputs* das IAs, e criar sistemas transparentes capazes de explicar a sua tomada de decisão. Essas duas abordagens também podem servir como parâmetro para sustentar um comportamento ético nos sistemas.<sup>216</sup>

O autor também menciona como devemos atuar com ações afirmativas em conjunto com a regulação para melhorar os dados que serão inseridos, e sua relação com as decisões automatizadas e o exercício do Direito à Explicação, que estão em relação constante uma com a outra, informando:

Outro ponto importante de destacar é que os cidadãos devem estar a par da situação e dos seus direitos de privacidade e propriedade dos dados, para entender o tratamento e ter o mínimo de conhecimento da utilização de um grande conjunto de dados e todas as decisões que envolvam um

<sup>215</sup> CRAWFORD, Kate; CALO, Ryan. **There is a blind spot in AI research**. Nature, v. 538, n. 7625, p. 312, 2016.

<sup>216</sup> SCHWEDE, Matheus Antes. **Inteligência Artificial Tendenciosa - Discriminação Racial pelo Viés Algorítmico**. Londrina: Thoth, 2023, p. 140-141.

funcionamento com base de dados. Ainda, os cidadãos também podem verificar se as decisões automatizadas estão de acordo com os padrões [éticos da sociedade, por isso se fala da transparência explicável ao público.<sup>217</sup>

Schwede descreve como o conceito de *accountability* - prestação de contas ao indivíduo destinatário da decisão automatizada, ou seja, requisitos e elementos que podem auxiliar sobremaneira o Direito à Explicação da decisão algorítmica, composto por boas práticas, *compliance* digital, governança da IA e ações afirmativas, podem proporcionar um ambiente mais ético e confiável para utilização da Inteligência Artificial e possibilitará maior exercício de aferição e prestação de contas para decisões automatizadas, pois:

Um estado democrático exige que as decisões sejam explicadas, o que parece ser difícil de conciliar com um sistema autônomo. Em áreas como a justiça criminal, as decisões não podem ser adotadas com base apenas em levantamentos de estatísticas de decisões passadas, mas sim devem ser analisados os casos concretos. Por exemplo, relatar os motivos que levaram à determinada decisão. Porém, há de se considerar que somente isso não basta. É necessário exigir a chamada *accountability* desses sistemas. Assim, como *fairness* e transparência são necessários e se completam para que se alcance uma IA confiável ao público e às instituições, a chamada *accountability* que pode ser falada no sentido de “responsabilização” ou “prestação de contas” pelas decisões algorítmicas complementa esses dois fundamentos acima citados, andando “de mãos dadas” com a transparência, pois é por meio dela que será contestado aquilo que se teve como resultado e que será estudado a seguir. 2.3. ACCOUNTABILITY. Como dito acima, *accountability* é uma das peças fundamentais para alcançar uma inteligência artificial que tenha confiança da população. [...] Nesse livro, pensa-se na ideia de uma “responsabilização” no sentido de “prestar contas” acerca dos resultados proferidos pelos algoritmos.<sup>218</sup>

A importância desse capítulo, sobre os meios pelos quais desenvolvimento e utilização das Inteligências Artificiais devem ser acompanhados pela conformidade digital (*Compliance Digital*) e pela Governança da IA, proporcionando condições de maior assertividade (acurácia) e prestação de contas (*accountability* para Explicação), tem conexão não somente com o ideal de mitigar riscos e fragilidades, mas também na aferição sobre como decisões automatizadas podem causar (no sentido preventivo) danos, ou caso esse danos sejam causados, seja possível verificar o nexo de causalidade existente entre o *output* (resultado ou decisão da máquina),

<sup>217</sup> BRASIL. Câmara dos Deputados. **Projeto de Lei n.º 2.338 de 2023**. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco. Brasília-DF: Senado Federal, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1742240889313&disposition=inline>. Acesso em: 27 nov. 2023.

<sup>218</sup> SCHWEDE, Matheus Antes. **Inteligência Artificial Tendenciosa - Discriminação Racial pelo Viés Algorítmico**. Londrina: Thoth, 2023, p. 106.

sejam inclusive por vieses, preconceitos, problemas com dados ou decorrentes do processamento e aprendizado da IA, e o dever de indenizar.

O Direito à Explicação poderá auxiliar as partes da relação jurídica decorrente da decisão automatizada na responsabilização, e por isso a pesquisa estudou maneiras de prevenir tais problemas e tornar o Direito à Explicação mais confiável e robusto.

No presente e “Em pouco tempo, as principais decisões que dizem respeito à vida das pessoas serão tomadas preponderantemente com base em tratamento automatizado de dados”, e compartilha em sua obra “Mas os algoritmos podem falhar. E causar danos.”<sup>219</sup>

O autor retro mencionado continua seu pensamento e entende que “Emerge daí interessante questão de responsabilidade civil a partir dessa utilização indevida de algoritmos sobre cujo funcionamento pouco se sabe, diante da opacidade técnica.”<sup>220</sup>

O dever de cuidado – ou diligência – para as organizações é fundamental, e decorre do adimplemento de obrigações de fazer constituídas pelo *Compliance* Digital e pela Governança de IA, que poderão ser auxiliados pelos mecanismos propostos neste capítulo.

E no que tange a responsabilidade dos agentes que desenvolvem, e daqueles que se utilizam dos sistemas de Inteligência Artificial, na esfera interna, Medon ainda escreve:

Veja-se, em primeiro lugar, a responsabilidade dita interna. Para nortear esta investigação, imagine-se um exemplo, como tantos já analisados, de um determinado agente empresarial, que decida implementar no âmbito de uma sociedade empresária sistemas de diferentes tipos de algoritmos comandados por Inteligência Artificial utilizados para diversos fins, como a avaliação de oportunidades de investimento, ranqueamento de consumidores e até mesmo reconhecimento facial. Suponha-se que um desses sistemas algoritmos tenha falhado e, além de causar danos aos consumidores, tenha imposto danos seríssimos à sociedade empresária, num investimento malfeito. Qual seria o regime de responsabilidade civil a ele aplicável?<sup>221</sup>

E continua o autor ao especificar sobre o não cumprimento de medidas preventivas, boas práticas, governança e ações afirmativas, ao qual denomina como “violação ao dever de diligência”.<sup>222</sup>

Em defesa de que a responsabilidade civil dos administradores por decisões tomadas com base em sistemas de inteligência artificial é de natureza subjetiva, como consequência de

---

<sup>219</sup> MEDON, Filipe. **Inteligência Artificial e Reponsabilidade Civil: autonomia, riscos e solidariedade**. Salvador: Juspodivm, 2020, p. 220.

<sup>220</sup> *Ibidem*, p. 255.

<sup>221</sup> *Ibidem*, p. 309.

<sup>222</sup> *Ibidem*, p. 309.

uma violação ao dever de diligência, tanto pela *culpa in eligendo* na escolha da tecnologia, como pela *culpa in vigilando* decorrente de monitoramento da tecnologia e pela sua má atuação.

Isso demonstra que para utilizarmos Inteligências Artificiais institucionalmente, de forma transparente e ética, é importante que conformidade e integridade digital, governança de dados, de inteligência artificial e ações afirmativas são parte do dever do gestor para com os indivíduos destinatários das decisões automatizadas, nas diversas vertentes em que ela atua dentro de um ecossistema e ambiente.

Deveres só podem ser analisados com exatidão ante o caso concreto, pois se conformam a qualquer relação contratual, inclusive a existente entre organizações, Inteligência Artificial com sua decisão e seus usuários, e, portanto, existe um dever de cuidado aplicável a todos as instituições e suas plataformas digitais e *softwares*.

As ações de *compliance* digital, governança da IA e demais ações afirmativas pesquisadas e citadas (sem esgotar o tema) têm importância para desenvolvimento, monitoramento, atualização, ética, robustez, transparência e conformidade para Inteligência Artificial, promovendo meios e formas de concretização mais eficaz do Direito à Explicação.

Componente importante desta pesquisa e ao qual não poderia deixar de citar neste capítulo é a instituição de Comitê de Ética para desenvolvimento e implementação de Inteligência Artificial como parte do estabelecimento de confiança, responsabilidade e como o próprio nome diz, ética. O Comitê de Ética é órgão importante da governança de Inteligência Artificial.

Ferferbaum, Klafke e Latini relatam:

Nesta seção, definimos esses comitês como estruturas colegiadas e independentes, responsáveis pelo direcionamento dos assuntos relacionados à ética em sistemas de IA nas organizações. Por sistemas de IA, adotamos um conceito amplo e alinhado com propostas regulatórias recentes, afastando-se da noção de que IA é apenas aquela que aprende.<sup>223</sup>

Portanto, é um órgão colegiado independente e responsável pelos sistemas de IA e decisões relacionadas as diretrizes éticas de projetos, programas, riscos, gestão, monitoramento, decisão com o objetivo de garantir ética e responsabilidade.

---

<sup>223</sup> FERFERBAUM, Marina; KLAFKE, Guilherme Forma; LATINI, Lucas Maldonado Diz. **Comitês de Ética para o desenvolvimento e implementação de IA confiável e responsável**. In: FACHINETTI, Aline Fuke. **Inteligência Artificial Responsável: programa de governança e melhores práticas organizacionais**. São Paulo: Thomson Reuters, 2024, p. 219.

Comitês de Ética “podem ter uma finalidade educacional, englobando as finalidades de informação e orientação.” E “também podem ter a finalidade mais relevante para a governança das organizações, participando efetivamente dos processos de tomada de decisão em políticas, projetos e programas.”<sup>224</sup>

Reforça a importância do dever de informação contido no processo de prestar conta aos usuários, conforme descreve Bioni conceitua: “A *accountability* é um roteiro não apenas para que o rito de um processo de tomada de decisão seja considerado justo, mas também, para aumentar a probabilidade de que seu resultado o seja”.<sup>225</sup>

Reforça, Vainzof, a pesquisa ao manifestar seu entendimento sobre o que foi pesquisado e escrito nesse capítulo, tornando o desenvolvimento e utilização das IA ético:

Mesmo no caso de ausência de legislações específicas sobre IA, boas práticas empresariais baseadas em responsabilização e prestação de contas podem e devem ser implementadas para sistemas de IA com base em diretrizes éticas, que são pautadas em direitos e garantias individuais. Valores éticos em que nossa sociedade se baseia devem ser plenamente integrados na forma como a IA evolui. Ou seja, organizações devem implementar programa ético de desenvolvimento ou de utilização de sistemas de IA abarcando elementos essenciais de responsabilidade e conformidade com as normas gerais (CF, CDC, CC, LGPD, Marco Civil da Internet, Lei do Cadastro Positivo, apenas para citar alguns exemplos), confirmando a eficácia do referido programa e garantindo sua melhoria contínua, além terem evidências e serem capazes de demonstrar a eficácia dele internamente ou perante terceiros.<sup>226</sup>

O Regulamento Europeu 1689 de 12 de julho de 2024<sup>227</sup> que cria regras harmonizadas em indicar para as organizações que efetivamente sejam efetivados os termos do Regulamento

<sup>224</sup> FERFERBAUM, Marina; KLAFKE, Guilherme Forma; LATINI, Lucas Maldonado Diz. **Comitês de Ética para o desenvolvimento e implementação de IA confiável e responsável.** In: FACHINETTI, Aline Fuke. **Inteligência Artificial Responsável: programa de governança e melhores práticas organizacionais.** São Paulo: Thomson Reuters, 2024, p. 229.

<sup>225</sup> BIONI, Bruno Ricardo. **Regulação e Proteção de Dados Pessoais – O Princípio da *Accountability*.** 1.<sup>a</sup> ed. São Paulo. Forense. 2022, p. 221-222.

<sup>226</sup> VAINZOF, Rony. **Aplicações e Regulação da Inteligência Artificial.** In: VAINZOF, Rony; SERAFINO, Danielle; STEINWASCHER, Aline. **Legal Innovation: o Futuro do Direito e o Direito do Futuro.** São Paulo: Thomson Reuters, 2022, p. 248.

<sup>227</sup> Vide: (4) do TFUE, A IA é uma família de tecnologias em rápida evolução que contribui para um vasto conjunto de benefícios económicos, ambientais e sociais em todo o leque de indústrias e atividades sociais. Ao melhorar as previsões, otimizar as operações e a repartição de recursos e personalizar as soluções digitais disponibilizadas às pessoas e às organizações, a utilização da IA pode conferir importantes vantagens competitivas às empresas e contribuir para progressos sociais e ambientais, por exemplo, nos cuidados de saúde, na agricultura, na segurança alimentar, na educação e na formação, nos meios de comunicação social, no desporto, na cultura, na gestão das infraestruturas, na energia, nos transportes e na logística, nos serviços públicos, na segurança, na justiça, na eficiência energética e dos recursos, na monitorização ambiental, na preservação e recuperação da biodiversidade e dos ecossistemas e na atenuação das alterações climáticas e adaptação às mesmas. (5) (6) (7) (8) Ao mesmo tempo, em função das circunstâncias relativas à sua aplicação, utilização e nível de evolução tecnológica específicos, a IA pode criar riscos e prejudicar interesses públicos e direitos fundamentais protegidos pela legislação da União. Esses prejuízos podem ser materiais ou imateriais, incluindo danos físicos, psicológicos,

e das implementações conformes com frameworks de Governança de IA para desenvolvimento e utilização da tecnologia da IA protegendo direitos e preservando princípios e valores fundamentais e democráticos conquistados pela história do ser-humano.

Equilíbrio entre inovação, desenvolvimento e regulação aplicadas a Inteligência Artificial como comenta Cantarini, quando o tema vinha sendo discutido em termos de desafios regulatórios e riscos, equilibrando a inovação e tecnologia com o desenvolvimento humano:

A nova regulamentação segue a ótica já traçada pela GDPR, por trazer uma regulamentação forte, em comparação com a regulamentação apenas moderada dos EUA, e fraca no Brasil, segundo parte da doutrina, seguindo-se a perspectiva de “human rights by design”, “beneficial AI”, “AI for good” e “HumanCentered AI”, ou seja, visa-se trazer um balanceamento entre desenvolvimento tecnológico, de modo a não obstar a inovação, de um lado, e a proteção dos valores democráticos, direitos humanos e fundamentais, de outro lado.<sup>228</sup>

E ao tratar do tema risco-regulação encontra-se alinhado com desenvolvimento e uso ético, confiável e robusto, propostos pela Governança de sistemas de Inteligência Artificial, segue a autora:

A “regulação do risco”, típica do direito ambiental, possui elemento central o risco, vinculado a processos de reunião de informação e cognição de riscos, e criação de instrumentos de regulação *ex ante*, como códigos de conduta, certificações, auditorias independentes, elaboração de documentos DPIA – Relatório de Impacto de proteção de dados e LIA – Avaliação do Legítimo Interesse.<sup>229</sup>

---

sociais ou económicos. O Tendo em conta o grande impacto que a IA pode ter na sociedade e a necessidade de criar confiança, é fundamental que a IA e o respetivo regime regulamentar sejam desenvolvidos em conformidade com os valores da União consagrados no artigo 2.º do Tratado da União Europeia (TUE), com os direitos e liberdades fundamentais consagrados nos Tratados e, nos termos do artigo 6.º do TUE, com a Carta. Como condição prévia, a IA deverá ser uma tecnologia centrada no ser humano. Deverá servir de instrumento para as pessoas, com o objetivo último de aumentar o bem-estar humano. A fim de assegurar um nível elevado e coerente de proteção dos interesses públicos nos domínios da saúde, da segurança e dos direitos fundamentais, deverão ser estabelecidas regras comuns aplicáveis a todos os sistemas de IA de risco elevado. Essas normas deverão ser coerentes com a Carta, não discriminatórias e estar em consonância com os compromissos comerciais internacionais da União. Deverão também ter em conta a Declaração Europeia sobre os Direitos e Princípios Digitais para a Década Digital e as Orientações Éticas para uma IA de Confiança do Grupo de Peritos de Alto Nível em IA (AI HLEG, do inglês High-Level Expert Group on Artificial Intelligence).” **JORNAL OFICIAL DA UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828**, p. 1-144. Disponível em: [https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689). Acesso em: 29 nov. 2024.

<sup>228</sup> CANTARINI, Paola. Inteligência Artificial: Desafios Regulatórios e Risquificação. In: PINHO, Anna Carolina et al. **Discussões sobre Direito na Era Digital**. 1.ª ed. Rio de Janeiro: GZ, 2021, p. 548.

<sup>229</sup> *Ibidem*, p. 550.

Ações afirmativas como pré-processar os dados de treinamento (ou seja, alterar alguns dos dados, remover algumas variáveis que correlacionam com a variável sensível, ou alterar alguns valores de destino previstos para que uma restrição de equidade seja satisfeita) pode auxiliar a otimizar seu algoritmo de classificação e/ou permitir que a máquina acesse os atributos sensíveis enquanto aprende com os dados de treinamento, isso tem como objetivo e prevenir problemas nas saídas (decisões) proferidas pelas máquinas, como aponta Bent:

5. Implementando Soluções de Justiça Como o exposto acima, os estudiosos de aprendizado de máquina estão trabalhando em maneiras interessantes, complexas e criativas de abordar o problema fundamental da equidade nos algoritmos de aprendizado de máquina. As principais propostas exigem algum tipo de ajuste de conscientização sobre características protegidas em algum ponto das etapas de aprendizado de máquina para restringir ou alterar a saída que o algoritmo teria gerado caso estivesse completamente sem restrições. Os meios pelos quais essas soluções de equidade são realmente implementadas no processo variam. Possibilidades incluem: pré-processar os dados de treinamento (ou seja, alterar alguns dos dados, remover algumas variáveis que correlacionam com a variável sensível, ou alterar alguns valores de destino previstos para que uma restrição de equidade seja satisfeita); impor algum tipo de penalidade por injustiça à medida que a máquina aprende padrões e otimiza seu algoritmo de classificação ou permitir que a máquina acesse os atributos sensíveis enquanto aprende com os dados de treinamento, mas depois escondê-los da máquina quando ela faz novas previsões. Para fins da lei antidiscriminação, o método de implementação pode se mostrar importante. O momento em que a solução de equidade e a variável sensível são inseridas no processo pode ser relevante sob a atual doutrina de ação afirmativa. E se os tribunais estiverem preocupados com a aparência do processo e os potenciais efeitos de balkanização, então a transparência da solução de equidade e a capacidade de identificar os indivíduos que são favorecidos ou prejudicados pela restrição de equidade podem ser importantes.<sup>230</sup>

Ações afirmativas e, Governança, para Inteligência Artificial são certamente meios mitigadores de riscos, mas sua principal função é tornar as tecnologias eficazes, éticas, confiáveis, e suas decisões automatizadas sejam proferidas da melhor maneira possível, como manifesta Reolon<sup>231</sup>, ao concluir artigo em que comenta a governança de IA normatizada no Regulamento Europeu (IA Act) para quem “A adoção de políticas, a gestão de riscos, as exigências relativas à documentação, gestão de incidentes entre outros são todos fatores já utilizados por muito tempo em todo mercado e com eficácia comprovada.”

<sup>230</sup> BENT, Jason R. **Is algorithmic affirmative action legal?** *The Georgetown Law Journal*, v. 108, p. 803, 2019. Disponível em: <https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2020/04/Is-Algorithmic-Affirmative-Action-Legal.pdf>. Acesso em: 1.º dez. 2024.

<sup>231</sup> REOLON, Carlos Eduardo Passerani. Governança técnica dos algoritmos de IA. In: VAINZOF, Rony *et al.* **Comentários ao EU AI Act: Uma abordagem prática e teórica do Artificial Intelligence Act da União Européia**. 1.ª ed. São Paulo: Revista dos Tribunais, 2024, p. 256.

Ações afirmativas tem como escopo proteger públicos vulneráveis, que poderão ser prejudicados pelas decisões automatizadas discriminatórias, sendo mais uma ferramenta auxiliar para que organizações possam prevenir problemas advindos das saídas (*outputs*) proferidos, e Requião e Costa:

A comunidade jurídica especializada no tema tem se voltado para o estabelecimento de um conjunto de regras harmônicas e requisitos específicos para minimizar o risco da ocorrência de discriminações algorítmicas. No atual estado da arte, é possível vislumbrar duas frentes de combate que se complementam: uma voltada à regulação dos sistemas de IA, sobretudo quanto à concepção, desenvolvimento e utilização de algoritmos, e outra voltada para a proteção dos dados pessoais, considerando que as decisões algorítmicas discriminatórias envolvem necessariamente a coleta, o tratamento e o compartilhamento de dados pessoais dos indivíduos. [...] Trata-se de um conjunto de medidas políticas sistematizadas num documento (incluindo mobilização de investimentos privados e públicos) que visa permitir que a Europa se torne a economia dos dados mais atrativa, segura e dinâmica do mundo, melhorando a qualidade das decisões automatizadas e, conseqüentemente, trazendo mais confiança na utilização de tais sistemas. Nessa linha, o “Livro Branco” definiu uma visão clara para a futura trajetória da IA na Europa: um ecossistema de excelência e confiança, que serviu de preparação para a recente Proposta de Regulamento do Parlamento Europeu e do Conselho, que estabelece regras harmonizadas em matéria de inteligência artificial, lançada em abril de 2021. Esta nova Proposta traz um conjunto de regras em matéria de não discriminação, com requisitos específicos que visam minimizar o risco de decisões algorítmicas enviesadas, detalhando as obrigações de testagem *ex ante*, gestão de riscos, documentação e supervisão humana ao longo do ciclo de vida dos sistemas de IA. Além disso, estabelece que, caso continuem a ocorrer violações dos direitos fundamentais em determinados sistemas de IA, as pessoas afetadas têm acesso a vias eficazes de recurso graças à garantia da transparência e da rastreabilidade dos sistemas, associadas a fortes controles posteriores do algoritmo. A Proposta estabelece uma abordagem baseada no risco, vedando ou restringindo (total ou parcialmente) a utilização de determinados sistemas de IA ou submetendo-os a obrigações mais rígidas de transparência. Dentro desse cenário, algumas práticas de IA particularmente prejudiciais foram de plano proibidas, como é o caso da utilização de sistemas de identificação biométrica à distância para efeitos de manutenção da ordem pública. Considerou-se que tais tecnologias têm sido um campo fértil para a ocorrência de discriminações algorítmicas, violando diretamente os direitos fundamentais dos indivíduos.<sup>232</sup>

O papel exercido pela gestão de organizações na promoção de instrumentos propostos neste capítulo permite aferir responsabilidades dos humanos na decisão automatizada da IA, e com isso permitir aferição quando da Explicação e da Explicabilidade quanto a riscos, efeitos e conseqüências dessas decisões.

---

<sup>232</sup> REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Revista Eletrônica de Direito Civil**, Rio de Janeiro, ano 11, n. 3, p. 7-9, 2022. Disponível em: <http://civilistica.com/discriminacao-algoritmica/>. Acesso em: 1.º dez. 2024.



Com efeito, a pesquisa e as sugestões de *compliance* e governança descritas neste capítulo poderão contribuir e conferir meios eficazes e eficientes para responder (por parte das organizações) e expressar (documentar) o Direito à Explicação e a Explicabilidade das decisões automatizadas para indivíduos, ao Estado, e até mesmo internamente, pois são exemplos de desenvolvimento e utilização de Inteligência Artificial mais assertiva, transparente, robusta e ética.

A pesquisa finaliza este capítulo trazendo instrumento de auditoria algorítmica como uma das formas de governar a inteligência artificial e as decisões automatizadas, como: o da IBM<sup>233</sup>, cuja ferramenta de código aberto que ajuda a detectar e mitigar vieses em modelos de IA.

A pesquisa verifica como regulações e normas *ex ante* são imprescindíveis para concretização de mecanismos que promovam a diminuição dos riscos, desenvolvimento e uso da tecnologia da Inteligência Artificial ética, confiável, interpretável, que permita as organizações consigam implementar, monitorar e atualizar as decisões automatizadas e conferir executoriedade para o Direito à Explicação e também Explicabilidade da IA.

---

<sup>233</sup> IBM. **Introduction to the AI Framework for z/OS**, 27 feb. 2025. Disponível em: <https://www.ibm.com/docs/en/zos/3.1.0?topic=zos-introduction-ai-framework>. Acesso em: 13 dez. 2024.

## 4 EXPLICAÇÃO E EXPLICABILIDADE DA INTELIGÊNCIA ARTIFICIAL

### 4.1 DIREITO À EXPLICAÇÃO

A pesquisa demonstrou até este capítulo que vários são os desafios para o Direito, como apontam Chiarello, Rocha e Moraes Rosa<sup>234</sup>, refletem que “O Direito está com o desafio de prover segurança para o desenvolvimento de tecnologias e que elas sejam introduzidas na rotina ainda que não saibam, exatamente, como será essa interação.”

*Big data* Brevini e Pasquale tratam do tema da pesquisa:

Muita coisa aconteceu desde 2015 que valida e desafia os principais temas do livro. Cambridge Analytica, a pontuação de crédito social e o fiasco do 'robodebt' colocaram a tomada de decisão algorítmica nos holofotes. O Regulamento Geral sobre a Proteção de Dados intensificou os debates sobre a proteção de dados. O emergente 'direito a uma explicação' da Europa sobre o perfilamento automatizado provocou longos debates sobre o processamento de informações pessoais. No entanto, exemplos recorrentes de injustiças impulsionadas por algoritmos levantam a questão de saber se a transparência – o valor normativo fundamental em *The Black Box Society* – é um primeiro passo em direção a uma implantação mais emancipatória de algoritmos e inteligência artificial (IA), é uma demanda facilmente desviada, ou realmente piora as coisas ao racionalizar a ordenação algorítmica dos assuntos humanos. Para abordar essas questões, este simpósio apresenta o trabalho de pensadores de destaque que exploraram a interação entre política, economia e cultura em domínios ordenados algorítmicamente por gerentes, burocratas e trabalhadores de tecnologia. É difícil pensar em um aspecto da vida que não tenha sido afetado pelo uso de algoritmos, automação e big data, incluindo medicina, educação, bem-estar, votação, namoro, comunicação, aplicação da lei, guerra e cibersegurança. No entanto, à medida que a tomada de decisão algorítmica se torna cada vez mais prevalente em nossas vidas públicas e privadas, tanto no nível local quanto global, ela frequentemente permanece nas mãos de intermediários de poder, como corporações e governos, inacessível a pesquisadores ou ao público em geral. Ao trazer cientistas sociais e especialistas jurídicos para o diálogo, nosso objetivo é tanto esclarecer os fundamentos teóricos dos estudos críticos de algoritmos quanto destacar a importância da pesquisa engajada, que traduz os insights da academia em uma agenda emancipatória para a reforma das leis e políticas.<sup>235</sup>

<sup>234</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha; ROSA, Alexandre de Moares. Direito para Humanos Robotizados ou Direito dos Robôs Humanizados? **Revista Novos Estudos Jurídicos**, Itajaí, v. 28, n. 3, p. 548, 2023. Disponível em: <https://periodicos.univali.br/index.php/nej/article/view/19985>. Acesso em: 24 abr. 2024.

<sup>235</sup> Texto Original: “Much has happened since 2015 that vindicates and challenges the book’s main themes. Cambridge Analytica, social credit scoring, and the ‘robodebt’ debacle have vaulted algorithmically inflected decisionmaking into the headlines. The General Data Protection Regulation has intensified debates over data protection. Europe’s emerging ‘right to an explanation’ of automated profiling has provoked lengthy debates on the processing of personal information. Yet recurring examples of algorithmically driven injustices raise the question of whether transparency– the foundational normative value in *The Black Box Society*– is a first step towards a more emancipatory deployment of algorithms and artificial intelligence (AI), is an easily deflected demand, or actually worsens matters by rationalizing the algorithmic ordering of human affairs. To address these questions, this symposium features the work of leading thinkers who have explored the interplay of politics,

No campo das relações digitais, é necessário ao jurista, tanto acadêmico como operador, pensar não somente nas relações jurídicas do ponto de vista material e processual, é necessário diversidade de pensamento, saber científico e computacional (mesmo que básico), multidisciplinariedade de conhecimento cultural, social, antropológico, econômico para que entenda como se aplicam as normas, o mercado, a tecnologia se conectam com o Estado e a Sociedade.

O desenvolvimento da autodeterminação, da proteção e da privacidade dos dados, a segurança da informação são alguns dos mecanismos (conforme o capítulo anterior demonstra) da transparência e possível explicação aos cidadãos (usuários), e permitem operacionalizar o exercício desse Direito, com potencial revisão dos resultados (*outputs*) e decisões automatizadas em casos de comprovado viés ou discriminação que não atendam aos enunciados normativos e éticos de uma Sociedade em suas relações digitais, que certamente avançarão cada vez mais.

O Direito à explicação como uma das formas de obtenção de informação deve encontrar amparo nas mais diversas formas de transparência e resposta sobre a decisão da máquina, e como essa decisão impactou na vida do indivíduo dela destinatário.

O acesso a essa informação, como dito linhas atrás pode ser feito por uma séria de documentos e exposição de métodos e lógicas sobre como os dados do indivíduo, do ambiente, da economia etc. foram tratados e inseridos para que um software proprietário (ou até mesmo sistema de inteligência artificial independente, como a Generativa), tomou sua decisão lógica, afetando também o ambiente coletivo, concorrencial e social.

O acesso e informação sobre coleta, tratamento e armazenamento de dados para titular por meio de pedido (Direito de Petição) tem amparo inicial no art. 18 da Lei n.º 13.709 de 2018.<sup>236</sup>

---

economics, and culture in domains ordered algorithmically by managers, bureaucrats, and technology workers. It is hard to think of an aspect of life that has not been affected by the use of algorithms, automation, and big data, including medicine, education, welfare, voting, dating, communication, law enforcement, warfare, and cyber-security. However, as algorithmic decision-making becomes ever more prevalent in our public and private lives at the local and global level, they all too often remain in the hands of power brokers such as corporations and governments, inaccessible to researchers or the broader public. By bringing social scientists and legal experts into dialogue, we aim both to clarify the theoretical foundations of critical algorithm studies and to highlight the importance of engaged scholarship, which translates the insights of the academy into an emancipatory agenda for law and policy reform.” PASQUALE, Frank; BREVINI, Benedetta. Revisiting the Black Box Society by rethinking the political economy of big data. *Sage Journals*, 20 oct. 2020, p. 1. Disponível em: <https://journals.sagepub.com/doi/10.1177/2053951720935146>. Acesso em: 9 dez. 2024.

<sup>236</sup> “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; II – acesso aos dados; III – correção de dados incompletos, inexatos ou desatualizados; IV – anonimização, bloqueio ou eliminação de dados desnecessários, excessivos ou tratados em desconformidade com o disposto nesta Lei; V – portabilidade dos dados a outro fornecedor de serviço ou produto, mediante requisição expressa e observados os

O artigo 19 da Lei Geral de Proteção de Dados inicia o Direito à Explicação quando define que o titular tem Direito a confirmação de existência e acesso a dados pessoais quando por ele requisitados:

Art. 19. A confirmação de existência ou o acesso a dados pessoais serão providenciados, mediante requisição do titular: I – em formato simplificado, imediatamente; ou II – por meio de declaração clara e completa, que indique a origem dos dados, a inexistência de registro, os critérios utilizados e a finalidade do tratamento, observados os segredos comercial e industrial, fornecida no prazo de até 15 (quinze) dias, contado da data do requerimento do titular. § 1º Os dados pessoais serão armazenados em formato que favoreça o exercício do direito de acesso. § 2º As informações e os dados poderão ser fornecidos, a critério do titular: I – por meio eletrônico, seguro e idôneo para esse fim; ou II – sob forma impressa. § 3º Quando o tratamento tiver origem no consentimento do titular ou em contrato, o titular poderá solicitar cópia eletrônica integral de seus dados pessoais, observados os segredos comercial e industrial, nos termos de regulamentação da autoridade nacional, em formato que permita a sua utilização subsequente, inclusive em outras operações de tratamento.

Goulart<sup>237</sup> ao comentar sobre o artigo da LGPD acima, afirma que “alguns critérios objetivos para cumprimento dos direitos de confirmação da existência e acesso aos dados”, e temos assim o estabelecimento de uma explicação sobre como um agente de tratamento coletou, tratou e armazenou dados do usuário.

O Direito a revisão de uma decisão automatizada está enunciado no artigo 20 da Lei Geral de Proteção de Dados, veja-se:

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,

---

segredos comercial e industrial, de acordo com a regulamentação do órgão controlador; VI – portabilidade dos dados a outro fornecedor de serviço ou produto, mediante requisição expressa, de acordo com a regulamentação da autoridade nacional, observados os segredos comercial e industrial; VII – eliminação dos dados pessoais tratados com o consentimento do titular, exceto nas hipóteses previstas no art. 16 desta Lei; VIII – informação das entidades públicas e privadas com as quais o controlador realizou uso compartilhado de dados; IX – informação sobre a possibilidade de não fornecer consentimento e sobre as consequências da negativa; X – revogação do consentimento, nos termos do § 5º do art. 8º desta Lei. § 1º O titular dos dados pessoais tem o direito de peticionar em relação aos seus dados contra o controlador perante a autoridade nacional. § 2º O titular pode opor-se a tratamento realizado com fundamento em uma das hipóteses de dispensa de consentimento, em caso de descumprimento ao disposto nesta Lei. § 3º Os direitos previstos neste artigo serão exercidos mediante requerimento expresso do titular ou de representante legalmente constituído, a agente de tratamento.” BRASIL. **Lei n.º 13.709, de 14 de agosto de 2018.** Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília-DF: Presidência da República, 2025. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709compilado.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm). Acesso em: 6 abr. 2025.

<sup>237</sup> GOULART, Guilherme Damásio. Comentários à Lei Geral de Proteção de Dados. In: MARTINS, Guilherme Magalhães; LONGHI, João Victor Rozatti; FALEIROS JÚNIOR, José Luiz de Moura. **Comentários à Lei Geral de Proteção de Dados**. 1.ª ed. Indaiatuba: Foco, 2022, p. 263.

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.

Goulart comenta também o artigo 20 da Lei Geral de Proteção de Dados, descreve a ideia e a importância contida no artigo acima contido:

A tutela dos direitos do titular diante da tomada de decisões automatizadas é outro ponto de bastante importância na disciplina de proteção de dados pessoais. A ideia é oferecer uma proteção ao titular de dados pessoais contra não somente ingerências extremas, mas também contra processos automatizados de tratamento de dados que possam, pela impessoalidade e pela escala de tais tratamentos, causar impactos negativos na vida dos sujeitos. [...] Muitas atividades cotidianamente utilizadas pelas pessoas tratam dados pessoais, em maior ou menor grau, com a utilização de perfis e decisões automatizadas para alavancar seus negócios. [...] Note-se que o art. 20 aborda duas questões: a tomada de decisões automatizadas que afetem os seus interesses (de maneira geral) e as questões relacionadas às decisões automatizadas específicas na definição de perfis (a chamada perfilização ou *profinling*).<sup>238</sup>

Na Europa, o *General Data Protection Regulation* (GDPR), tem uma maior amplitude do que na Lei Geral de Proteção de Dados acima descrita no que se refere a Explicação da decisão automatizada e sua revisão humana, conforme Dourado:

O GDPR garante direitos a todas as pessoas que sejam influenciadas por decisões automatizadas. Pelo regulamento europeu, os titulares de dados têm direito de saberem quando há decisões automatizadas que os afetem significativamente ou produzam efeitos na sua esfera jurídica e de receberem “informações úteis relativas à lógica subjacente, bem como a importância e as suas consequências previstas de tal tratamento” (artigos 13º, 14º e 15º). Além disso, todo titular de dados tem o “direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis”, a não ser que a decisão seja necessária para celebração ou execução de um contrato ou for baseada no consentimento explícito do titular – nesses casos, o GDPR dá ao titular de dados o direito de solicitar intervenção humana e de contestar a decisão automatizada (artigo 22º).<sup>239</sup>

<sup>238</sup> GOULART, Guilherme Damásio. Comentários à Lei Geral de Proteção de Dados. In: MARTINS, Guilherme Magalhães; LONGHI, João Victor Rozatti; FALEIROS JÚNIOR, José Luiz de Moura. **Comentários à Lei Geral de Proteção de Dados**. 1.ª ed. Indaiatuba: Foco, 2022, p. 268.

<sup>239</sup> DOURADO, Daniel de Araújo. Direito à Explicação e Regulação da Inteligência Artificial na Saúde: perspectiva da LGPD. In: AITH, Fernando; DALLARI, Annaluz Bolivar. **LGPD na Saúde Digital**. São Paulo: Thomson Reuters, 2022, p. 115-130.

No regulamento europeu, portanto, a normatização mais restrita do Direito à Explicação possui diferenças do enunciado no artigo 20 da lei geral de proteção de dados brasileira, que é mais aberta a possibilidade de solicitação da explicação, pois como a pesquisa demonstra, dados pessoais não somente permitem identificar (ou. Tornar identificável) um indivíduo, mas são necessários para uma decisão que o afete diretamente em qualquer contexto, e como o final do caput do artigo 20 da LGPD menciona “os aspectos de sua personalidade”, há possibilidade de solicitar Direito à Explicação em diversos contextos em que o usuário possa solicitar informações referentes a decisão automatizada.

Expõe Dourado ainda, que “O direito à explicação está vinculado aos limites da transparência algorítmica”, sendo essa definida pelo autor como:

A transparência de um sistema de IA se concentra principalmente no processo, ou seja, trata-se de permitir que as pessoas entendam como os algoritmos são desenvolvidos e implantados em termos gerais. Eventualmente pode incluir elementos sobre fatores que de uma previsão ou decisão específica, mas em regra não inclui o compartilhamento de códigos ou. Conjuntos de dados. Logo, a existência de alguma opacidade é inevitável. Essa opacidade tem relação com a dimensão de caixa-preta, pela complexidade dos sistemas, mas não só. Há opacidade imposta por sigilo corporativo ou de estado (intencional), pois o compartilhamento de códigos ou conjuntos de dados específicos pode revelar segredos comerciais ou divulgar dados confidenciais do usuário, e opacidade decorrente do próprio “analfabetismo técnico” dos usuários.<sup>240</sup>

A pesquisa demonstra que o entendimento acima não é – com permissão para discordar concernente a opacidade, uma premissa que possa ser considerada verdadeira.

Isto porque com os instrumentos de governança de dados e governança de inteligência artificial, e seus documentos, é possível dar respostas mais assertivas sem que seja necessário apresentar códigos ou programações de software (seja de desenvolvimento ou estrutura de rede neural) de uma I.A., sendo possível, portanto, apresentar Explicação sobre dados coletados e inseridos nos sistemas que influenciaram na decisão automatizada, conferindo a explicação necessária.

Desafio que se coloca para o Direito à Explicação é entender como expressar ao usuário o ponto de vista intrínseco da decisão automatizada e sua exteriorização (internamente a matemática e, externamente, o efeito de fato que o Direito confere valor), como leciona por Frazão:

---

<sup>240</sup> DOURADO, Daniel de Araújo. Direito à Explicação e Regulação da Inteligência Artificial na Saúde: perspectiva da LGPD. *In*: AITH, Fernando; DALLARI, Annaluz Bolivar. **LGPD na Saúde Digital**. São Paulo: Thomson Reuters, 2022, p. 121.

Objetiva-se aqui entender como decisões automatizadas, que utilizam tecnologia de inteligência artificial (em versões avançadas das máquinas de Turing), podem representar um desafio ao direito, por colocarem em conflito duas ordens de pensamento diversas: a linguagem matemática dos algoritmos e a linguagem natural, própria do direito.<sup>241</sup>

Para Frazão acima, o potencial de revisão tem como pressuposto a explicação: “Reavaliado como elo entre modalidades regulatórias distintas, o direito à explicação se representa como uma tradução para a linguagem. Natural dos elementos da infraestrutura matemática que levaram à decisão automatizada.”<sup>242</sup>

Outra questão refutada pela pesquisa, e que será tratada no próximo capítulo são as inteligências artificiais explicáveis (XAI), sistemas desenvolvidos para explicar decisões e resultados das máquinas.

Terceiro ponto que merece consideração, é o do citado “analfabetismo do usuário”, que não justifica alegação de opacidade, uma vez que é vulnerável e tem o direito de explicação quando uma decisão afete seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade, nos termos da parte final da cabeça do artigo 20 da lei geral de proteção de dados, não sendo necessário informar questões atinentes a segredo de negócio ou propriedade industrial, que pode ser superada com a assinatura de termo de confidencialidade e sigilo, bem como dados do usuário lhe pertencem e devem ser informados, de acordo com as premissas e princípios da legislação de proteção de dados brasileira.

E quanto a revisão da decisão automatizada pós explicação, ainda a depender de regulamentação normativa da Autoridade Nacional de Proteção de Dados, pois não há na LGPD a exata enunciação sobre referido procedimento para os agentes de tratamento de dados para com o usuário, que pode ser conferido (Direito à Explicação) independente de revisão ou alteração humana.

Conforme Godoy afirma:

Não se estabeleceu, porém, de que modo, exatamente, este procedimento de explicação e de revisão se deve dar. Em especial não se consagrou menção expressa a que se interpusesse, em seu curso, o exame, por pessoa natural, do pleito de explicação ou de revisão. E tal o que se releva porque, originalmente, ressalva a este direito do titular de dados havia no texto projetado, tendo sido,

---

<sup>241</sup> FRAZÃO, Ana; GOETTENAUER, Carlos. O jogo da imitação jurídica: o direito à revisão de decisões algorítmicas como um mecanismo para a necessária conciliação entre linguagem natural e infraestrutura matemática. *In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo et al. O Direito Civil na Era da Inteligência Artificial*. 1.ª ed. São Paulo: Thomson Reuter, 2020, p. 46.

<sup>242</sup> *Ibidem*, p. 59.

todavia, dele retirado por iniciativa específica. [...] Resultado então é a redação atual, de início transcrita, em que não se expressa o direito do titular a que o procedimento de explicação e de revisão das decisões tomadas unicamente com base em tratamento automatizado de seus dados pessoais conte com a intervenção de pessoa natural.<sup>243</sup>

A separação acima exposta é motivada para que se entenda que Direito à Explicação é uma coisa, a revisão da decisão automatizada por humano é dependente e subsequente a concretização do direito à explicação. Se a explicação foi efetivada e exposta por meio de documento com clareza e acompanhada de instrumentos que a subsidiem não está submissa e independente da revisão, que poderá ou não ocorrer.

Algoritmos e suas decisões permeiam nossa existência e interações constantemente, pois os indivíduos são submetidos a resultados proferidos pelos sistemas inteligentes após esses apreenderem com dados pessoais e corporativos, bem como demais informações que são inseridas (*inputs*), disseminadas dentro das redes neurais, propagadas e retropropagadas no processo de aprendizado de máquina (*machine learning*), para posteriormente decidirem para condução de nossa vida, diretamente em aplicações e aplicativos utilizados em aparelhos, ou nas decisões dos diferentes agentes econômicos (privados e públicos).

Parte fundamental para a pesquisa é conceito informado por Floridi de que: “O direito à explicação é essencial para manter a confiança nas decisões automatizadas e promover a transparência nas operações de IA”.<sup>244</sup>

Floridi afirma que o direito à explicação em sistemas de IA é crucial para assegurar que os indivíduos compreendam as bases das decisões automatizadas que os afetam, o que é fundamental para a justiça e a responsabilidade.<sup>245</sup>

O autor acima ainda nos alerta sobre as implicações éticas quando argumenta que a implementação do direito à explicação deve equilibrar a transparência com a proteção de propriedade intelectual e a privacidade de dados.<sup>246</sup>

A tomada de decisão por parte dos sistemas de Inteligência Artificial poderá ser objeto de explicação aferição de potenciais riscos sociais e individuais de opacidade, vieses e discriminações algorítmicas, mas:

---

<sup>243</sup> GODOY, Cláudio Luiz Bueno de Godoy. A intervenção humana no procedimento de explicação e de revisão de decisões automatizadas. In: CHINELLATO, Silmara J. de A. et al. **Inteligência Artificial: Visões Interdisciplinares e Internacionais**. 1.ª ed. São Paulo: Almedina Brasil, 2023, p. 287.

<sup>244</sup> FLORIDI, Luciano. **The Right to Explanation: Ethics, Transparency, and Accountability in AI Systems**. Oxford: Oxford University Press, 2018, p. 54.

<sup>245</sup> FLORIDI, Luciano. Ethical Implications of the Right to Explanation in AI Systems. **Journal of Ethics in Technology**, v. 25, n. 1, p. 45-60, 2020.

<sup>246</sup> FLORIDI, Luciano. Balancing Transparency and Privacy: The Right to Explanation in the Age of AI. **Journal of Digital Ethics**, v. 18, n. 2, p. 150-170, 2021.



O Direito à Explicação é algo que transborda a seara jurídica em si mesma, a governança dos dados e dos algoritmos podem prover a explicabilidade e prevenir possíveis ou eventuais problemas advindos das decisões automatizadas da Inteligência Artificial. Algumas perspectivas podem ser objeto de reflexão sobre o tema, iniciando-se pela “Explicação, que por si só, não é algo simples, e há esforços tanto da filosofia, da psicologia, como das ciências cognitivas em definir o que seja o explicar.”<sup>247</sup>

Sistemas computacionais complexos, como citado anteriormente, são capazes de aprender com os padrões de dados de privacidade e personalidade dos indivíduos, e “É sabido que as pessoas empregam certos vieses cognitivos, estereótipos e expectativas em relação ao processo de Explicação. É preciso saber o nível de complexidade com o qual e com quem explicar, sob o risco de nada explicar”.<sup>248</sup>

O raciocínio é de que o Direito à Explicação passa pela crença dos indivíduos que: “[...] acreditam que a explicação de um fenômeno psicológico envolve fatos que são relacionados com o cérebro do que quando envolve fatos que não o são, mesmo quando estas informações neurocientíficas são irrelevantes para a explicação do fenômeno em questão”.<sup>249</sup>

Na etapa de captura dos dados e seu tratamento, a elaboração de preços personalizados moldados por algoritmos, cujas decisões tomadas e impressionante velocidade influenciam o titular de dados.<sup>250</sup>

A utilização dos sistemas de IA não podem ser opacos, e a explicação é justamente uma forma de garantir transparência, “que deverá ocorrer sob dois prismas: (i) substancial (exposição de critérios); e (ii) formal (procedimentos realizados), tomando por base a pessoa do titular dos dados para assegurar a clareza e adequação das informações”.<sup>251</sup>

A Lei Geral de Proteção de Dados Pessoais (Lei n.º 13.709/2018), em seu artigo 20 dispõe sobre o Direito de:

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.

<sup>247</sup> ABRUSIO, Juliana. **Proteção de Dados na Cultura do Algoritmo**. 1.ª ed. São Paulo: D’Plácido, 2020, p. 65. (Coleção Direito privado, v. 1).

<sup>248</sup> *Ibidem*, p. 65.

<sup>249</sup> EKUNI, Roberta; POMPÉIA, Sabine. O impacto da divulgação científica na perpetuação de neuromitos na educação. **Revista de Biologia**, v. 15, n. 1, p. 22, 2016. Disponível em: <https://www.revistas.usp.br/revbiologia/article/view/114837/112576>. Acesso em: 24 abr. 2024.

<sup>250</sup> KOGA, Bruno Yudi Soares. **Precificação Personalizada**. 1.ª ed. São Paulo: Almedina, 2021, p. 294.

<sup>251</sup> *Ibidem*, p. 295.

O dispositivo confere o Direito do titular de dados saber como a Inteligência Artificial decidiu em sua relação jurídica junto a uma organização, e como esta organização – por meio da decisão automatizada – decidiu sobre possível (ou não) negócio jurídico entre ela e o referido titular destinatário da decisão.

Esse entendimento é compartilhado por Magrani e Oliveira, manifestam a possibilidade de aferir quando, se, e como uma decisão automatizada pode prejudicar ou discriminar o titular de dados, e como o Direito à Explicação é também estabelecido no Regulamento Europeu de Proteção de Dados Pessoais, a *General Data Protection Regulation* (GDPR):

Esse dispositivo consagra o direito à explicação (*right to explanation*) do titular dos dados no caso de decisões automatizadas e tem por objetivo auxiliá-lo contra a possibilidade de tratamento automatizado possivelmente discriminatório. Contudo, a previsão de que a revisão deveria ser feita por “pessoa natural” foi retirada da lei por meio de veto presidencial. A GDPR também possui previsão sobre esse direito, mas é mais protetiva do usuário ao dizer que decisões totalmente automatizadas devem ser a exceção, cabível apenas nos casos previstos no artigo 22 da regulação. Trata-se de direito de grande importância no cenário das novas tecnologias, uma vez que as empresas estão transferindo a tomada de decisões importantes antes feitas por seres humanos para máquinas. Assim, por exemplo, a análise de crédito em um banco pode ser feita unicamente de forma automatizada, ou a definição do valor de seguro de vida, e tais decisões podem se basear em sistemas de ranking ou de riscos preditivos, utilizando aspectos potencialmente discriminatórios e de desigualdade injusta, como gênero, local de residência. E raça. O ideal, portanto, é que o indivíduo tenha direito a que uma pessoa natural investigue o processo de tomada de decisão e =lhe dê explicações sobre o motivo do veredito automatizado.<sup>252</sup>

A percepção a respeito das decisões das máquinas é a de que tais vereditos sejam resultados neutros e orgânicos, mas a personalização dos dados e os resultados individualizados são fundamentais para a decisão automatizada.

Os autores do artigo supramencionado, manifestam-se no sentido de que seres humanos devam avaliar a metodologia e aplicação dos dados na tomada de decisão proferida por inteligência artificial.

No entanto, como se verá no capítulo destinado a Explicabilidade de Inteligência Artificial, temos atualmente estudos e ferramentas de Inteligências Artificiais (*Explainable*

---

<sup>252</sup> MAGRANI, Eduardo; OLIVEIRA, Renana Medeiros de. Desafios Regulatórios da Internet das Coisas: reflexões sobre LGPD, o consentimento e o direito à explicação. In: CRESPO, Marcelo Xavier de Freitas *et al.* **Compliance no Direito Digital**. São Paulo: Thomson Reuters Brasil, 2020, p. 81-82.

*Artificial Intelligence – XAI*) que permitem com assertividade e clareza explicar como se deu a tomada decisória de outra Inteligência Artificial.

Claro que não se trata de usarmos tais sistemas de inteligência artificial em todas as questões, mas com todo respeito, ao invés de pensarmos no Direito a Explicação como apenas e tão somente um relatório escrito por um indivíduo para explicar a decisão automatizada (com base nos documentos comprobatórios de Boas Práticas de *Compliance* Digital e Governança de IA), mas também a revisão de uma decisão por meio de pessoas capacitadas que podem aferir eventuais discrepâncias ou detalhes que podem modificar a decisão (ou não) em favor de um indivíduo-usuário destinatário de uma decisão automatizada.

O Direito à Explicação tem como escopo dar cumprimento ao Princípio da Transparência, verificando os resultados (decisões) automatizadas em sistemas de Inteligência Artificial. O cumprimento de normas e a satisfação do jurisdicionado sobre o conjunto de informações utilizadas na tomada decisória para que possam ser aferidas são relatos onde se verificam possíveis ocorrências de discriminação de indivíduos ou grupos de indivíduos.

O Direito à Explicação de tais decisões tomadas é o exercício de uma manifestação de vontade de uma pessoa natural ou jurídica de saber e obter acesso aos critérios ou resultados proferidos por máquinas que aprendem, pelos dados, que forma classificatória (n-gramas) e quais são os cálculos de probabilidades, pesos e metodologias serão aplicáveis.

De acordo com Floridi:

O direito à explicação deve ser visto não apenas como uma obrigação legal, mas como uma oportunidade para reforçar a compreensão e a confiança do público. Isso inclui fornecer explicações que sejam significativas e acessíveis a uma ampla variedade de usuários, considerando a complexidade dos sistemas de IA.<sup>253</sup>

E a pesquisa demonstra que o cumprimento desses ditames éticos de acesso à informação, de transparência, confiança e robustez que tornam inteligível, compreensível, aceitável e válido.

Abrusio, cita artigo de WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. **International Data Privacy Law**, v. 7, n. 2, p. 76-99, 2017, afirma:

---

<sup>253</sup> FLORIDI, Luciano. **Accountability and the Right to Explanation in AI: A Multidimensional Approach**. Cambridge: Cambridge University Press, 2019, p. 112.

Para os autores Sandra Wachter, Brent Mittelstadt e Luciano Floridi, o reconhecimento de um direito à explicação implica que os controladores de dados tenham esclarecido quais deveres exatamente devem ser cumpridos em decorrência, de modo que lhes impor multas por negligência em relação aos seus deveres, sem o devido esclarecimento, conflita com os princípios de um julgamento justo e em consonância com os fundamentos do Estado Democrático de Direito.<sup>254</sup>

### O Direito à Explicação, de acordo com Abrusio:

Implica na possibilidade do interessado obter *intervenção humana* por parte do responsável para manifestar seu ponto de vista e contestar a decisão, nos casos em que se produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar. [...] Cria-se um instrumento, portanto, em prol da defesa da pessoa humana, visando evitar decisões discriminatórias. É dizer, “a *ratio* do instituto, portanto, parece fornecer um meio de contraste para evitar ou conter preconceitos à esfera jurídica da parte interessada que possam derivar de tais decisões.”<sup>255</sup>

A ideia acima está direcionada a princípio para indivíduos, usuários, pessoas físicas de plataformas e de toda economia digital.

Ressalvamos, porém, que decisões automatizadas também são tomadas para relações entre agentes de mercados, Estado e outras relações jurídico institucionais, devendo o Direito à Explicação ser ampliado para um Direito de caráter geral para cumprir a finalidade e a adequação do processamento dos dados (princípio que é aplicado de forma geral quando se trata de Governança de Dados e Governança de IA).

Com efeito, o Direito à Explicação poderá garantir transparência e confiança no desenvolvimento, fornecimento e utilização dos sistemas computacionais de Inteligência Artificial.

A decisão automatizada pode ser considerada opaca ou enviesada, considerando-se o pressuposto do não atendimento ao interesse público, nem tão pouco aos elementos do ato administrativo aferíveis de forma simples e clara (a falta de clareza é justamente uma opacidade), a saber: (i) legalidade; (ii) motivação e/ou motivo; (iii) competência; (iv) finalidade.

Não podemos nos esquecer o nexo causal como “o elo coerente criado entre dois (ou mais) fatos ocorridos em tempo e lugar específico que deflagram uma relação indissociável de

<sup>254</sup> WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. **International Data Privacy Law**, v. 7, n. 2, p. 76-99, 2017. Disponível em: <https://academic.oup.com/idpl/article/7/2/76/3860948>. Acesso em: 28 abr. 2024.

<sup>255</sup> ABRUSIO, Juliana; RAIZ, Diogo. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. São Paulo: Revista dos Tribunais, 2022, p. 173.

ocorrência e lógica decorrente entre os fatos analisados”<sup>256</sup>, ainda mais em processos matemáticos como os resultados da Inteligência Artificial.

Ética em sistemas algorítmicos, ou seja, Ética Algorítmica dos sistemas de Inteligência Artificial e os *Softwares* a elas integrados devem cumprir ditames morais e valores sociais de igualdade, liberdade e equilíbrio. Todo o alcance de aplicação de valores éticos, morais e do Direito (princípios, valores, costumes e fontes do Direito) precisam compor objeto de ampla discussão na aplicação da I.A.

O intercepto adicionado em uma equação linear, ou seja, é o parâmetro adicional na Rede Neural usada para ajustar a saída junto da soma ponderada das entradas (ou o peso padrão que a inteligência artificial recebeu durante esse processo de aprendizado), com resultados decisórios não aferíveis ou passíveis de modificação (após aferição de opacidade, por exemplo), que devem ser considerados no processo decisório e que demandam explicação aos usuários de um determinado sistema de I.A. e seu aprendizado.

Pode-se obter explicação de um sistema, seu uso, os dados que o formaram e as decisões que não constituam descumprimento de Direitos Fundamentais e Constitucionais como à proteção de dados, liberdade, igualdade.

A fórmula da tomada de decisões em sistemas de Inteligência Artificial por meio do exercício do Direito à Explicação, está intimamente vinculada a maneira como esses sistemas são governados, aferidos em conformidade com a governança e a transparência.

É preciso cumprir os princípios da precaução e transparência para decisões na esfera judicial providas da Inteligência Artificial, pois a decisão automatizada deve expor e transmitir níveis de confiança, de *fairness*, de não discriminação de indivíduos ou grupos de indivíduos, e a explicação é a resposta que garante ao jurisdicionado que o sistema e o aprendizado estão adequados ao julgamento e ao *output* de decisões.

O Direito conferido ao titular de dados pode requerê-lo do responsável pelo tratamento dos dados<sup>257</sup> imputando à Inteligência Artificial (em aplicativo, software integrado, plataforma, sistema independente ou generativa utilizada internamente) que tenha ciência das informações, dados, critérios práticos, compreensíveis da tomada de decisão autônoma, de forma clara e relatada de maneira clara no que tange aos elementos do processo decisório.

---

<sup>256</sup> FRANÇA, Phillip Gil. **Ato Administrativo, Consequencialismo e Compliance: Gestão de Riscos, Proteção de Dados e Soluções para o Controle Judicial na Era da IA**. 4.<sup>a</sup> ed. São Paulo: Revista dos Tribunais, 2019, p. 123.

<sup>257</sup> RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 172.

Ao solicitar e exercer o Direito à Explicação, o jurisdicionado pretende que sejam respondidas questões como o cumprimento das normas, a valoração das provas e as condições instrumentais do caso concreto.

O Viés Algorítmico, a discriminação, assim definido como algo que possa ser provado como desigual, desequilibrado, ilegal ou irregular, quando além de provocar sentimentos de injustiça, diminuição, separação e segregação de um indivíduo (ou grupo) em um determinado contexto, no presente caso em uma decisão processual, pode comprovar que houve decisão em desequilíbrio com tais valores em um processo.

Discriminação vem do termo “discriminar”, que é o ato de distinguir, diferenciar ou segregar algo com base em algum critério ou categoria específica.<sup>258</sup>

A “imparcialidade do julgador é difícil de ser alcançada na medida em que, mesmo que venha tomar uma decisão considerada correta, ela pode ser enviesada, sem sequer perceber essa possibilidade e, ainda, podem ocorrer ilusões cognitivas”.<sup>259</sup>

O questionamento sobre determinada decisão automatizada para cumprimento do Direito à Explicação trafega pelo desenvolvimento do sistema a ser utilizado, seu aprendizado, qual tipo de decisão a Inteligência Artificial permitirá a solução mais adequada.

O aprendizado da máquina pode ser objeto de auditoria, atendendo ao Princípio da Transparência, e com isso sabe-se como a I.A. decidiu, se foi com a lei correta aplicável ao caso, alinhada com as normas e com as demais fontes do Direito inerentes às decisões judiciais, permitindo, inclusive, a verificação pelo Poder Judiciário e, dessa forma, fazendo-se cumprir o Direito à Explicação.

A fundamentação das decisões automatizadas tomadas por Inteligência Artificial, principalmente de Aprendizado Profundo (*Deep Learnings*), intensifica a discussão e reflexão sobre a Ética Algorítmica, vieses, discriminações; a Explicabilidade deve ser considerada um dos pilares fundamentais para utilização, proporcionando a aferição de tais decisões, buscando evitar, assim, resultados inesperados prejudiciais ou que coloquem o jurisdicionado em risco de dano de difícil ou impossível reparação, prejudicando-o individual ou coletivamente.

Afinal, as pessoas são vulneráveis à luz de tecnologias como a IA porque, na medida em que sua agência aumenta, delega-se mais tarefas a elas e todos se tornam mais dependentes.

---

<sup>258</sup> PARANHOS, Mário Cosac Oliveira. **Viés Algorítmico: uma análise sobre discriminações automatizadas**. Rio de Janeiro: Lumen Iuris, 2022, p. 8.

<sup>259</sup> SCHWEDE, Matheus Antes. **Inteligência Artificial Tendenciosa - Discriminação Racial pelo Viés Algorítmico**. Londrina: Thoth, 2023, p. 29.

Assim, as novas vulnerabilidades tecnológicas não são meramente tecnológicas, tornam-se vulnerabilidades humanas, existenciais.<sup>260</sup> A Inteligência Artificial é falível, e sua programação pode conter escrita enviesada ou incorreta na linguagem computacional, na equação matemática ou em dados que são imputados – e direcionados no aprendizado –, gerando conclusões erradas e de possível desconformidade, o que somente pode ser analisado quando se solicita uma explicação.

A empresa IBM (*International Business Machine*)<sup>261</sup> disponibiliza uma ferramenta de Explicabilidade de Inteligência Artificial, que demonstra como algoritmos podem explicar as Decisões Automatizadas proferidas por Inteligências Artificiais.

A regulação proposta pelo Projeto de Lei n.º 2.338 de 2023, conforme complementação do voto<sup>262</sup> da Comissão Interna para inclusão de ementas, resolveu em suas considerações já registradas no relatório apresentado em 18 de junho de 2024 e na complementação de voto de 4 de julho de 2024, e as alterações decorrentes do acatamento das emendas acima indicadas, apresenta-se substitutivo com os seguintes aprimoramentos principais, sendo:

VII. supressão do “direito à determinação humana” que passa a ser aplicável apenas a IAs de alto risco, o que é englobado e precisado pelos direitos de explicação, revisão e contestação – o chamado devido processo informacional em decisões automatizadas por sistemas de IAs.

O Direito à Explicação, conforme o item acima indicado, ficou assim redigido no Projeto de Lei n.º 2.338/2023.<sup>263</sup>

Com a redação do projeto, a pesquisa aponta para que o Direito à Explicação ficará relegado somente as decisões automatizadas em Inteligências Artificiais de alto risco, limitando

<sup>260</sup> COECKELBERGH, Mark. **Ética na inteligência artificial**. Tradução: Clarisse de Souza, Edgar Lyra, Matheus Ferreira e Waldyr Delgado. 1.ª ed. São Paulo: Ubu Editora, 2024, p. 10.

<sup>261</sup> IBM. **Introduction to the AI Framework for z/OS**, 27 feb. 2025. Disponível em: <https://www.ibm.com/docs/en/zos/3.1.0?topic=zos-introduction-ai-framework>. Acesso em: 13 dez. 2024.

<sup>262</sup> Complementação do voto. BRASIL. Senado Federal. **Comissão Temporária Interna sobre Inteligência Artificial no Brasil**. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9852013&ts=1732829652162>. Acesso em: 13 dez. 2024.

<sup>263</sup> “Seção II - Dos Direitos da Pessoa e Grupos Afetados por Sistema de IA de Alto Risco. Art. 6º Pessoa ou grupo afetado por sistema de IA de alto risco tem os seguintes direitos: I - direito à explicação sobre a decisão, recomendação ou previsão feitas pelo sistema; II - direito de contestar e de solicitar a revisão de decisões, recomendações ou previsões de sistema de IA; e III - direito à revisão humana das decisões, levando-se em conta o contexto, risco e o estado da arte do desenvolvimento tecnológico. § 1º A explicação solicitada no âmbito do inciso I, respeitando o segredo comercial e industrial, incluirá informações suficientes, adequadas e inteligíveis, nos termos do Regulamento. § 2º Os direitos previstos nesta seção serão implementados considerando o estado da arte do desenvolvimento tecnológico, devendo o agente do sistema de IA sempre implementar medidas eficazes e proporcionais.” BRASIL. Senado Federal. **Comissão Temporária Interna sobre Inteligência Artificial no Brasil**. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9852013&ts=1732829652162>. Acesso em: 13 dez. 2024.

o exercício de um direito que deveria ser previsto para todo e qualquer indivíduo que entende ter sido prejudicado por uma decisão automatizada, evitando-se assim potencial desigualdade em discriminação em qualquer contexto, e não somente nas decisões de alto risco como disposto e enunciado na norma como pretendido.

Dessa forma, o presente capítulo da pesquisa demonstra a possibilidade concretização do Direito à Explicação no ordenamento jurídico, tornando os sistemas de Inteligência Artificiais mais confiáveis quando conseguimos analisar e dar respostas sobre as decisões automatizadas.

A regulação *ex ante* da governança em inteligência artificial (IA) desempenha um papel crucial na mitigação de vulnerabilidades, também conhecidos como os “*backdoors*”, que podem comprometer a segurança e a confiabilidade dos sistemas de IA. *Backdoors* são inserções maliciosas ou não intencionais que permitem acessos não autorizados ou manipulam o comportamento do sistema, representando riscos significativos em aplicações críticas.

Implementar uma regulação preventiva estabelece diretrizes claras para o desenvolvimento e a implementação de sistemas de IA, exigindo práticas rigorosas de segurança e auditoria desde as fases iniciais do projeto, incluindo – mas não se limitando – a realização de análises de impacto algorítmico e a adoção de *frameworks* de governança que promovam a transparência e a responsabilidade desde o desenvolvimento até disponibilização no mercado.

## 4.2 EXPLICABILIDADE – XAI

A explicabilidade em sistemas de inteligência artificial (IA), especialmente aqueles considerados como “caixas pretas”, é um tema de crescente importância na ciência da computação e em campos aplicados.

A regulação *ex ante* reforça a importância da explicabilidade nos sistemas de IA. A exigência de que os modelos forneçam explicações claras e compreensíveis sobre seu funcionamento e decisões não apenas aumenta a transparência, mas também facilita a detecção de comportamentos anômalos ou inserções maliciosas.

Como destacado por Nannini et al. (2023)<sup>264</sup>, políticas que promovem a explicabilidade contribuem para uma supervisão humana mais eficaz, permitindo que

---

<sup>264</sup> NANNINI, Luca; BALAYN, Agathe; SMITH, Adam Leon. Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK. **ArXiv preprint**, 2023. Disponível em: <https://arxiv.org/abs/2304.11218>. Acesso em: 5 nov. 2025.



pesquisadores e reguladores identifiquem e corrijam vulnerabilidades potenciais nos sistemas de IA.

No entanto, é fundamental que as regulamentações considerem as limitações técnicas atuais das metodologias de explicabilidade. Chung *et al.*<sup>265</sup> alertam que, embora a explicabilidade seja central nas regulamentações de IA, as técnicas disponíveis podem, por vezes, fornecer explicações errôneas ou incompletas. Portanto, uma abordagem regulatória equilibrada deve reconhecer essas limitações e incentivar o desenvolvimento contínuo de métodos mais robustos de explicação, garantindo que as exigências regulatórias sejam tecnicamente viáveis e eficazes na prática.

Porém, não é uma tarefa tão simples assim conferir explicabilidade para Inteligência Artificial, objeto da Teoria da Complexidade citada por Penrose:

No entanto, existem várias coisas gerais que são conhecidas, ou conjecturadas, com relação a certas limitações *absolutas* do quanto a velocidade de um algoritmo pode aumentar. Acontece que, mesmo entre os problemas matemáticos que *são* algoritmos por natureza, existem alguns que são intrinsecamente muito mais difíceis de solucionar algoritmicamente que outros. Os problemas difíceis podem ser resolvidos somente por algoritmos muito vagarosos (ou talvez, por algoritmos que necessitem de um imenso espaço de armazenamento de dados etc.). A teoria que trata de questões desse tipo é conhecida *teoria da complexidade computacional*. A teoria da complexidade não se preocupa com a solução de problemas *particulares* de forma algorítmica, mas com famílias infinitas de problemas em que haveria um algoritmo geral para encontrar a resposta a todos os problemas de uma mesma família. Os problemas diferentes na família teriam “tamanhos” diferentes, nos quais o tamanho de um problema é medido por algum número natural *n*.<sup>266</sup>

Em virtude da enorme quantidade de dados (pessoais, corporativos, institucionais) muitos são os problemas em termos de “tamanho” em virtude ds atividades que são processadas, aprendidas e que se tornam decisões de máquina.

No que se refere a base dados, reflete Pasquale:

Sobre (e além de) Dados. Mesmo com esse compromisso, não podemos esquecer que o acesso aos dados é apenas o primeiro e menor passo em direção à equidade em um mundo de pontuação digital onipresente, onde muitas de nossas atividades diárias são processadas como "sinais" para recompensas ou penalidades, benefícios ou ônus. Decisões críticas não são tomadas com base nos dados em si, mas com base nos dados analisados *algoritmicamente*: ou

<sup>265</sup> CHUNG, Neo Christopher *et al.* False Sense of Security in Explainable Artificial Intelligence (XAI). **ArXiv preprint**, 2024. Disponível em: <https://arxiv.org/abs/2405.03820>. Acesso em: 5 nov. 2024.

<sup>266</sup> PENROSE, Roger. **A mente do novo imperador: Sobre computadores, mentes e as leis da física**. Tradução: Gabriel Cozzella. São Paulo: UNESP, 2023, p. 210.

seja, em cálculos codificados em software de computador. Sem uma compreensão clara dos algoritmos envolvidos – e o direito de contestar aqueles injustos – a divulgação. Dos dados subjacentes fará pouco para garantir a justiça reputacional. Aqui está um conceito familiar das promessas e armadilhas de um mundo "pontuado".<sup>267</sup>

Explicar um sistema requer o atendimento de algumas condições e atributos desse sistema, como segurança, responsabilidade, estabilidade, transparência.

Explicabilidade está relacionada com interpretabilidade de um sistema, mas como citam Russel e Norvig:

**Interpretabilidade:** dizemos que um modelo de aprendizado de máquina é **interpretável** se você puder inspecionar o modelo real e entender por que ele obteve uma resposta específica para determinada entrada e como a resposta mudaria quando a entrada muda.<sup>268</sup> (grifo nosso).

A interpretabilidade dos sistemas de IA, e suas decisões ou saídas, é um desafio para as XAI, como ponderado abaixo por Kaufaman, Junquillo e Reais:

Adicionalmente, a complexidade das correlações estabelecidas pelos algoritmos de IA nos dados transcende a capacidade cognitiva dos seres humanos, configurando o “problema da interpretabilidade” (ou opacidade, ou *black-box*).<sup>269</sup>

Com redes neurais profundas, uma questão que levanta preocupações significativas sobre confiança, responsabilidade e ética, fundada na complexidade algorítmica como maneira de não expor que ela não é totalmente isenta e não igualitária.

Em outras palavras, os sistemas computacionais, podem ser compreendidos por meio de duas facetas: seu desenho técnico e sua mecânica (como funcionam); seus objetivos, ou seja, a correlação entre entradas e saídas, que se estabelece dentro de um determinado contexto social (por que opera).

---

<sup>267</sup> “On (and beyond) Data Even with that commitment, we can’t forget that access to data is just the first and smallest step toward fairness in a world of pervasive digital scoring, where many of our daily activities are processed as “signals” for rewards or penalties, benefits or burdens. Critical decisions are made not on the basis of the data per se, but on the basis of data analysed *algorithmically*: that is, in calculations coded in computer software. Failing clear understanding of the algorithms involved – and the right to challenge unfair ones – disclosure of underlying data will do little to secure reputational justice. Here a familiar concept from promise and pitfalls of a “scored” world.” PASQUALE, Frank. **The Black Box Society. The Secret Algorithms That Control Money and Information**. Cambridge-MA: Harvard University Press, 2015, p. 21-22.

<sup>268</sup> RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022, p. 644.

<sup>269</sup> KAUFMAN, Dora. JUNQUILHO, Táina; REIS, Priscila. Externalidades Negativas da Inteligência Artificial: Conflitos entre limites da técnica e dos direitos humanos. **Revista de Direitos e de Garantias Fundamentais**, Vitória, v. 24, n. 3 p. 43-71, set./dez. 2023. Disponível em: <https://doi.org/10.18759/rdgf.v24i3.2198>. Acesso: 28 nov. 2024.

Ambos os aspectos são passíveis de escrutínio, não apenas pelo pessoal técnico especializado, mas também por equipes jurídicas, e regulatórias que desejam compreender as decisões particulares e as dinâmicas de poder envolvidas na criação e operação dos sistemas.<sup>270</sup>

Conceituando o termo explicabilidade, Russel e Norvig descrevem-na da seguinte maneira:

**Explicabilidade:** um modelo explicável é aquele que pode ajudar você a entender “*por que* esta saída foi produzida para esta entrada”. Em nossa terminologia, a interpretabilidade deriva da inspeção do modelo real, enquanto a explicabilidade pode ser fornecida por um processo separado. Ou seja, o próprio modelo pode ser uma caixa-preta difícil de entender, mas um módulo de explicação pode resumir o que o modelo faz. [...] As explicações são uma forma de criar confiança, e algumas regulamentações, como o GDPR (Regulamento Geral de Proteção de Dados) europeu, exigem que os sistemas forneçam explicações. Em outras palavras, os sistemas computacionais, podem ser compreendidos por meio de duas facetas: seu desenho técnico e sua mecânica (como funcionam); seus objetivos, ou seja, a correlação entre entradas e saídas, que se estabelece dentro de um determinado contexto social (por que opera). Ambos os aspectos são passíveis de escrutínio, não apenas pelo pessoal técnico especializado, mas também por equipes jurídicas, e regulatórias que desejam compreender as decisões particulares e as dinâmicas de poder envolvidas na criação e operação dos sistemas.<sup>271</sup> (grifo nosso).

Explicabilidade de sistemas de IA é, portanto, à capacidade de explicar de maneira compreensível as decisões ou ações tomadas por um modelo de IA, ou seja, tem na transparência seu nexo relacional de duas partes.

Pode a Inteligência Artificial ser dotada de Explicabilidade (pelo menos aparente e em nível mais superficial) quando se considera que um determinado número e tipo de usuários disponibilizam uma determinada informação ou razão que tornem seu funcionamento (da IA) claro de entendimento.<sup>272</sup>

Abrusio discorre sobre o Princípio da Inteligência Artificial Explicável:

Existe uma reticência, por parte dos seres humanos, em adotar técnicas que não sejam interpretáveis e compreensíveis a si mesmos. É particularmente difícil confirmar naquilo que não se pode compreender. Em meio a esse cenário de discussões, surge a proposta de uma inteligência artificial explicável, cujo assunto tem ganhado cada vez mais força. Também identificado pela sigla XAI (*Explainable Artificial intelligence*). [...] Em outras palavras, pelo princípio da inteligência artificial explicável, a componente da explicabilidade deve estar presente nos sistemas inteligentes, gerando confiança ao usuário. Isso porque para além de serem *explicáveis*, é essencial

<sup>270</sup> FEFERBAUM, Marina *et al.* **Ética, Governança e Inteligência Artificial**. São Paulo: Almedina, 2023, p. 94.

<sup>271</sup> *Ibidem*, p. 94.

<sup>272</sup> RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 180.

que os sistemas de inteligência artificial, especialmente aqueles que envolvem alto risco, sejam confiáveis. [...] Note-se que a componente de explicabilidade não implica em ter conhecimento técnicos de programação algorítmica. É dizer, o modelo XAI é pensado e voltado ao usuário comum.<sup>273</sup>

Riscos, danos e efetivos ilícitos podem ser mitigados por meio de mecanismos de *Compliance*, Governança e *Accountability*, cujos relatórios, documentos e boas práticas podem certamente auxiliar no dever de transparência e informação, tornando a Inteligência Artificial Explicável.

Não somente para usuários pessoas naturais, mas também para todas as relações a que se destinam, evitando problemas colaterais das relações entre instituições de mercado e Estado, conferindo, assim, a possibilidade de mitigar problemas concorrenciais, de insumos, de Abuso de Poder Concorrencial e Econômico, uma vez que os dados não são somente pessoais, os agentes de mercado e o Estado transferem dados e são usuários de sistemas computacionais inteligentes também em seus negócios e perante seu ecossistema de relações jurídicas.

As palavras de Chiarello, Rocha e Moraes Rosa, “As preocupações acerca do controle, auditoria, verificação, interferência e modificação de um agente inteligente permeiam todos os imperativos éticos, todavia, encontram no componente “Explicabilidade” seu principal defensor”, pois é a Explicabilidade que permitirá entender a Inteligência Artificial em seu “estado da arte”, bem como se as medidas do Capítulo anterior são suficientes para entender decisões automatizadas ou parte delas, mesmo sabendo-se que mitigando riscos, e certamente afirmando e promovendo o desenvolvimento e uso ético da IA como premissa de “IA for Good”, reforçando dessa forma princípios como o da autonomia, não-maleficência, confiança, transparência etc.

Chiarello, Rocha e Moraes Rosa:<sup>274</sup>

Tidos como “imperativos éticos” - i) Respeito da autonomia humana ii) Prevenção de danos iii) Equidade iv) Explicabilidade<sup>20</sup> - estes princípios não possuem uma hierarquia entre si, o que, inclusive, pode gerar conflitos entre eles. As preocupações acerca do controle, auditoria, verificação, interferência e modificação de um agente inteligente permeiam todos os imperativos éticos, todavia, encontram no componente “Explicabilidade” seu principal defensor. É o imperativo ético cujo objetivo mais se alinha com o intuito de promover a adoção já que prevê que “a entidade responsável pela decisão deve ser identificável e os processos decisórios explicáveis”. A explicabilidade é

<sup>273</sup> RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 175.

<sup>274</sup> PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha; ROSA, Alexandre de Moares. Direito para Humanos Robotizados ou Direito dos Robôs Humanizados? **Revista Novos Estudos Jurídicos**, Itajaí, v. 28, n. 3, p. 544, 2023. Disponível em: <https://periodicos.univali.br/index.php/nej/article/view/19985>. Acesso em: 24 abr. 2024.

crucial para criar e manter a confiança dos utilizadores nos sistemas de IA. Tal significa que os processos têm de ser transparentes, as capacidades e a finalidade dos sistemas de IA abertamente comunicadas e as decisões — tanto quanto possível — explicáveis aos que são por elas afetados de forma direta e indireta. Sem essas informações, não é possível contestar devidamente uma decisão. Nem sempre é possível explicar por que razão um modelo gerou determinado resultado ou decisão (e que combinação de fatores de entrada contribuiu para esse efeito). Estes casos são designados por algoritmos de «caixa negra» e exigem especial atenção. Nessas circunstâncias, podem ser necessárias outras medidas da explicabilidade (p. ex., a rastreabilidade, a auditabilidade e a comunicação transparente sobre as capacidades do sistema), desde que o sistema, no seu conjunto, respeite os direitos fundamentais. O grau de necessidade da explicabilidade depende em grande medida do contexto e da gravidade das consequências de um resultado errado ou inexato.<sup>275</sup>

Explicabilidade tem relação com transparência e acesso a informações sobre como os dados e informações sobre inseridos, com que propósito, mas também como determinado algoritmo foi desenvolvido, seu propósito, monitoramento, ajustes, atualizações, e se um agente econômico possui ferramentas como as anteriormente propostas no capítulo referente a Governança da Inteligência Artificial poderá conferir a explicabilidade necessária.

Os sistemas de inteligência artificial, especialmente ao lidar com fontes de captação de dados, e diversas correspondências subsequentes precisam incluir conceitos de proteção de dados desde sua origem.<sup>276</sup>

Essa origem pode ser a transferência de um Agente de Tratamento de Dados Controlador para um Agente de Tratamento de Dados Operador, e vice-versa, por isso a ideia de amplitude da concretização do Princípio da Explicabilidade da IA (XAI) acima, reforçando os processos de Governança de Dados e *Compliance* Digital – agregados aos relatórios e auditorias algorítmicas, pois o *Privacy by Design* é uma das ferramentas de aplicação da Explicabilidade inerente aos dados.

Desta forma, a empresa atenderá, ainda, o importante princípio e vetor para o exercício da sua atividade econômica, traduzido pela sua função social, a qual para além de norma interpretativa e integrativa, implica em deveres positivos que orientam os interesses dos diversos sujeitos afetados pela empresa, como é o caso dos consumidores e dos titulares de dados de dados usuários dos sistemas, bem como consolida os princípios constitucionais da valorização do trabalho e da livre iniciativa. Não à toa, a função da empresa é imbuída de função sistematizadora do ordenamento jurídico, compatibilizando os vários

<sup>275</sup> COMISSÃO EUROPEIA. Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial (GPAN IA). **Orientações éticas para uma IA de confiança**, jun. 2018, p. 14. Disponível em: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60435](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60435). Acesso em: 29 nov. 2024.

<sup>276</sup> RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 177.

interesses envolvidos na atividade econômica, sem perder de vista a preservação da empresa e da atividade lucrativa que assim a qualifica.<sup>277</sup>

Concebe-se um ideal de existência e consolidação da Inteligência Artificial cujos efeitos possam ser intensificados com maior segurança jurídica e transparência para todos os envolvidos no sistema social, abaixo apresentamos entendimento que se encontra alinhado com esse pesquisador:

O modelo de inteligência artificial explicável implica, ainda, em desenvolver e manter a adequada governança, com a criação e aplicação constante de mecanismos de erros e inadequações sistêmicas, como vieses discriminatórios, com recorrentes avaliações de conformidades, incluindo auditorias como meio de efetivo a elevar o padrão ético e leal das tecnologias de inteligência artificial.<sup>278</sup>

A necessidade de explicabilidade surge não apenas por uma questão de eficiência operacional, mas também para atender aos requisitos legais e éticos.

Os modelos de caixa preta, particularmente aqueles baseados em aprendizado de máquina profundo, são notórios por sua falta de transparência. Eles operam através de camadas complexas de algoritmos que podem dificultar a interpretação de suas decisões. A explicabilidade em sistemas de inteligência artificial (IA), especialmente aqueles considerados como "caixas pretas", é um tema de crescente importância na ciência da computação e em campos aplicados.

A explicabilidade pode resolver o efeito Black box apontado por Pasquale quando conclui seu livro e descreve no último parágrafo:

Os serviços de caixa-preta são frequentemente maravilhosos de se ver. Mas nossa sociedade de caixa-preta tornou-se perigosamente instável, injusta e improdutiva. Nem os *quants* de Nova York nem os engenheiros da Califórnia podem entregar uma economia sólida ou uma sociedade segura. Essas são as tarefas de uma cidadania, que só pode desempenhar seu papel tão bem quanto entende os riscos.<sup>279</sup>

<sup>277</sup> RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA.** Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 179.

<sup>278</sup> *Ibidem*, p. 180.

<sup>279</sup> "Black box services are often wondrous to behold. But our black box society has become dangerously unstable, unfair, and unproductive. Neither New York quants nor California engineers can deliver a sound economy or a secure society. Those are the tasks of a citizenry, which can perform its job only as well it understands the stakes." PASQUALE, Frank. **The Black Box Society. The Secret Algorithms That Control Money and Information.** Cambridge-MA: Harvard University Press, 2015, p. 218.

A natureza talvez pouco transparente de muitos modelos de IA, como redes neurais profundas, levanta preocupações significativas sobre confiança, responsabilidade e ética. Este texto abordará o conceito de explicabilidade, os desafios associados e as abordagens atuais para tornar esses sistemas mais transparentes e compreensíveis.

Rudin afirma que: "As decisões humanas na construção de modelos de IA, desde a escolha de variáveis até a definição de métricas de performance, moldam diretamente como estes sistemas irão se comportar em produção".<sup>280</sup>

A explicabilidade de sistemas de IA refere-se à capacidade de explicar de maneira compreensível as decisões ou ações tomadas por um modelo de IA.

Isso é especialmente crítico em áreas como medicina, finanças e justiça, onde decisões erradas podem ter consequências graves. Os modelos de caixa preta, particularmente aqueles baseados em aprendizado de máquina profundo, são notórios por sua falta de transparência. Eles operam através de camadas complexas de algoritmos que podem dificultar a interpretação de suas decisões.

Isso pode criar uma barreira onde os usuários e desenvolvedores não conseguem entender completamente os mecanismos internos dos modelos, o que pode levar a um ceticismo quanto à sua confiabilidade.

Para combater a opacidade das caixas pretas, várias abordagens têm sido exploradas, como: (i) Métodos de Aprendizado de Máquina Interpretável: técnicas como LIME (*Local Interpretable Model-agnostic Explanations*) permitem que os usuários vejam quais entradas influenciaram a saída de um modelo, oferecendo uma forma de explicabilidade "local"; (ii) Ferramentas de Visualização: cujas aplicações e ferramentas que visualizam o processamento interno dos modelos de IA, como *TensorBoard* para redes neurais *TensorFlow*, ajudam os pesquisadores a entender melhor o comportamento do modelo; (iii) Regulações e Diretrizes: as quais Organizações e Governos estão começando a implementar diretrizes que exigem que sistemas de IA sejam acompanhados de explicações compreensíveis.

A União Europeia, por exemplo, incluiu a explicabilidade como um dos pilares do seu regulamento geral de proteção de dados (GDPR).

A aplicação de técnicas de explicabilidade já mostrou resultados promissores em vários setores. Em saúde, sistemas explicáveis podem ajudar médicos a entenderem as

---

<sup>280</sup> RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, v. 1, p. 208, 2019. Disponível em: <https://www.nature.com/articles/s42256-019-0048-x>. Acesso em: 24 abr. 2024.

recomendações de tratamento baseadas em IA potencialmente aumentando a aceitação de tais sistemas no diagnóstico e tratamento.<sup>281</sup>

A busca pela explicabilidade em sistemas de IA de caixa preta é fundamental para garantir que essas tecnologias sejam utilizadas de maneira responsável e ética.

Embora existam desafios significativos, os avanços nas técnicas de interpretabilidade e as regulamentações emergentes estão ajudando a pavimentar o caminho para sistemas de IA mais transparentes e confiáveis.

Importa para a pesquisa considerar que, em 14 de março de 2023, a *OpenAI* divulgou a atualização do *ChatGPT* com nova versão do seu algoritmo de inteligência artificial generativa – o GPT-4.

No documento técnico oficial, a empresa revelou diversos malefícios e riscos da tecnologia: (i) que esse novo algoritmo não é totalmente confiável, pois sofre de “alucinações digitais”; (ii) gerar conteúdo potencialmente prejudicial; (iii) reproduzir vários preconceitos e visões de mundo discriminatórias; iv) fornecer informações detalhadas sobre como conduzir atividades ilegais, incluindo o desenvolvimento de armas biológicas. Um trecho crucial deste documento técnico da OpenAI diz que o GPT-4 e sistemas de IA semelhantes terão um potencial ainda maior para reforçar ideologias, visões de mundo, verdades e inverdades para bloqueá-las, impedindo futuras contestações, reflexões e melhorias.<sup>282</sup>

Isso torna a Generativa um desafio para o conceito de explicabilidade, dada sua capacidade de aprendizado profundo, aliada à interação humana que possui vieses cognitivos.

As LLMs (Large Language Models), com sua capacidade de compreender e gerar linguagem humana, têm um papel importante na explicação de decisões automatizadas de inteligências artificiais, tornando-as mais transparentes e compreensíveis para os usuários.

Ao utilizar LLMs para gerar explicações claras, concisas e personalizadas, podemos aumentar a transparência e a confiança nas decisões automatizadas, promovendo uma IA mais responsável e ética. É crucial integrar as LLMs de forma estratégica, considerando as necessidades dos usuários e as características específicas do sistema de IA em questão.

Os métodos de avaliação da explicabilidade em sistemas de inteligência artificial são essenciais para garantir a transparência e confiabilidade desses sistemas. Através da avaliação, é possível verificar se as explicações fornecidas são compreensíveis e satisfatórias para os

---

<sup>281</sup> HOLZINGER, Andreas *et al.* **What do we need to build explainable AI systems for the medical domain?**, 28 dec. 2017. Available at: <https://arxiv.org/pdf/1712.09923>. Access in: 24 abr. 2024.

<sup>282</sup> OPENAI. **GPT-4 Technical Report**, 27 mar. 2023, p. 49. Disponível em: <https://cdn.openai.com/papers/gpt-4.pdf>. Acesso em: 27 mar. 2024.



usuários, permitindo uma análise crítica de como a explicabilidade está sendo implementada na prática e identificando possíveis melhorias. Além disso, os métodos de avaliação também auxiliam na comparação e escolha entre diferentes abordagens de explicabilidade, contribuindo para o avanço da área.

As métricas e critérios de avaliação da explicabilidade em sistemas de inteligência artificial são fundamentais para medir e avaliar a qualidade e eficácia das explicações fornecidas. Estas métricas e critérios podem incluir aspectos como clareza, consistência, precisão e utilidade das explicações, possibilitando uma avaliação objetiva do desempenho dos sistemas de IA em relação à explicabilidade.

A definição de métricas e critérios claros e bem estabelecidos é crucial para garantir a comparabilidade e replicabilidade dos resultados, permitindo a evolução e aprimoramento contínuo das práticas de explicabilidade, aponta a pesquisa com exemplo a ferramenta para análise da explicabilidade dos modelos de *machine learning* como o da Google.<sup>283</sup>

---

<sup>283</sup> GOOGLE CLOUD. **Introdução à Vertex Explainable AI**. Disponível em: <https://cloud.google.com/vertex-ai/docs/explainable-ai/overview?authuser=1&hl=pt-br>. Acesso em: 7 dez. 2024.

## CONCLUSÃO

A pesquisa realizada evidencia a complexidade e a importância crescente da do Direito à Explicação e da possibilidade da Explicabilidade da Inteligência Artificial (IA).

A pesquisa - dentre outros desafios, concluiu sobre: (i) o papel da responsabilidade exercida pelos humanos na decisão automatizada, na revisão e supervisão das decisões de I.A., discutindo os benefícios e riscos de delegar ou compartilhar a tomada de decisão com as máquinas; (ii) o uso de instrumentos de *compliance* e governança de Inteligência Artificial como método *ex ante* que permitem tornar decisões automatizadas mais transparentes e éticas; (iii) implicações éticas, jurídicas e sociais do Direito à Explicação e a possibilidade de Explicabilidade em Inteligência Artificial. As decisões humanas e comportamentos expressos nas relações tecnológicas são transformados em dados que integram e alteram a compreensão da IA sobre o mundo que a circunda, e, portanto, devem os usuários e seres humanos ponderar sobre a constante produção de dados de personalidade e decisão pessoal quando de suas interações tecnológicas.

As decisões automatizadas não podem ou devem conter obscuridades ou desconfianças, portanto, devem ser transparentes, confiáveis e explicáveis, alcançando o interesse público pautado nos princípios da legalidade, impessoalidade, moralidade, publicidade, eficiência, transparência, bem como Direitos Fundamentais Humanos.

Dados inconsistentes, que estejam mal alocados e que, portanto, são mal imputados em sistemas de Inteligência Artificial darão resultados equivocados, inconsistentes, possibilitando assim vieses decisórios que contém problemas éticos e/ou discriminatórios, e portanto poderão prejudicar cidadãos, agentes de mercado, organizações e instituições públicas, enfim, toda a sociedade.

Em um mundo cada vez mais dependente de sistemas automatizados, a transparência e a compreensibilidade das decisões automatizadas proferidas por sistemas de IA tornaram-se questões fundamentais que perpassam dimensões éticas, jurídicas, econômicas e sociais.

O *Compliance* Digital, segurança da informação, governança de dados e a governança de inteligência artificial emergem como ferramentas essenciais para mitigar riscos e promover confiança nos sistemas de IA. Tais abordagens não apenas buscam aumentar a transparência e a confiança no desenvolvimento e utilização da Inteligência Artificial, e também estabelecem um marco ético e responsável para a aplicação de tecnologias de inteligência artificial.

A Governança da Inteligência Artificial objetiva garantir a confiabilidade ética das decisões e resultados, evitando que os sistemas não estejam consonantes com legislações e

normas jurídico sociais, diminuindo a exposição a ataques e incidentes que possibilitem desconfiança, vieses discriminatórios, e não explicação dos resultados e decisões automatizadas.

Os métodos positivos e boas práticas da regulação *ex ante* pesquisadas visam sugerir (sem esgotar) formas de cumprir o objetivo de transparência, e apresentar documentnos e medidas de explicação sobre decisões resultantes de Inteligência Artificial, diminuindo afastando dúvidas, obscuridades, opacidades, vieses e discriminações, vez que aferíveis.

Parte integrante do Direito à Explicação é a possibilidade de desenvolver e utilizar sistemas de Inteligência Artificial Explicáveis – XAI, para que produzam averiguações e emitam documentos capazes de conferir como ocorreu o aprendizado e o resultado de uma Inteligência Artificial.

Como conclusão, a pesquisa e a dissertação demonstraram que é possível o Direito à Explicação das decisões automatizadas, verificando como – e se – uma Decisão Automatizada possui riscos economicos e sociais decorrentes de vieses, discriminações etc., que podem advir do desenvolvimento, do aprendizados, dos dados coletados e inseridos, do perfilamento e padronização e que influenciam os resultados dos sistemas de Inteligência Artificial.

A pesquisa também conclui que regulações *ex ante* e mecanismos positivos para desenvolvimento e utilização de Inteligência Artificial contribuem para o Direito à Explicação e para auxiliar a Explicabilidade.

Problemas com aprendizado e resultados podem ser mitigados e revisados, não somente para os sistemas de alto risco, mas para todo e qualquer sistema que possa prejudicar um indivíduo ou grupo de indivíduos, bem como a Sociedade e o próprio mercado quando utilizam-se de dados para Poder Dominante que são inseridos em Inteligências Artificiais que produzem resultados que formam a base de decisões organizacionais.

A problemática das "caixas pretas" não pode ser subestimada, especialmente quando decisões automatizadas impactam diretamente a vida das pessoas em áreas críticas como medicina, direito, segurança e economia de mercado, porém a pesquisa demonstra que existem pesquisas e instrumentos que auxiliam na explicação, e na explicabilidade, dos sistemas de Inteligência Artificial.

A pesquisa aponta que temos instrumento para exercício do Direito à Explicação, não é apenas uma questão técnica, mas um imperativo ético para conter eventuais discriminações ou má utilização da IA.

Podemos obter maior grau de explicação de decisões automatizadas quando temos controles preventivos, boas práticas e implementação de programas (*Frameworks*) governança

para utilização dessas ferramentas computacionais, iniciando tal tarefa quando o uso de dados coletados da vida privada, dados pessoais de identificação, e dados corporativos que subsidiam as decisões não constituam máculas aos Direitos Fundamentais, sociais e econômicos, cada vez mais utilizados no perfilamento individual, de grupos e de empresas, e que predeterminam os resultados exarados (decisões automatizadas, *Outputs*) pela Inteligência Artificial.

Sistemas de IA verdadeiramente explicáveis e éticos apenas começaram, e requer um esforço contínuo e colaborativo de pesquisadores, desenvolvedores, legisladores e sociedade civil.

A Inteligência Artificial Explicável (XAI) representa um avanço significativo nesse contexto, oferecendo uma ponte crucial entre a complexidade técnica dos sistemas de IA e a necessidade humana de compreensão.

A busca pela explicabilidade em sistemas de IA de caixa preta é fundamental para garantir que essas tecnologias sejam utilizadas de maneira responsável e ética. Embora existam desafios significativos, os avanços nas técnicas de interpretabilidade e as regulamentações emergentes estão ajudando a pavimentar o caminho para sistemas de IA mais transparentes e confiáveis.

A pesquisa demonstra que a Explicabilidade é possível, e que mecanismos de governança, regulação e tecnologia XAI permitirão entender as decisões da Inteligência Artificial.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABRUSIO, Juliana. O princípio da Inteligência Artificial Explicável e a Proteção de Dados. *In: RAIS, Diogo; ABRUSIO, Juliana. Privacidade, mercado e cidadania: uma conexão a partir da IA*. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022, p. 169-180.
- ABRUSIO, Juliana. **Proteção de Dados na Cultura do Algoritmo**. 1.<sup>a</sup> ed. São Paulo: D'Plácido, 2020. (Coleção Direito privado, v. 1).
- ALPAYDIN, Ethem. **Introduction to Machine Learning**. 4<sup>th</sup> ed. Cambridge: MIT Press, 2020.
- AMOROSO, Fabrício Steinle. **Inteligência artificial explicável com LIME e SHAP aplicada à rede neural convolucional**. 2023. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Faculdade de Ciências, Universidade Estadual Paulista (UNESP), Bauru, 2023. Disponível em: <http://hdl.handle.net/11449/239119>. Acesso em: 24 abr. 2024.
- ANDERSON, David *et al.* **Ethics & Algorithms Toolkit**. Available at: <https://ethicstoolkit.ai/>. Access in: 7 dez. 2024.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 42001:2024 - Tecnologia da informação — Inteligência artificial — Sistema de gestão**, 18 abr. 2024. Disponível em: <https://www.abntcatalogo.com.br/pnm.aspx?Q=eXkvMUMzazlVVWZzRjN1eEJxQVpoOU05WkVaamF6ZDIBVnRIem5zSDBkRT0=>. Acesso em: 7 dez. 2024.
- BAGNOLI, Vicente. **Direito econômico e concorrencial – O Poder Econômico e seus Limites Jurídicos**. 9.<sup>a</sup> ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2022.
- BAGNOLI, Vicente; CRISTÓFARO, Pedro Paulo Salles. **Jurisprudência do CADE Comentada**. São Paulo: Revista dos Tribunais, 2019.
- BARBOSA, Mafalda Miranda *et al.* **Direito Digital e Inteligência Artificial: Diálogos entre Brasil e Europa**. São Paulo: Foco, 2021.
- BARBOSA, Nathalia Sartarello. O Reflexo da sociedade do hiperconsumo no Instagram e a Responsabilidade Civil dos Influenciadores Digitais. **Revista de Direitos Culturais**, Santo Ângelo, v. 13, n. 30, p. 73-88, mai./ago. 2018. Disponível em: [https://www.researchgate.net/publication/328376203\\_O\\_REFLEXO\\_DA\\_SOCIEDADE\\_DO\\_HIPERCONSUMO\\_NO\\_INSTAGRAM\\_E\\_A\\_RESPONSABILIDADE\\_CIVIL\\_DOS\\_INFLUENCIADORES\\_DIGITAIS](https://www.researchgate.net/publication/328376203_O_REFLEXO_DA_SOCIEDADE_DO_HIPERCONSUMO_NO_INSTAGRAM_E_A_RESPONSABILIDADE_CIVIL_DOS_INFLUENCIADORES_DIGITAIS). Acesso em: 29 abr. 2024.
- BAROCAS, Solon; SELBST, Andrew D. Big Data's Disparate Impact. **California Law Review**, v. 104, n. 671, p. 671-732, 2016. Available at: <https://ssrn.com/abstract=2477899> or <http://dx.doi.org/10.2139/ssrn.2477899>. Access in: 29 abr. 2024.
- BARRETO JÚNIOR, Irineu Francisco; COSTA, Bruna Marangoni Brancaleone. Hiperexposição Pessoal nas Redes Sociais e seus Reflexos no Direito. **Revista de Direito, Governança e Novas Tecnologias**, Salvador, v. 4, n. 1, p. 92–108, jan./jun. 2018. Disponível

em: <https://www.indexlaw.org/index.php/revistadgnt/article/view/4299/pdf>. Acesso em: 24 abr. 2024.

BARRETO JÚNIOR, Irineu Francisco; MOLINA, Fernanda Zampieri. Capitalismo de plataforma: a ameaça ao direito à autodeterminação informativa na Sociedade da Informação. **Revista Brasileira de Estudos Políticos**, Belo Horizonte, n. 125, p. 243-278, jul./dez. 2022. Disponível em: <https://pos.direito.ufmg.br/rbep/index.php/rbep/article/view/852/657>. Acesso em: 24 abr. 2024.

BARRETO JÚNIOR, Irineu Francisco; NASCIMENTO, Ariane Azevedo Carvalho do; FULLER, Greice Patrícia. Lei Geral de Proteção de Dados Pessoais: Efetividade Jurídica do Consentimento do Titular para Tratamento dos Registros. **Revista de Constitucionalização do Direito Brasileiro – RECONTO**, v. 3, n. 2, p. 1-23, jul./dez. 2020. Disponível em: <https://revistareconto.com.br/index.php/reconto/article/view/44/42>. Acesso em: 24 abr. 2024.

BARRETO JÚNIOR, Irineu Francisco; UELZE, Hugo Barroso. Cibercultura e o Poder Informacional na Esfera Pública: Crise do Paradigma Positivista no Direito Tributário. **Revista Prima Facie**, Universidade Federal da Paraíba, João Pessoa, v. 19, n. 40, p. 1-29, 2020. Disponível em: <https://periodicos.ufpb.br/index.php/primafacie/article/view/51047>. Acesso em: 24 abr. 2024.

BASTOS Alberto; CAUBIT, Rosângela. **Gestão de Segurança da Informação. SAI 27001 e 27002 - Uma Visão Prática**. Rio Grande do Sul: Zouk, 2009.

BECHARA, Antoine; DAMASIO, Antonio R. The somatic marker hypothesis: Twenty years later. **Nature Reviews Neuroscience**, v. 22, n. 4, p. 273-289, 2021.

BECHARA, Antoine; DAMASIO, Hanna; DAMASIO, Antonio R. Emotion, Decision Making and the Orbitofrontal Cortex. **Cereb Cortex**, v. 10, n. 3, p. 295-307, mar. 2000. Available at: DOI: 10.1093/cercor/10.3.295. Access in: 24 apr. 2024.

BENSOUSSAN, Fábio Guimarães; GOUVÊA, Marcus de Freitas; LUZ, Reinaldo Diogo. **Manual de Direito Econômico**. 7.<sup>a</sup> ed. rev., atual. e ampl. São Paulo: Juspodivm, 2023.

BENT, Jason R. **Is algorithmic affirmative action legal? The Georgetown Law Journal**, v. 108, p. 803-853, 2019. Disponível em: <https://www.law.georgetown.edu/georgetown-law-journal/wp-content/uploads/sites/26/2020/04/Is-Algorithmic-Affirmative-Action-Legal.pdf>. Acesso em: 1.º dez. 2024.

BIONI, Bruno Ricardo. **Proteção de Dados Pessoais – A Função e os Limites do Consentimento**. 3.<sup>a</sup> ed. São Paulo: Forense, 2022.

BIONI, Bruno Ricardo. **Regulação e Proteção de Dados Pessoais – O Princípio da Accountability**. 1.<sup>a</sup> ed. São Paulo: Forense, 2022.

BLUM, Renato Opice; MALDONADO, Viviane Nóbrega. **Comentários a GDPR, Regulamento Geral de Proteção de Dados da União Europeia**. São Paulo: Thomson Reuters, 2018.

BRASIL. Câmara dos Deputados. Comissão de Constituição e Justiça e de Cidadania. **Parecer acerca da Proposta de Emenda à Constituição nº 17/2019**, p. 1. Disponível em: [https://www.camara.leg.br/proposicoesWeb/prop\\_mostrarintegra;jsessionid=node01vqwnuiffbkdp1v8m3yk5aotpe37875732.node0?codteor=1791032&filename=Tramitacao-PEC+17/2019](https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=node01vqwnuiffbkdp1v8m3yk5aotpe37875732.node0?codteor=1791032&filename=Tramitacao-PEC+17/2019). Acesso em: 7 dez. 2024.

BRASIL. Câmara dos Deputados. **Projeto de Lei n.º 2.338 de 2023**. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco. Brasília-DF: Senado Federal, 2023. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1742240889313&disposition=inline>. Acesso em: 27 nov. 2023.

BRASIL. **Lei n.º 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília-DF: Presidência da República, 2025. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/L13709compilado.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm). Acesso em: 6 abr. 2025.

BRASIL. Ministério das Relações Exteriores. **Declaração de Osaka dos Líderes do G20**, 7 jan. 2021. Disponível em: [https://www.gov.br/mre/pt-br/canais\\_atendimento/imprensa/notas-a-imprensa/2019/declaracao-de-osaka-dos-lideres-do-g20](https://www.gov.br/mre/pt-br/canais_atendimento/imprensa/notas-a-imprensa/2019/declaracao-de-osaka-dos-lideres-do-g20). Acesso em: 10 nov. 2024.

BRASIL. Senado Federal. **Comissão Temporária Interna sobre Inteligência Artificial no Brasil**. Disponível em: <https://legis.senado.leg.br/sdleg-getter/documento?dm=9852013&ts=1732829652162>. Acesso em: 13 dez. 2024.

BRENNAN, Karen; RESNICK, Mitchel. **New Frameworks for Studying and Assessing the Development of Computational Thinking**. American Educational Research Association, 2024, p. 1-25. Available at: <https://scratched.gse.harvard.edu/ct/files/AERA2012.pdf>. Access in: 24 apr. 2024.

BRIDGER, Darren. **Neuromarketing: como a neurociência aliada ao design pode aumentar o engajamento e a influência sobre os consumidores**. Tradução: Afonso Celso da Cunha Serra. 1.ª ed., 2.ª reimp. São Paulo: Autêntica Business, 2019.

BUCCI, Eugênio. Reflexões sobre ética & inteligência maquina. *In: SANTAELLA, Lúcia. Simbioses do Humano & Tecnologias: Impasses, Dilemas, Desafios*. São Paulo: Editora Universidade de São Paulo/IEA-USP, 2022, p. 91-102.

BUSINESS SOFTWARE ALLIANCE. **Business Software Alliance Transition Priorities**. Available at: <https://www.bsa.org/pt>. Acesso em: 7 dec. 2024.

CANTARINI, Paola. Inteligência Artificial: Desafios Regulatórios e Risquificação. *In: PINHO, Anna Carolina et al. Discussões sobre Direito na Era Digital*. 1.ª ed. Rio de Janeiro: GZ, 2021, p. 548.

CÉSAR, Ana Maria Roux Valentini; BOGGIO, Paulo; CAMPANHÃ, Camila. **Neuroeconomia: uma visão geral sobre o tema**. Disponível em: <https://www.researchgate.net/publication/307575716>. Acesso em: 11 nov. 2024.

CHUNG, Neo Christopher *et al.* False Sense of Security in Explainable Artificial Intelligence (XAI). **ArXiv preprint**, 2024. Disponível em: <https://arxiv.org/abs/2405.03820>. Acesso em: 5 nov. 2024.

COECKELBERGH, Mark. **Ética na inteligência artificial**. Tradução: Clarisse de Souza, Edgar Lyra, Matheus Ferreira e Waldyr Delgado. 1.<sup>a</sup> ed. São Paulo: Ubu Editora, 2024.

COMISSÃO EUROPEIA. **Abordagem europeia da inteligência artificial**. Disponível em: <https://digital-strategy.ec.europa.eu/pt/policies/european-approach-artificial-intelligence>. Acesso em: 13 jun. 2024.

COMISSÃO EUROPEIA. Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial (GPAN IA). **Orientações éticas para uma IA de confiança**, jun. 2018, p. 1-55. Disponível em: [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60435](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60435). Acesso em: 29 nov. 2024.

COSTA, Bruna Marangoni Brancaleone; BARRETO JÚNIOR, Irineu Francisco. **Revista de Direito, Governança e Novas Tecnologias**, Salvador, v. 4, n.1, p. 92-108, jan./jun. 2018. Disponível em: <https://www.indexlaw.org/index.php/revistadgnt/article/view/4299/pdf>. Acesso em: 24 abr. 2024.

COUTINHO, Diego Araújo. **O consumidor e modulação algorítmica de comportamento: a influência da Inteligência Artificial por meio de algoritmos no poder decisório do consumidor**. São Paulo: Dialética, 2021.

CRAWFORD, Kate; CALO, Ryan. **There is a blind spot in AI research**. *Nature*, v. 538, n. 7625, p. 311-313, 2016.

DONEDA, Danilo Cesar Maganhoto. **Da Privacidade à Proteção de Dados Pessoais**. Rio de Janeiro: Renovar, 2006.

DONEDA, Danilo Cesar Maganhoto. **Da Privacidade à Proteção de Dados Pessoais: elementos da formação da nova cultura jurídica**. São Paulo: Revista dos Tribunais, 2017.

DONEDA, Danilo Cesar Maganhoto. O direito fundamental à proteção de dados pessoais. *In: SOUZA, Allan Rocha de et al. Direito Digital: Direito Privado e Internet*. 4.<sup>a</sup> ed. Indaiatuba: Foco, 2021, p. 33-49.

DONEDA, Danilo Cesar Maganhoto *et al.* Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar – Revista de Ciências Jurídicas**, Fortaleza, v. 23, n. 4, p. 1-17, 2018. Disponível em: <https://ojs.unifor.br/rpen/article/view/8257/pdf>. Acesso em: 24 abr. 2024.

DOURADO, Daniel de Araújo. Direito à Explicação e Regulação da Inteligência Artificial na Saúde: perspectiva da LGPD. *In: AITH, Fernando; DALLARI, Annaluz Bolivar. LGPD na Saúde Digital*. São Paulo: Thomson Reuters, 2022, p. 115-130.

EKUNI, Roberta; POMPÉIA, Sabine. O impacto da divulgação científica na perpetuação de neuromitos na educação. **Revista de Biologia**, v. 15, n. 1, p. 21-28, 2016. Disponível em: <https://www.revistas.usp.br/revbiologia/article/view/114837/112576>. Acesso em: 24 abr. 2024.



EUROPE. **General Data Protection Regulation**. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504>. Access in: 14 nov. 2022.

EUROPEAN ARTIFICIAL INTELLIGENCE ACT. The AI Act Explorer. **Official Journal**, 13 jun. 2024. Available at: <https://artificialintelligenceact.eu/ai-act-explorer/>. Access in: 13 dec. 2024.

EYSENCK, Michel W.; EYSENCK, Christine. **Inteligência Artificial x Humanos: o que a ciência cognitiva nos ensina ao colocar frente a frente a mente humana e a IA**. Tradução: Gisele Klein. Revisão Técnica: Vitor Geraldi Haase. Porto Alegre: Artmed, 2023.

FACCHINI NETO, Eugênio; SOARES, Flaviana Rampazzo. **Responsabilidade Civil pela Violação ao Dever de Proteção de Dados na LGPD**. In: PINHO, Anna Carolina *et al.* **Discussões sobre direito na era digital**. 1.<sup>a</sup> ed. Rio de Janeiro: GZ, 2021, p. 237.

FALEIROS JÚNIOR, José Luiz de Moura. **Administração Pública Digital: proposições para o aperfeiçoamento do Regime Jurídico Administrativo na sociedade da informação**. 2.<sup>a</sup> ed. Indaiatuba: Foco, 2024.

FEFERBAUM, Marina *et al.* **Ética, Governança e Inteligência Artificial**. São Paulo: Almedina, 2023.

FERFERBAUM, Marina; KLAFKE, Guilherme Forma; LATINI, Lucas Maldonado Diz. **Comitês de Ética para o desenvolvimento e implementação de IA confiável e responsável**. In: FACHINETTI, Aline Fuke. **Inteligência Artificial Responsável: programa de governança e melhores práticas organizacionais**. São Paulo: Thomson Reuters, 2024, p. 215-236.

FERRARI, Isabela; BECKER, Daniel. Direito à Explicação e Decisões Automatizadas: Reflexões sobre o Princípio do Contraditório. In: NUNES, Dierle; LUCON, Paulo Henrique dos Santos; WOLKART, Erik Navarro. **Inteligência Artificial e Direito Processual: Os Impactos da Virada Tecnológica no Direito Processual**. Salvador: Juspodivm, 2021, p. 277-303.

FERRAZ, Sérgio; VENTURINI, Otávio; GASIOLA, Gustavo Gil. **Proteção de Dados Pessoais e Compliance Digital**. Cuiabá: Umanos Editora, 2022.

FERREIRA, Marcelo; LOPES, Marcos. **Para Conhecer Linguística Computacional**. São Paulo: Contexto, 2023.

FIELDING, Roy Thomas. **Architectural Styles and the Design of Network-based Software Architectures**. Dissertação (Doutorado em Ciência da Computação) – University of California, Irvine, 2000, p. 1-24. Disponível em: <https://www.proquest.com/openview/fc2d064044b971dda476dfb429a2b344/1?pq-origsite=gscholar&cbl=18750&diss=y>. Acesso em: 2 dez. 2024.

FISHER, Carl Erik; CHIN, Lisa; KLITZMAN, Robert. Defining neuromarketing: Practices and professional challenges. **Nature Reviews Neuroscience**, v. 11, n. 4, p. 348, 2010.

FLORIDI, Luciano. **Accountability and the Right to Explanation in AI: A Multidimensional Approach**. Cambridge: Cambridge University Press, 2019.

FLORIDI, Luciano. Addressing Bias in Artificial Intelligence: A Holistic Ethical Framework. **Ethics and Information Technology**, v. 23, n. 1, p. 15-27, 2021.

FLORIDI, Luciano. Balancing Transparency and Privacy: The Right to Explanation in the Age of AI. **Journal of Digital Ethics**, v. 18, n. 2, p. 150-170, 2021.

FLORIDI, Luciano. **Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide**. Tradução: Massimo Durante. Milano – Itália: Raffaello Cortina Editore, 2022.

FLORIDI, Luciano. Ethical Implications of the Right to Explanation in AI Systems. **Journal of Ethics in Technology**, v. 25, n. 1, p. 45-60, 2020.

FLORIDI, Luciano. **Ethics of Artificial Intelligence**. Oxford: Oxford University Press, 2020.

FLORIDI, Luciano. **The Right to Explanation: Ethics, Transparency, and Accountability in AI Systems**. Oxford: Oxford University Press, 2018.

FRANÇA, Phillip Gil. **Ato Administrativo, Consequencialismo e Compliance: Gestão de Riscos, Proteção de Dados e Soluções para o Controle Judicial na Era da IA**. 4.<sup>a</sup> ed. São Paulo: Revista dos Tribunais, 2019.

FRAZÃO, Ana. **Fundamentos da proteção de dados pessoais – Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados**. In: FRAZÃO, Ana. TEPEDINO, Gustavo; OLIVA, Milena Donato. **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. 1.<sup>a</sup> ed. São Paulo: Thomson Reuters Brasil, 2019, p. 26-38.

FRAZÃO, Ana; CUEVA, Ricardo Villas Bôas Cueva. **Compliance e Políticas de Proteção de Dados**. São Paulo: Revista dos Tribunais, 2021.

FRAZÃO, Ana; GOETTENAUER, Carlos. O jogo da imitação jurídica: o direito à revisão de decisões algorítmicas como um mecanismo para a necessária conciliação entre linguagem natural e infraestrutura matemática. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo *et al.* **O Direito Civil na Era da Inteligência Artificial**. 1.<sup>a</sup> ed. São Paulo: Thomson Reuter, 2020, p. 45-63.

FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito – Ética, Regulação e Responsabilidade**. São Paulo: Revista dos Tribunais, 2019.

GABRIEL FILHO, Oscar. **Inteligência Artificial e aprendizagem de máquina: aspectos teóricos e aplicações**. São Paulo: Blucher, 2023.

GLIMCHER, Paul W. *et al.* Neuroeconomics: Decision making and the brain. **Economic Analysis & Policy**, v. 40, n. 1, p. 111-117, mar. 2010. Available at: file:///C:/Users/Samsung/Downloads/Neuroeconomics\_Decision\_making\_and\_The\_Brain.pdf. Access in: 24 abr. 2024.

GODOY, Cláudio Luiz Bueno de Godoy. A intervenção humana no procedimento de explicação e de revisão de decisões automatizadas. *In*: CHINELLATO, Silmara J. de A. *et al.* **Inteligência Artificial: Visões Interdisciplinares e Internacionais**. 1.<sup>a</sup> ed. São Paulo: Almedina Brasil, 2023, p. 286.

GOMES, Luiz Flávio Autran Monteiro; GOMES, Carlos Francisco Simões. **Princípios e Métodos da Tomada de Decisão**. 6.<sup>a</sup> ed. São Paulo: Atlas, 2019.

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

GOOGLE CLOUD. **Introdução à Vertex Explainable AI**. Disponível em: <https://cloud.google.com/vertex-ai/docs/explainable-ai/overview?authuser=1&hl=pt-br>. Acesso em: 7 dez. 2024.

GOULART, Guilherme Damásio. Comentários à Lei Geral de Proteção de Dados. *In*: MARTINS, Guilherme Magalhães; LONGHI, João Victor Rozatti; FALEIROS JÚNIOR, José Luiz de Moura. **Comentários à Lei Geral de Proteção de Dados**. 1.<sup>a</sup> ed. Indaiatuba: Foco, 2022, p. 263.

GROVER, Shuchi; PEA, Roy. Computational Thinking in K–12: A Review of the State of the Field. **Educational Researcher**, v. 42, n. 1, p. 38-43, 2024.

HAN, Byung-Chul. **Infocracia: digitalização e a crise da democracia**. Tradução: Gabriel S. Philipson. Petrópolis: Vozes, 2022.

HAN, Byung-Chul. **Sociedade do Cansaço**. Tradução: Ênio Paulo Giachini. 2.<sup>a</sup> ed. Petrópolis: Vozes, 2017.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2.<sup>a</sup> ed. New York: Springer, 2017.

HAYKIN, Simon. **Redes Neurais: princípios e prática**. Tradução: Paulo Martins Engel. 2.<sup>a</sup> ed. Porto Alegre: Bookman, 2001.

HINTEZBERGEN, Jule *et al.* **Fundamentos da Segurança da Informação: com base na ISSO 27001 e na ISSO 27002**. Tradução: Alan de Sá. Rio de Janeiro: Brasport, 2018, p. 27.

HOLZINGER, Andreas *et al.* **What do we need to build explainable AI systems for the medical domain?**, 28 dec. 2017. Available at: <https://arxiv.org/pdf/1712.09923>. Access in: 24 abr. 2024.

IBM. **Introduction to the AI Framework for z/OS**, 27 feb. 2025. Disponível em: <https://www.ibm.com/docs/en/zos/3.1.0?topic=zos-introduction-ai-framework>. Acesso em: 13 dez. 2024.

JORNAL OFICIAL DA UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial e que altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as**

**Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828**, p. 1-144. Disponível em: [https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=OJ:L_202401689). Acesso em: 29 nov. 2024.

JOU, Graciela Inchausti de; SPERB, Tania Mara. Teoria da Mente: diferentes abordagens. **Psicologia: Reflexão e Crítica**, v. 12, n. 2, 1999. Disponível em: <https://www.scielo.br/j/prc/a/H7Bb5zCwRFqfLK8BNrf7ZGS/?lang=pt>. Acesso em: 24 abr. 2024.

KAHNEMAN, Daniel. **Rápido e Devagar: duas formas de pensar**. Tradução: Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.

KAHNEMAN, Daniel; SIBONY, Olivier; SUNSTEIN, Cass R. **Ruído**. Tradução: Cássio de Arantes. 1.<sup>a</sup> ed. Rio de Janeiro: Onketiva, 2021.

KAUFMANN, Dora. **Desmestificando a inteligência artificial**. São Paulo: Autêntica, 2022.

KAUFMAN, Dora. JUNQUILHO, Táina; REIS, Priscila. Externalidades Negativas da Inteligência Artificial: Conflitos entre limites da técnica e dos direitos humanos. **Revista de Direitos e de Garantias Fundamentais**, Vitória, v. 24, n. 3 p. 43-71, set./dez. 2023. Disponível em: <https://doi.org/10.18759/rdgf.v24i3.2198>. Acesso: 28 nov. 2024.

KLEINBERG, Jon *et al.* Algorithms as discrimination detectors. **Proceedings of the National Academy of Sciences (PNAS)**, v. 117, n. 48, p. 30096-30100, 1.<sup>o</sup> dec. 2020. Available at: <https://www.pnas.org/doi/epdf/10.1073/pnas.1912790117>. Access in: 1.<sup>o</sup> maio 2023.

KLEINBERG, Jon *et al.* Discrimination in the Age of Algorithms. **Journal of Legal Analysis**, v. 10, p. 113-174, 2019. Available at: <https://academic.oup.com/jla/article/doi/10.1093/jla/laz001/5476086>. Access in: 1.<sup>o</sup> maio 2023.

KLEINBERG, Jon *et al.* Human Decisions and Machine Predictions. **The Quarterly Journal of Economics**, v. 133, n. 1, p. 237-293, 2018.

KOGA, Bruno Yudi Soares. **Precificação Personalizada**. 1.<sup>a</sup> ed. São Paulo: Almedina, 2021.

KOGA, Bruno Yudi Soares. **Tecnologias persuasivas, decisões e sua liberdade: Como as tecnologias persuasivas afetam suas decisões e tolem sua liberdade, formando um cabresto digital**. Londrina: Thoth, 2024.

LAGE, Fernanda de Carvalho. **Manual de Inteligência Artificial no Direito Brasileiro**. 1.<sup>a</sup> ed. Salvador: Juspodivm, 2021.

LINDOSO, Maria Cristine Branco. **Discriminação de Gênero no Tratamento Automatizado de Dados Pessoais**. 1.<sup>a</sup> ed. São Paulo: Editora Processo, 2021.

LOURENÇO, Flávia de Faria. Data Mining, Big Data e o Direito à Privacidade. *In*: BLUM, Renato Opice; ABRUSIO, Juliana; WALDER, Priscila Ungaretti de Godoy. **Temas relevantes de direito digital**. São Paulo: Editora IASP, 2020.

LOURIDAS, Panos. **Algorithms**. Cambridge-MA: The Mit Press, 2019.

MAGALHÃES, Guilherme Martins; LONGHI, João Victor Rozatti; FALEIROS JÚNIOR, José Luiz de Moura. **Comentários à Lei Geral de Proteção de Dados Pessoais: Lei 13.709/2018**. 1.<sup>a</sup> ed. São Paulo: Foco, 2022.

MAGRANI, Eduardo; OLIVEIRA, Renana Medeiros de. Desafios Regulatórios da Internet das Coisas: reflexões sobre LGPD, o consentimento e o direito à explicação. *In*: CRESPO, Marcelo Xavier de Freitas *et al.* **Compliance no Direito Digital**. São Paulo: Thomson Reuters Brasil, 2020, p. 65-105.

MANOEL, Sérgio Silva. **Governança de Segurança da Informação: como criar oportunidades para o seu negócio**. 1.<sup>a</sup> ed. Rio de Janeiro: Brasport, 2014.

MARQUES, Elaine Cristina; SILVA, Victor Hugo Cunha. Capítulo 1 - Vieses Algorítmicos nas Aplicações de Inteligência Artificial: uma análise dos projetos de lei brasileiros. *In*: VIGLIAR, José Marcelo Menezes. **Inteligência artificial: aspectos jurídicos**. São Paulo: Almedina, 2023, p. 17-41.

MARTINS, Pedro Bastos Lobo. **Profiling na Lei Geral de Proteção de Dados: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica**. 1.<sup>a</sup> ed. Indaiatuba: Foco, 2022. (Coleção Persona – coordenador: Brunello Souza Stancioli).

MASSENO, Manuel David. Como a União Europeia procura proteger os cidadãos-consumidores em tempos de Big Data. *In*: Martins, Guilherme Magalhães; LONGHI, João Victor Rozatti. **Direito Digital: Direito Privado e Internet**. 4.<sup>a</sup> ed. Indaiatuba: Foco, 2021, p. 496.

McCARTHY, John. **Wha its Artificial intelligence?** Computer Science Department. Stanford: Stanford University CA 94305, 12 nov. 2007, p. 1-15. Disponível em: <https://www-formal.stanford.edu/jmc/whatisai.pdf>. Acesso em: 2 dez. 2024.

MEDON, Filipe. **Inteligência Artificial e Responsabilidade Civil: autonomia, riscos e solidariedade**. Salvador: Juspodivm, 2020.

MELO JÚNIOR, Fábio Antero Pulpa; OLIVEIRA, Mauri Aparecido de. **Backpropagation e Redes Neurais – Volume 1: MLP, RPROP, QPROP, LMBP, DELTA-BAR-DELTA, MATLABR E PYTHON™**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2024.

MENEZES, Marcos Antonio de. A Inteligência Artificial versus a Inteligência Humana. **Saber Humano**, v. 13, n. 22, p. 220-239, jan./jun. 2023. Disponível em: <https://saberhumano.emnuvens.com.br/sh/article/view/576/590>. Acesso em: 29 abr. 2024.

MERCEDES-BENZ GROUP. **Como a Mercedes-Benz usa a Inteligência Artificial (IA) – Duas Letras e Quatro Princípios**. Disponível em: <https://group.mercedes-benz.com/responsibility/compliance/digital/ki-guidelines.html>. Acesso em: 28 abr. 2024.

MEZZARROBA, Orides; MONTEIRO, Cláudia Servilha. **Manual de Metodologia da Pesquisa no Direito**. 8.<sup>a</sup> ed. São Paulo: SaraivaJur, 2019.

MISES, Ludwig Von. **Ação Humana**. São Paulo: Instituto Ludwig Von Mises Brasil, 2023.

MITTELSTADT, Brent *et al.* The ethics of algorithms: Mapping the debate. **Big Data & Society**, jul./dec. 2016. Available at: [https://www.researchgate.net/publication/309322060\\_The\\_Ethics\\_of\\_Algorithms\\_Mapping\\_the\\_Debate](https://www.researchgate.net/publication/309322060_The_Ethics_of_Algorithms_Mapping_the_Debate). Access in: 29 abr. 2024.

MOMO, Maria Vitória Galvan; BONE, Leonardo Castro de. A Proteção de Dados Pessoais Enquanto Direito Fundamental: um Olhar à Luz do Direito Comparado. *In:* COMÉRIO, Murilo Siqueira; JUNQUILHO, Tainá Aguiar. **Direito e Tecnologia: Um debate multidisciplinar**. Rio de Janeiro: Lumen Iuris, 2021, p. 23-53.

MOREIRA, Adilson José. **Tratado de Direito Antidiscriminatório**. 1.<sup>a</sup> ed. São Paulo: Editora Contracorrente, 2020.

MOREIRA, Adilson José *et al.* **Direito antidiscriminatório e relações raciais: Práticas excludentes, perspectivas críticas, medidas inclusivas**. 1.<sup>a</sup> ed. São Paulo: Editora Conhecimento, 2023.

MOREIRA, Diogo Rais Rodrigues; BARBOSA, Nathalia Sartarello. O Reflexo da sociedade do hiperconsumo no Instagram e a Responsabilidade Civil dos Influenciadores Digitais. **Revista de Direitos Culturais**, Santo Ângelo, v. 13, n. 30, p. 73-88, mai./ago. 2018. Disponível em: [https://www.researchgate.net/publication/328376203\\_O\\_REFLEXO\\_DA\\_SOCIEDADE\\_DO\\_HIPERCONSUMO\\_NO\\_INSTAGRAM\\_E\\_A\\_RESPONSABILIDADE\\_CIVIL\\_DOS\\_INFLUENCIADORES\\_DIGITAIS](https://www.researchgate.net/publication/328376203_O_REFLEXO_DA_SOCIEDADE_DO_HIPERCONSUMO_NO_INSTAGRAM_E_A_RESPONSABILIDADE_CIVIL_DOS_INFLUENCIADORES_DIGITAIS). Acesso em: 29 abr. 2024.

MORIN, Christophe. Neuromarketing: The new science of consumer behavior. **Society**, v. 48, n. 2, p. 131-135, 2011.

NANNINI, Luca; BALAYN, Agathe; SMITH, Adam Leon. Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK. **ArXiv preprint**, 2023. Disponível em: <https://arxiv.org/abs/2304.11218>. Acesso em: 5 nov. 2025.

NEUFELD, Carmem Beatriz; BRUST, Priscila Goergewn; STEIN, Lilian Milnitsky. Bases Epistemológicas da Psicologia Cognitiva Experiental. **Psicologia: Teoria e Pesquisa**, v. 27, n. 1, p. 103-112, jan./mar. 2011. Disponível em: <https://www.scielo.br/j/ptp/a/DJVCxtgRTDRG6y45VJHbgzS/?format=pdf&lang=pt>. Acesso em: 29 abr. 2024.

NEVES, Clarissa Eckert Baeta; SAMIOS, Eva Machado Barbosa. **Niklas Luhmann: a Nova Teoria dos Sistemas**. 1.<sup>a</sup> ed. Porto Alegre: Editora da Universidade Federal do Rio Grande do Sul, Goethe-Institut/ICBBA, 1997. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/242209/000152200.pdf?sequence=1&isAllowed=y>. Acesso em: 24 abr. 2024.

OLIVEIRA, M. T.; SILVA, R. A. Marcadores somáticos em transtornos de ansiedade: uma revisão sistemática. **Revista Brasileira de Psiquiatria**, v. 46, n. 1, p. 45-58, 2024.

OPENAI. **GPT-4 Technical Report**, 27 mar. 2023, p. 1-100. Disponível em: <https://cdn.openai.com/papers/gpt-4.pdf>. Acesso em: 27 mar. 2024.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). Governing with Artificial Intelligence. **OECD Artificial Intelligence Papers**, 13 jun. 2024. Available at: [https://www.oecd.org/en/publications/governing-with-artificial-intelligence\\_26324bc2-en.html](https://www.oecd.org/en/publications/governing-with-artificial-intelligence_26324bc2-en.html). Access in: 7 dec. 2024.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). OECD Framework for the Classification of AI systems. **OECD Digital Economy Papers**, 22 feb. 2022. Available at: [https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems\\_cb6d9eca-en.html](https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html). Access in: 13 dec. 2024.

OSOBA, Osonde; WELSER IV, William. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. **Rand Corporation**, 2017, p. 1-45. Disponível em: [https://www.rand.org/content/dam/rand/pubs/research\\_reports/RR1700/RR1744/RAND\\_RR1744.pdf](https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf). Acesso em: 13 jun. 2024.

PALHARES, Felipe; PRADO, Luis Fernando; VIDIGAL, Paulo. **Compliance Digital e LGPD**. São Paulo: Thomson Reuters, 2021.

PAPERT, Seymour; SOLOMON, Cynthia. Twenty Things to do with a Computer. **Artificial Intelligence Education**, v. 15, n. 3, p. 115-130, 2024.

PARANHOS, Mário Cosac Oliveira. **Viés Algorítmico: uma análise sobre discriminações automatizadas**. Rio de Janeiro: Lumen Iuris, 2022.

PASQUALE, Frank. Algorithms, healthcare, and ethics: Understanding the implications of machine decision-making. **Health Policy Journal**, v. 15, n. 3, p. 87-104, 2019.

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

PASQUALE, Frank; BREVINI, Benedetta. Revisiting the Black Box Society by rethinking the political economy of big data. **Sage Journals**, 20 oct. 2020. Disponível em: <https://journals.sagepub.com/doi/10.1177/2053951720935146>. Acesso em: 9 dez. 2024.

PASETTI, Marcelo. **Inteligência Artificial aplicada ao Direito Tributário. Um Novo Modelo na Construção de uma Justiça Fiscal?** São Paulo: Lumem Juris, 2019.

PENROSE, Roger. **A mente do novo imperador: Sobre computadores, mentes e as leis da física**. Tradução: Gabriel Cozzella. São Paulo: UNESP, 2023.

PEREIRA, João Sérgio dos Santos Soares. **A padronização decisória na era da inteligência artificial: uma possível leitura hermenêutica e da autonomia do direito**. Belo Horizonte: Casa do Direito, 2021. (Coleção Direito Processual na Ordem Constitucional).

PERSONAL DATA PROTECTION COMMISSION. **Singapore's Approach to AI Governance**. Available at: <http://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework>. Access in: 7 dec. 2024.

PICAZIO, Joseph Rodrigo Amorim. **Inteligência Artificial e seus Impactos nos Direitos Sociais: um panorama da convergência e dos desafios da IA para a efetividade dos direitos e garantias fundamentais**. São Paulo: Dialética, 2024.

PINHO, Anna Carolina *et al.* **Discussões sobre direito na era digital**. 1.<sup>a</sup> ed. Rio de Janeiro: GZ, 2021.

PINKER, Steven. **Racionalidade: o que é, porque parece estar em falta, por que é importante**. Tradução: Waldéa Barcellos. 1.<sup>a</sup> ed. Rio de Janeiro: Intrínseca, 2022.

PINTO, Felipe Chiarello de Souza; AZZARI, Bruna; SILVA, Amanda Scalisse. Ética e Integridade nas Instituições de Ensino Superior: A importância da Implementação de Programas de *Compliance* nas Universidades. **Revista de Direito Brasileira**, Florianópolis, v. 26, n. 10, p. 457-470, mai./ago. 2020. Disponível em: <https://www.indexlaw.org/index.php/rdb/article/view/6252/5122>. Acesso em: 24 abr. 2024.

PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Desafios Internacionais da Aplicação da Inteligência Artificial no Direito. **Revista Justiça de Direito**, v. 35, n. 2, p. 6-29, mai./ago. 2021. Disponível em: <https://seer.upf.br/index.php/rjd/article/view/13040/114116181>. Acesso em: 24 abr. 2024.

PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha. Inovação e sua Principal Barreira Jurídica. **Revista Eletrônica do Curso de Direito da UFSM (Universidade Federal de Santa Maria)**, v. 15, n. 2, 2020. Disponível em: <https://periodicos.ufsm.br/revistadireito/article/view/41419/pdf>. Acesso em: 24 abr. 2024.

PINTO, Felipe Chiarello de Souza; GARCIA, Lara Rocha; ROSA, Alexandre de Moares. Direito para Humanos Robotizados ou Direito dos Robôs Humanizados? **Revista Novos Estudos Jurídicos**, Itajaí, v. 28, n. 3, p. 536-553, 2023. Disponível em: <https://periodicos.univali.br/index.php/nej/article/view/19985>. Acesso em: 24 abr. 2024.

PINTO, Felipe Chiarello de Souza; OLIVEIRA, Gabriela Franklin de. Não acredite em tudo que vê: *deepfake pornography* e responsabilidade civil no ordenamento jurídico brasileiro. **Revista Direito & Política**, v. 18, n. 2, p. 427-451, 2023. Disponível em: <https://periodicos.univali.br/index.php/rdp/article/view/19869>. Acesso em: 24 abr. 2024.

PREMACK, David; WOODRUFF, Guy. Does the chimpanzee have a theory of mind? **The Behavioural and Brain Science**, v. 1, n. 4, p. 515-526, 1978. Disponível em: <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1E96B02CD9850016B7C93BC6D2FEF1D0/S0140525X00076512a.pdf/does-the-chimpanzee-have-a-theory-of-mind.pdf>. Acesso em: 29 abr. 2024.

QUINELATO, Pietra Daneluzzi. **Preços Personalizados à luz da Lei Geral de Proteção de Dados**. 1.<sup>a</sup> ed. São Paulo: Foco, 2022.

QUINELATO, Pietra Daneluzzi; FALEIROS JÚNIOR, José Luiz de Moura. Economia Comportamental e Inteligência Artificial na publicidade Veiculada em Mercados Ricos em Dados. In: PINHO, Anna Carolina *et al.* **Discussões sobre Direito na Era Digital**. 1.<sup>a</sup> ed. Rio de Janeiro: GZ, 2021, p. 637-654.



RAHMAN, Was. **Inteligência artificial e aprendizado de máquina**. 1.<sup>a</sup> ed. São Paulo: SENAC, 2022.

RAIS, Diogo; ABRUSIO, Juliana. **Privacidade, mercado e cidadania: uma conexão a partir da IA**. Prefácio: Felipe Chiarello. São Paulo: Revista dos Tribunais, 2022.

RAIS, Diogo; PRADO FILHO, Francisco Octávio de Almeida. **Direito público digital: o Estado e as novas tecnologias – desafios e soluções**. São Paulo: Revista dos Tribunais, 2020.

REOLON, Carlos Eduardo Passerani. Governança técnica dos algoritmos de IA. *In*: VAINZOF, Rony *et al.* **Comentários ao EU AI Act: Uma abordagem prática e teórica do Artificial Intelligence Act da União Européia**. 1.<sup>a</sup> ed. São Paulo: Revista dos Tribunais, 2024, p. 243-257.

REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Revista Eletrônica de Direito Civil**, Rio de Janeiro, ano 11, n. 3, p. 7-9, 2022. Disponível em: <http://civilistica.com/discriminacao-algoritmica/>. Acesso em: 1.<sup>o</sup> dez. 2024.

RODOTÀ, Stefano. **A Vida na Sociedade da Vigilância: a Privacidade Hoje**. 1.<sup>a</sup> ed. Rio de Janeiro: Renovar, 2008.

ROLLS ROYCE. **O Marco Aletheia – Ajudando a construir confiança na inteligência artificial**. Disponível em: <https://www.rolls-royce.com/innovation/the-aletheia-framework.aspx>. Acesso em: 27 abr. 2024.

RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, v. 1, p. 206- 215, 2019. Disponível em: <https://www.nature.com/articles/s42256-019-0048-x>. Acesso em: 24 abr. 2024.

RUSSELL, Stuart. **Human Compatible: Artificial Intelligence and the Problem of Control**. Westminster: Penguin Books. 2020.

RUSSEL, Stuart. **Inteligência Artificial a nosso favor: Como manter o controle sobre a tecnologia**. Tradução: Berilo Vargas. 1.<sup>a</sup> ed. São Paulo: Companhia das Letras, 2021.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência Artificial - Uma Abordagem Moderna**. Tradução: Daniel Vieira, Flávio Soares Corrêa da Silva. 4.<sup>a</sup> ed. São Paulo: LTC, 2022.

SANTAELLA, Lúcia. Desafios e dilemas da ética na inteligência artificial. *In*: GUERRA FILHO, Willis Santiago Guerra Filho *et al.* **Direito e Inteligência Artificial: Fundamentos – Volume 1 - Inteligência Artificial, Ética e Direito**. Rio de Janeiro: Lumen Juris, 2021, p. 109-136.

SANTANA, Paulo Campanha. O papel do Direito no Mundo do Algoritmo. *In*: AVANCI, Thiago Felipe S. **O Futuro do Direito: o que esperar do Direito e das Tecnologias, Estudo em homenagem ao professor Angelo Viglianisi Ferraro**. Londrina: Thoth, 2022, p. 51.

SAJJACHOLAPUNT, Pitch; BALL, Linden J. The influence of banner advertisements on attention and memory: human faces with averted gaze can enhance advertising effectiveness.

**Frontiers in Psychology**, v. 5, article 166, p. 1-11, 3 mar. 2014. Disponível em: <https://www.researchgate.net/publication/260760911>. Acesso em: 27 abr. 2024.

SCHWEDE, Matheus Antes. **Inteligência Artificial Tendenciosa - Discriminação Racial pelo Viés Algorítmico**. Londrina: Thoth, 2023.

SHARDA, Ramesh; DELEN, Dursun; TURBAN, Efraim. **Business Intelligence e Análise de Dados para Gestão do Negócio**. Tradução: Ronald Saraiva de Menezes. Revisão: Ângela Brodbeck. 4.<sup>a</sup> ed. Porto Alegre: Bookman, 2019.

SICSU, Abraham Laredo; SAMARTINI, André; BARTH, Nelson Lerner. **Técnicas de Machine Learning**. São Paulo: Blucher, 2023.

SILVA, Rodrigo da Guia; TEPEDINO, Gustavo *et al.* **O Direito na Era da Inteligência Artificial**. 1.<sup>a</sup> ed. São Paulo: Thomson Reuters, 2020.

SILVER, David *et al.* Mastering the game of Go with deep neural networks and tree search. **Nature**, v. 529, n. 7587, p. 484-489, 2016.

SOUZA, Alan Rocha de *et al.* **Direito digital: Direito Privado e Internet**. 4.<sup>a</sup> ed. Indaiatuba: Foco, 2021.

SUTTON, Richard S.; BARTO, Andrew G. **Reinforcement Learning: An Introduction**. 2.<sup>a</sup> ed. Cambridge: MIT Press, 2018.

TAULLI, Tom. **Generativa AI: How ChatGPT and Other AI Tools Will Revolutionize Business**. New York: Apress, 2023.

TATEOKI, Victor Augusto. **O Uso dos Dados Pessoais como Mecanismo de Persuasão no Processo de Tomada de Decisão dos Usuários de Internet**. Rio de Janeiro: Lumen Iuris, 2021.

TEGMARK, Max. **Vida 3.0: o ser humano na era da inteligência artificial**. Tradução: Petê Rissati. São Paulo: Benvirá, 2020.

TEIXEIRA FILHO, Sócrates Arantes. **Segurança da Informação Descomplicada**. 1.<sup>a</sup> ed. Brasília-DF: 2015.

TEIXEIRA, Tarcísio. **Direito Digital e Processo Eletrônico**. 8.<sup>a</sup> ed. São Paulo: SaraivaJur, 2024.

THALER, Richard H.; SUNSTEIN, Cass R. **Nudge: Como tomar melhores decisões sobre saúde, dinheiro e felicidade**. Tradução: Ângelo Lessa. 1.<sup>a</sup> ed. Rio de Janeiro: Companhia das Letras, 2023.

TURING, Alan M. **Mind a Quarterly Review of Psychology and Philosophy**. Disponível em: <https://academic.oup.com/mind/article/LIX/236/433/986238>. Acesso em: 2 dez. 2024.

UNITED STATES. Department of Commerce. **AI Risk Management Framework**.

Gaithersburg, MD: National Institute of Standards and Technology, 26 jul. 2024. Disponível em: <https://www.nist.gov/itl/ai-risk-management-framework>. Acesso em: 7 dez. 2024.

UNITED STATES. NIST Trustworthy and Responsible AI NIST AI 600-1. **Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile**. NIST – National Institute of Standards and Technology, 25 jul. 2024. Available at: <https://doi.org/10.6028/NIST.AI.600-1>. Acesso: 13 dez. 2024.

URAFSKY, Daniel; MARTIN, James H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3.<sup>a</sup> ed. Hoboken: Pearson, 2023.

VAINZOF, Rony. Aplicações e Regulação da Inteligência Artificial. In: VAINZOF, Rony; SERAFINO, Danielle; STEINWASCHER, Aline. **Legal Innovation: o Futuro do Direito e o Direito do Futuro**. São Paulo: Thomson Reuters, 2022, p. 248.

VAINZOF, Rony *et al.* **EU AI Act: Uma abordagem prática e teórica do Artificial Intelligence Act da União Européia**. São Paulo: Thomson Reuters, 2024.

VASWANI, Ashish *et al.* Attention is all you need. **31st Conference on Neural Information Processing Systems (NIPS)**. Long Beach – CA: 2017, p. 1-11. Available at: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Access in: 25 jan. 2025.

VEALE, Michael; VAN KLEEK, Max; BINNS, Reuben. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. **CHI Conference on Human Factors in Computing Systems**, n. 440, p. 1-14, 21 apr. 2018.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. **International Data Privacy Law**, v. 7, n. 2, p. 76-99, 2017. Disponível em: <https://academic.oup.com/idpl/article/7/2/76/3860948>. Acesso em: 28 abr. 2024.

WARREN, Samuel D.; BRANDEIS, Louis D. The Right to Privacy. **Harvard Law Review**, v. 4, n. 5, p. 193-220, 15 dec. 1890. Disponível em: <https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf>. Acesso em: 28 abr. 2024.

WING, Jeannette M. Computational thinking and thinking about computing. **Philosophical Transactions of the Royal Society a Mathematical, Physical and Engineering Sciences**, v. 378, p. 3717-3725, 2023.

YADAV, Aman; STEPHENSON, Chris; HONG, Hai. Computational Thinking for Teacher Education. **Communications of the ACM**, v. 60, n. 4, p. 55-62, 1.º apr. 2017.