



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE DIREITO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO
CURSO DE MESTRADO EM DIREITO, ESTADO E CONSTITUIÇÃO

GABRIEL DE ARAÚJO OLIVEIRA

**PERFILAMENTO E RACISMO ALGORÍTMICO: DESAFIOS À
PROTEÇÃO DA POPULAÇÃO NEGRA À LUZ DO ORDENAMENTO
JURÍDICO BRASILEIRO**

Brasília – DF
2025

GABRIEL DE ARAÚJO OLIVEIRA

**PERFILAMENTO E RACISMO ALGORÍTIMICO: DESAFIOS À
PROTEÇÃO DA POPULAÇÃO NEGRA À LUZ DO ORDENAMENTO
JURÍDICO BRASILEIRO**

Dissertação de Mestrado apresentada
ao Programa de Pós-Graduação em
Direito da Faculdade de Direito,
Universidade de Brasília, como requisito
parcial à obtenção do título de Mestre
em Direito, Estado e Constituição.

Orientador: Prof. Dr. Alexandre Kehrig
Veronese Aguiar.

Brasília – DF

2025

GABRIEL DE ARAÚJO OLIVEIRA

**PERFILAMENTO E RACISMO ALGORÍTIMICO: DESAFIOS À
PROTEÇÃO DA POPULAÇÃO NEGRA À LUZ DO ORDENAMENTO
JURÍDICO BRASILEIRO**

Dissertação de Mestrado apresentada
ao Programa de Pós-Graduação em
Direito da Faculdade de Direito,
Universidade de Brasília, como requisito
parcial à obtenção do título de Mestre em
Direito, Estado e Constituição.

BANCA EXAMINADORA

Prof. Dr. Alexandre Kehrig Veronese Aguiar
Orientador – Faculdade de Direito/Universidade de Brasília

Prof. Dr. Fabiano Hartmann Peixoto
Examinador Interno – Faculdade de Direito/Universidade de Brasília

Prof.^a Dr.^a . Tainá Aguiar Junquilha
Examinadora Externa – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa
(IDP)

Brasília, 9 de setembro de 2025.

*Dedico esta dissertação de mestrado à
minha mãe, Edirani Santos de Araújo, ao
meu pai, Edvaldo dos Anjos Oliveira, às
minhas avós, Lindaura Santos de Araújo
e Valdelice dos Anjos Oliveira, bem
como à minha irmã, Kelly Cristine de
Araújo Oliveira, e às minhas sobrinhas,
Carolina Oliveira de Jesus e Esther de
Araújo Pessoa.*

Gabriel de Araújo Oliveira.

AGRADECIMENTOS

Agradeço, primeiramente, à minha família, pelo apoio incondicional, paciência e incentivo contínuo durante toda a jornada deste mestrado. Sua compreensão e o encorajamento foram fundamentais para que eu pudesse superar os desafios e chegar até o final desse processo.

Aos meus amigos, expresso minha gratidão pela parceria constante, pelos momentos de descontração e pelo suporte emocional, que trouxeram leveza e equilíbrio a esta trajetória.

Ao meu orientador, Alexandre Veronese, registro meu profundo agradecimento por todo apoio e compreensão ao longo dessa jornada.

Aos membros da banca examinadora, Professor Fabiano Hartmann e Professora Tainá Junquilha, sou grato pelas avaliações cuidadosas, sugestões valiosas e discussões que certamente enriquecerão significativamente esta dissertação, proporcionando novos olhares e aprimoramentos ao estudo.

Agradeço à Universidade de Brasília (UnB), pelo ambiente acadêmico estimulante, pela infraestrutura disponibilizada e pelo suporte essencial ao desenvolvimento da pesquisa.

Este trabalho é o resultado do apoio e da colaboração de todos vocês, e sou imensamente grato por cada contribuição recebida.

*"Say a prayer for what has been
We'll be the ones to purify our fathers' sins
American requiem
Them old ideas
Are buried here
Amen!"*
Beyoncé

RESUMO

A partir de uma perspectiva interdisciplinar, que parte da intersecção entre direito, tecnologia e discriminação, o presente trabalho objetiva analisar o perfilamento automatizado enquanto uma manifestação do racismo algorítmico e, por consequência, os esforços do Brasil para regulamentação dessa técnica através de normas existentes no ordenamento jurídico pátrio, a exemplo da Lei Geral de Proteção de Dados (LGPD), e da iminente regulamentação da Inteligência Artificial (IA) como uma forma de combate ao racismo estrutural. No primeiro capítulo, são feitas considerações técnicas sobre categorias como IA, dados pessoais, big data, algoritmos e, não menos importante, aprendizado de máquina e IA generativa. No segundo capítulo, discorre-se sobre o racismo como construção social e estruturante da sociedade capitalista e, por conseguinte, analisa-se o fenômeno racismo algorítmico, notadamente a partir de manifestações no perfilamento automatizado. Por fim, no terceiro e último capítulo, são analisados mecanismos presentes na legislação de proteção de dados pessoais, assim como em iniciativas globais e nacionais, a exemplo do Projeto de Lei n. 2.338/2023, formas de inibir violações a direitos fundamentais e de contribuir para a proteção da população negra ante os riscos discriminatórios do perfilamento automatizado.

Palavras-chave: Perfilamento. Racismo algorítmico. Inteligência artificial – IA. Regulação. Lei Geral de Proteção de Dados – LGPD.

ABSTRACT

From an interdisciplinary perspective, which begins at the intersection of law, technology, and discrimination, this paper aims to analyze automated profiling as a manifestation of algorithmic racism and, consequently, Brazil's efforts to regulate this technique through existing standards in the national legal system, such as the Brazilian General Data Protection Law (LGPD), and the imminent regulation of Artificial Intelligence (AI) as a way to combat structural racism. The first chapter presents technical considerations on categories such as AI, personal data, big data, algorithms, and, no less importantly, machine learning and generative AI. The second chapter discusses racism as a social and structuring construct of capitalist society and, consequently, analyzes the phenomenon of algorithmic racism, notably through its manifestations in automated profiling. Finally, the third and final chapter analyzes mechanisms present in personal data protection legislation, as well as in global and national initiatives, such as Bill no. 2.338/2023, ways to inhibit violations of fundamental rights and to contribute to the protection of the Black population against the discriminatory risks of automated profiling.

Keywords: Profiling. Algorithmic racism. Artificial intelligence – AI. Regulation. Brazilian General Data Protection Law – LGPD.

LISTA DE FIGURAS

Figura 1 – Linha do tempo da IA com base em eventos históricos

Figura 2 – Estrutura de processamento no *machine learning* e no *deep learning*

Figura 3 – Linha do tempo das revoluções industriais

Figura 4 – Representação gráfica do “rosto” dos trabalhos mais bem e mal remunerados

Figura 5 – Estágio da regulação da inteligência artificial ao redor do mundo

Figura 6 – Tipos de sistemas de IA correspondentes a diferentes níveis de risco.

Tabela 1 – Algumas definições de inteligência artificial

Tabela 2 – Técnicas utilizadas em IA generativa

LISTA DE ABREVIATURAS E SIGLAS

AM – Aprendizado de Máquina

AP – Aprendizagem Profunda

ANPD – Autoridade Nacional de Proteção de Dados

GDPR – General Data Protection Regulation

IA - Inteligência Artificial

IAGen – Inteligência Artificial Generativa

LGPD – Lei Geral de Proteção de Dados Pessoais

LLM – Large Language Model

STF – Supremo Tribunal Federal

SUMÁRIO

1. INTRODUÇÃO	12
2. INTELIGÊNCIA ARTIFICIAL: ENTRE DADOS E ALGORITMOS	14
2.1. Uma breve introdução à inteligência artificial	14
2.2. Dados pessoais e big data	22
2.3. Algoritmos e aprendizado de máquina	29
2.4. Inteligência artificial generativa: da criação ao ineditismo	35
3. RACISMO ALGORÍTMICO E PERFILAMENTO	42
3.1. Vieses discriminatórios	42
3.2. Racismo algorítmico	48
3.3. Perfis e Perfilamento	56
4. REGULAÇÃO DO PERFILAMENTO NO ORDENAMENTO JURÍDICO BRASILEIRO	66
4.1. A regulação sob a ótica da proteção de dados	66
4.2. Perfilamento e direito de revisão	75
4.3. Perspectivas regulatórias e o Projeto de Lei n. 2.338/2023	80
5. CONSIDERAÇÕES FINAIS	87
6. REFERÊNCIAS	90

1. INTRODUÇÃO

Por que falar de raça quando falamos de dados pessoais, inteligência artificial e algoritmos? É com essa indagação que Johanna Monagreda (2024) parte em uma missão com o objetivo de interseccionar e explicar a relação entre categorias distintas, mas que guardam estreita conexão. Muitas podem ser as respostas para esse questionamento. Na economia digital, os dados pessoais são tratados como um recurso valioso, muitas vezes comparado ao "novo petróleo", por ser o combustível necessário ao desenvolvimento e funcionamento de algoritmos de inteligência artificial. É dessa forma que produtos, serviços e estratégias são criados diariamente. A raça entra na equação como uma das bases fundantes desse sistema complexo e imbricado. Assim, sem delongas, é suficiente dizer que sistemas de opressão característicos da sociedade capitalista, como o racismo, também estruturam o mundo digital, reproduzindo e automatizando desigualdades, exclusões, vigilâncias e discriminações contra grupos racializados, especialmente a população negra.

A ascensão da inteligência artificial e a expansão do processamento de grandes volumes de dados, ou big data, reconfiguraram as dinâmicas sociais, econômicas e políticas do século XXI, promovendo inovações que transformaram desde interações cotidianas até processos decisórios complexos. Nesse contexto, o perfilamento automatizado, técnica que utiliza algoritmos de inteligência artificial para inferir características, comportamentos e/ou preferências de indivíduos e grupos, tornou-se uma ferramenta central em setores como segurança pública, saúde, educação, marketing, finanças, entre outros.

No entanto, o uso indiscriminado de sistemas de IA tem revelado sérios desafios éticos e jurídicos, particularmente frente ao fenômeno do racismo algorítmico, que perpetua desigualdades estruturais ao incorporar e reproduzir vieses históricos e sociais, impactando desproporcionalmente a população negra. Estudos como os de Noble (2018) e Benjamin (2019), apontam que sistemas algorítmicos, ao refletirem dados enviesados e decisões de design orientadas por interesses econômicos, frequentemente reforçam injustiças e estereótipos raciais, com consequências danosas em áreas como justiça criminal e recrutamento.

Este trabalho propõe uma análise do perfilamento automatizado como uma manifestação do racismo algorítmico, com o objetivo de compreender como essa tecnologia contribui para a reprodução de desigualdades raciais. A pesquisa avalia,

ainda, os esforços do ordenamento jurídico brasileiro para enfrentar esses desafios, com foco na Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018) e no Projeto de Lei nº 2.338/2023, que busca regulamentar a IA no país. Adotando uma abordagem interdisciplinar, que articula direito, tecnologia e estudos sobre discriminação, este estudo visa identificar os riscos inerentes aos sistemas algorítmicos e analisar a eficácia dos instrumentos normativos disponíveis para proteger a população negra contra práticas discriminatórias, promovendo reflexões sobre o papel do direito na construção de uma sociedade mais equitativa.

A relevância deste estudo é reforçada pela crescente dependência global de tecnologias de IA, que podem ampliar desigualdades se não forem adequadamente regulamentadas. No Brasil, onde a população negra representa mais de 50% dos habitantes, mas enfrenta barreiras sistemáticas em diversas esferas, a ausência de regulamentações específicas e a opacidade dos sistemas algorítmicos agravam os riscos de discriminação. Este trabalho busca contribuir para o debate acadêmico e jurídico, oferecendo reflexões sobre como o ordenamento brasileiro pode enfrentar os desafios do racismo algorítmico, promovendo transparência, responsabilidade e proteção efetiva à população negra.

2. INTELIGÊNCIA ARTIFICIAL: ENTRE DADOS E ALGORITMOS

2.1. Uma breve introdução à inteligência artificial

A definição mais difundida sobre Inteligência Artificial (IA) na cultura popular tem estreita relação com a capacidade de máquinas e sistemas computacionais realizarem tarefas que, usualmente, são reproduzidas por seres humanos e, portanto, exigem inteligência humana. Essa visão, que está de certa forma alinhada à realidade que hoje se observa, pouco se aproxima daquela que outrora era nutrida pela população mundial durante o século XX e associada fortemente à superação da espécie humana pelas máquinas¹ – graças, em grande parte, às narrativas de ficção científica, como 2001: Uma Odisseia no Espaço, de Stanley Kubrick; Eu, Robô, de Isaac Asimov; e Matrix, de Lana e Lily Wachowski.

Embora a inteligência artificial se manifeste em contextos extraordinários tais quais os previstos no passado, como em carros autônomos e no setor de saúde através do diagnóstico e tratamento de doenças, sua presença é percebida em atividades triviais do dia a dia (e.g, por meio dos motores de busca, como o Google e o Bing, no atendimento automatizado – *chatbot* –, nas análises de créditos e na indicação personalizada de produtos/conteúdo na internet, especialmente em redes sociais e aplicativos de streaming). Trata-se de um campo de estudo vasto e uma área multifacetada, sendo um verdadeiro desafio defini-la sem diminuir ou limitar sua grandiosidade e importância na era digital.

Peter Norvig e Stuart Russell explicam que, historicamente, é possível deduzir, ao menos, quatro definições de inteligência artificial que estão dispostas em duas dimensões: as que se relacionam a processos de pensamento e raciocínio e as que se referem ao comportamento (Russell; Norvig, 2013). Tais dimensões se desdobram em nível de sucesso, o qual é mensurado em termos de fidelidade ao desempenho humano e ao conceito ideal de inteligência, a denominada racionalidade. Assim, segundo os autores, quatro são as estratégias seguidas para o estudo da inteligência artificial: a abordagem humana, que consiste pensar ou agir como um ser humano, e a abordagem da racionalidade, que, por outro lado, consiste em pensar ou agir de forma ideal, ou seja, racionalmente (Tabela 1).

¹ Essa perspectiva coincide com o conceito de Superinteligência Artificial (“Super IA”), que é um tipo hipotético de inteligência artificial que possui capacidades cognitivas muito superiores às dos seres humanos, abrangendo raciocínio, aprendizado, tomada de decisões e outras habilidades intelectuais.

Tabela 1 – Algumas definições de inteligência artificial

Pensando como um humano	Pensando racionalmente
<p>“O novo e interessante esforço para fazer os computadores pensarem (...) <i>máquinas com mentes</i>, no sentido total e literal.” (Haugeland, 1985)</p> <p>“[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado...” (Bellman, 1978)</p>	<p>“O estudo das faculdades mentais pelo uso de modelos computacionais.” (Charniak e McDermott, 1985)</p> <p>“O estudo das computações que tornam possível perceber, raciocinar e agir.” (Winston, 1992)</p>
Agindo como seres humanos	Agindo racionalmente
<p>“A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas.” (Kurzweil, 1990)</p> <p>“O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas.” (Rich and Knight, 1991)</p>	<p>“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole <i>et al.</i>, 1998)</p> <p>“AI... está relacionada a um desempenho inteligente de artefatos.” (Nilsson, 1998)</p>

Fonte: Russell; Norvig, 2013, p. 25.

Para Tainá Aguiar Junquilha, “a IA não é a reprodução exata do comportamento humano, mas apenas algumas funções específicas do cérebro humano”, logo, “defini-la passa por compreender como funcionam os modelos de aprendizagem da máquina e saber identificar qual a melhor técnica se ajusta ao problema a ser resolvido” (Junquilha, 2023, p. 30).

Na missão de definir o que seria inteligência artificial, há, ainda, aqueles que mesclam elementos para conferir ao campo uma definição tecnicamente precisa e abrangente, como é o caso do Parlamento Europeu, que assim a descreve:

(...) a capacidade de uma máquina exibir habilidades semelhantes às humanas, como raciocínio, aprendizado, planejamento e criatividade. A IA permite que sistemas técnicos percebam seu ambiente, lidem com o que percebem, resolvam problemas e ajam para alcançar um objetivo específico. O computador recebe dados – já preparados ou coletados por meio de seus próprios sensores, como uma câmera –, os processa e responde. Sistemas de IA são capazes de adaptar seu comportamento até certo ponto, analisando os efeitos de ações anteriores e trabalhando de forma autônoma² (tradução nossa).

² European Parliament News. **What is artificial intelligence and how is it used?** Parlamento Europeu, 4 de setembro de 2020. Disponível em: <https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used>. Acesso em: 14 fev. 2025.

E também da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que concebeu uma definição de inteligência artificial com o intuito de ser tecnologicamente neutra e aplicável a horizontes temporais de curto e longo prazo. É, assim, uma definição esculpida para ser abstrata e ampla o suficiente para abranger a maioria das definições mais precisas utilizadas pelas comunidades científica, empresarial e política:

Um sistema de IA é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir das informações recebidas, como gerar resultados, como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e adaptabilidade após a implantação.³ (tradução nossa).

O conceito de IA é dinâmico e evolui constantemente, assim como a tecnologia em si. Nesse sentido, a compreensão da inteligência, enquanto parte da experiência humana e que consiste no conjunto de habilidades (entender, pensar, raciocinar, interpretar e julgar) de uma espécie que permitem, de maneira geral, interagir com o ambiente e adaptar-se a novas situações⁴, é essencial para apreender o significado de inteligência artificial, ou seja, é o ponto de partida.

Para Dora Kaufman, a inteligência é encarada como “a capacidade de compreender ideias complexas, de se adaptar efetivamente ao ambiente, de aprender com a experiência, de se envolver em várias formas de raciocínio, de superar os obstáculos” (Kaufman, 2019, p. 16). Surge daí a questão se sistemas de IA são ou podem vir a ser considerados “conscientes”, uma vez que a consciência é um atributo humano e que se relaciona ao conceito de inteligência.⁵ John Searle, notório filósofo americano, defende que “o aspecto primário e mais essencial das mentes é a consciência”, a qual ele entende como “os estados de conhecimento ou percepção

³ OEDC. **Explanatory Memorandum on the Updated OECD Definition of An AI System**. Disponível em: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf. Acesso em: 5 jun. 2025.

⁴ INTELIGÊNCIA. In: **Dicionário da Academia Brasileira de Letras**. Disponível em: <https://servbib.academia.org.br/dlp/>. Acesso em: 5 junho de 2025. <<http://www.academia.com.br>>. Acessado em 12/02/2003.

⁵ Embora relacionáveis, os termos inteligência e consciência não são sinônimos. A inteligência refere-se, de modo geral, à capacidade de processar informações, resolver problemas, aprender com experiências, adaptar-se a novos contextos e realizar tarefas complexas. Já a consciência consiste na experiência subjetiva de estar ciente de si mesmo e do ambiente, incluindo sensações, pensamentos e emoções. Ambos os conceitos estão intrinsecamente associados experiência humana. Na literatura sobre tema, não existe consenso sobre o que seria inteligência e/ou consciência.

que começam quando acordamos de manhã depois de um sono sem sonhos e continuam durante o dia até que adormecemos novamente” (Searle, 2000, p. 26).

Para Searle, há três características comuns aos estados conscientes: são internos, qualitativos e subjetivos. Internos porque acontecem dentro do corpo, especificamente dentro do cérebro, de modo que “a consciência ocorre necessariamente dentro de um organismo ou de algum outro sistema” (Searle, 2000, p. 27). Qualitativos porque “há, para cada um deles, um determinado modo de senti-lo, uma determinada característica qualitativa” (Searle, 2000, p. 26). E, por fim, subjetivos por “serem sempre experimentados por um sujeito humano ou animal. Portanto, os estados conscientes têm o que se pode chamar de ‘ontologia da primeira pessoa’. Ou seja, existem apenas do ponto de vista de algum agente, organismo, animal ou pessoa que os possua. Os estados conscientes têm um modo de existência na primeira pessoa” (Searle, 2000, p. 27). Nessa perspectiva, a consciência é tida como um fenômeno biológico intrinsecamente humano, resultado de processos cerebrais, mas com características subjetivas únicas, como intencionalidade e qualia (experiências subjetivas).

Kate Crawford defende que a ideia de “inteligência” artificial é equivocada, pois a dita inteligência não é intrínseca aos sistemas de IA e sua capacidade vem de treinamento intensivo com grandes volumes de dados, o que exige intervenção humana constante (Crawford, 2021, p. 8). Stuart Russell apresenta ponto de vista similar ao de Crawford, no sentido de que haveria um erro de definição de IA, pois, ao contrário dos seres humanos, as máquinas inteligentes não são capazes de agir para alcançar objetivos próprios (Russell, 2019). Os objetivos são atribuídos aos sistemas inteligentes pelos seres humanos – e, nesse ponto, é que, para o autor, consiste a inteligência: é a capacidade de atingir os seus objetivos, de agir com sucesso. Portanto, à máquina faltaria consciência, a partir da percepção de si e do mundo ao seu redor, o que “impõe um limite à inteligência artificial, que acredita que nunca será dotada de consciência, razão pela qual entende que não poderá sentir e tampouco competir com a inteligência humana” (Marques, 2020, p. 23).

Marvin Minsky, no entanto, desafia essa corrente ao propor que a inteligência é não uma prerrogativa exclusivamente humana. No artigo “Communication with Alien Intelligence”, publicado em 1985, o precursor do campo da inteligência artificial argumenta que a inteligência, definida como a capacidade de resolver problemas difíceis de forma rápida e inovadora – como construir espaçonaves ou sistemas de

comunicação de longa distância –, é um traço que pode ser compartilhado por outras espécies inteligentes, incluindo alienígenas (Minsky, 1985). Não obstante a proposição de Minsky tenha sido apresentada ao mundo há muitos anos, e antes do período conhecido como verão da IA⁶, oferece uma perspectiva distinta para se enfrentar a questão, considerando especialmente que “as técnicas atuais de IA lidam com percepção, análise de texto, processamento de linguagem natural (PNL)⁷, raciocínio lógico, sistemas de apoio à decisão, análise de dados e análise preditiva” (Kaufman, 2022, p. 11).

É certo que nenhuma definição consegue abranger suficientemente todos os aspectos da tecnologia e de seu potencial (Steibel; Vicente; De Jesus, 2020). Primeiro porque existem variados tipos de inteligência e no campo dos estudos da mente, que se estendem por áreas como a psicologia, filosofia e neurologia, não existe um consenso a respeito do significado de inteligência (Lage, 2020). E, por assim ser, em segundo lugar, nem todo sistema de IA é capaz de observar, agir ou criar, ao passo que outros são capazes de não apenas se comportar conforme programado, como também ir além ao produzir resultados inesperados ou inexplicáveis.⁸

Vale dizer que a complexidade inerente à IA é consequência de sua construção como uma tecnologia que agrega conhecimento de áreas inúmeras, a exemplo da ciência da computação, matemática, engenharia, filosofia, economia, psicologia, entre outras (Russell; Norvig, 2013). Comumente atribui-se a figuras como os filósofos Aristóteles e Ramon Llull as bases da evolução gradual do pensamento humano sobre a sistematização do raciocínio, da lógica e da automação do conhecimento – resultando na origem histórica da IA (Silva, 2020). Aristóteles em suas proposições pretendeu “codificar” o processo de raciocínio humano por meio de silogismos, que são formas de dedução lógica que estruturam argumentos em premissas e conclusões. Essa sistematização do pensamento foi um marco porque forneceu um

⁶ Período de grande entusiasmo, investimento e avanços significativos na pesquisa e aplicação de inteligência artificial. Cita-se como exemplos o período entre 1990 e 2000 e de 2010 até os dias atuais.

⁷ Processamento de Linguagem Natural (PLN) é uma tecnologia da inteligência artificial que utiliza métodos formais para analisar, interpretar, manipular e gerar textos em linguagem humana, permitindo que máquinas compreendam e interajam de forma natural com idiomas humanos.

⁸ Segundo Tainá Aguiar Junquilha, “a temática tem aflorado no mundo e fez surgir amplas pesquisas no que se tem chamado de *Explainable AI* (XAI), as quais exploram e pretendem avançar na explicação das decisões tomadas pela IA, em especial as que utilizam os modelos de redes neurais, cujos resultados são per se opacos. produção científica na área de XAI tem aumentado nos últimos anos e se dedicado a estudar técnicas e metodologias para compreensão e inteligibilidade dos modelos de aprendizagem profunda” (2023, pp. 16-17).

método para representar e validar o raciocínio de maneira formal, algo essencial para a IA moderna, que depende de regras lógicas para processar informações.

Já Llull destacou-se por sua *Ars Magna* (A Grande Arte, tradução nossa), uma "engrenagem" teórica destinada a responder todos os problemas por meio de combinações sistemáticas de conceitos. A *Ars Magna* consistia em um sistema lógico-combinatório baseado em discos giratórios que representavam categorias filosóficas e teológicas (e.g., Deus, bem, verdade), permitindo gerar proposições a partir de suas interconexões. Llull acreditava que esse mecanismo poderia revelar verdades universais de forma mecânica, sem depender exclusivamente da intuição humana. Embora as contribuições tenham sido consideradas percussoras para suas respectivas épocas, idade antiga (século IV a.c.) e idade média (século XIII), eram limitadas em razão dos recursos disponíveis e do próprio marco temporal – visto que o avanço tecnológico era lento e as teorias demoravam a ganhar forma. Isso não impediu, no entanto, que fossem revisitadas e aprimoradas por seus sucessores, que não desistiram da ideia de simular ou automatizar o pensamento humano, sendo o caso de Gottfried Wilhelm Leibniz, Charles Babbage e Ada Lovelace.

Não obstante as contribuições advindas dos teóricos mencionados, o campo da inteligência artificial começou a se desenvolver efetivamente no século XX, por meio de trabalhos realizados entre as décadas de 1930 e 1950. Destaca-se, nesse período, o artigo escrito em 1943 por Warren McCulloch e Walter Pitts sobre Redes Neurais, no qual se sugeriu a possibilidade de simular o sistema nervoso humano utilizando estruturas de raciocínio artificial representadas por modelos matemáticos⁹; e o estudo pioneiro publicado em 1950 pelo cientista britânico Alan Turing, por meio do qual propôs-se critérios para avaliar a “inteligência” das máquinas.¹⁰ A obra de Turing trouxe ao centro do debate científico o questionamento sobre a capacidade de computadores raciocinarem como seres humanos. Por intermédio de um experimento denominado Teste de Turing (ou Jogo da Imitação), o cientista visava determinar se uma máquina tinha a capacidade de simular atitudes tão semelhantes às de uma pessoa a ponto de se tornar impossível diferenciá-las em conversas com interlocutores não visíveis, fossem eles pessoas ou máquinas. Se a máquina conseguisse confundir o examinador humano, ela seria considerada inteligente,

⁹ McCulloch, Warren Sturgis; Pitts, Walter. **A logical calculus of the ideas immanent in nervous activity**. The bulletin of mathematical biophysics, 5(4): 115–133, 1943.

¹⁰ Turing, Alan Mathison. **Computing machinery and intelligence**. In: Mind, v. 59, p. 433–460, 1950.

necessitando, para isso, de aptidões em compreensão da linguagem natural e pensamento racional.

Estes são tidos como trabalhos pioneiros sobre IA, produzidos em uma época em que eram praticamente inexistentes os esforços para criação de uma área específica. Porém, o marco zero da IA moderna ocorreu em 1955, durante a Conferência de Dartmouth, organizada por John McCarthy, então professor na Dartmouth College. McCarthy, junto com Marvin Minsky, Nathaniel Rochester e Claude Shannon, propôs um workshop de dois meses para explorar como máquinas poderiam simular inteligência humana. A proposta de McCarthy trazia pela primeira vez o termo inteligência artificial e assim enunciava:

Propusemos que um estudo de dois meses e dez homens sobre inteligência artificial fosse realizado durante o verão de 1956 no Dartmouth College, em Hanover, New Hampshire. O estudo era para prosseguir com a conjectura básica de que cada aspecto da aprendizagem ou qualquer outra característica da inteligência pode, em princípio, ser descrita tão precisamente a ponto de ser construída uma máquina para simulá-la. Será realizada uma tentativa para descobrir como fazer com que as máquinas usem a linguagem, a partir de abstrações e conceitos, resolvam os tipos de problemas hoje reservados aos seres humanos e se aperfeiçoem. Achamos que poderá haver avanço significativo em um ou mais desses problemas se um grupo cuidadosamente selecionado de cientistas trabalhar em conjunto durante o verão. (McCarthy et al., 1955, tradução nossa)

Este momento representou uma virada, pois definiu a IA como uma disciplina científica com objetivos claros e delimitados, entre os quais estavam: resolver problemas complexos reservados aos humanos; compreender a linguagem natural e a partir daí gerá-la; aprender com a experiência; formar conceitos e abstrações; realizar tarefas criativas e, não menos importante – mas certamente o mais desafiador –, simular aspectos do cérebro humano. Embora a conferência tenha definido esses objetivos, os resultados imediatos foram modestos devido às limitações tecnológicas da época – especialmente em virtude das máquinas com pouca capacidade de processamento – e à subestimação da complexidade da inteligência humana. Ainda assim, Fernanda de Carvalho Lage aponta avanços tecnológicos observados nos anos seguintes que ajudaram na maturação epistemológica do campo da Inteligência Artificial (Lage, 2020):

1966: Os anos de 1956 a 1966 foram plenos de avanços no campo da IA. Pode-se destacar: i) em 1958, John McCarthy desenvolve a linguagem de programação LISP, que se torna a linguagem de programação mais popular usada na pesquisa de inteligência artificial;

ii) em 1959, Arthur Samuel inventa o termo "aprendizado de máquina", relatando a programação de um computador 'para que ele aprenda a jogar um jogo de damas melhor do que o que pode ser jogado pela pessoa que escreveu o programa'; iii) em 1961, o primeiro robô industrial, Unimate, começa a trabalhar em uma linha de montagem em uma fábrica da General Motors em Nova Jersey; iv) contudo, o marco é o ano de 1966, quando Joseph Weizenbaum criou a ELIZA, o primeiro chatbot (embora o termo chatbot somente tenha sido cunhado em 1994) com o objetivo de responder as perguntas de tal forma que as pessoas tivessem a impressão de estar conversando com outra pessoa e não com um programa de computador. O software Eliza, opera utilizando uma técnica de repetir (reformular) trechos das frases dos usuários para parecer que tem um extenso vocabulário. Foi a primeira tentativa de criação de um software que pudesse tentar passar no Teste de Turing, vez que conseguia reconhecer cerca de 250 tipos de frases.

1972: O WABOT-1 foi o primeiro robô antropomórfico desenvolvido no mundo. Consistia em um sistema de controle de membros, um sistema de visão e um sistema de conversação. O WABOT-1 era capaz de se comunicar em japonês e medir distâncias e direções para os objetos usando receptores externos, ouvidos e olhos artificiais e uma boca artificial. Estima-se que o WABOT-1 tinha a faculdade mental de uma criança de um ano e meio.

Conforme destaca a autora, entre os anos de 1974 e 1984, o campo da IA passou por seu primeiro inverno, visto que, a despeito do entusiasmo inicial, o relatório *Lighthill*¹¹ apontou que as descobertas feitas até aquele momento não foram acompanhadas do impacto prometido. Como consequência, os investimentos e financiamentos na área de IA foram afetados (Lage, 2020, p. 31). O segundo inverno ocorreu entre os anos de 1987 e 1993, quando, mais uma vez, as expectativas e promessas não foram atendidas e se tornou caro financiar e investir em projetos e pesquisas que não davam o retorno esperado.

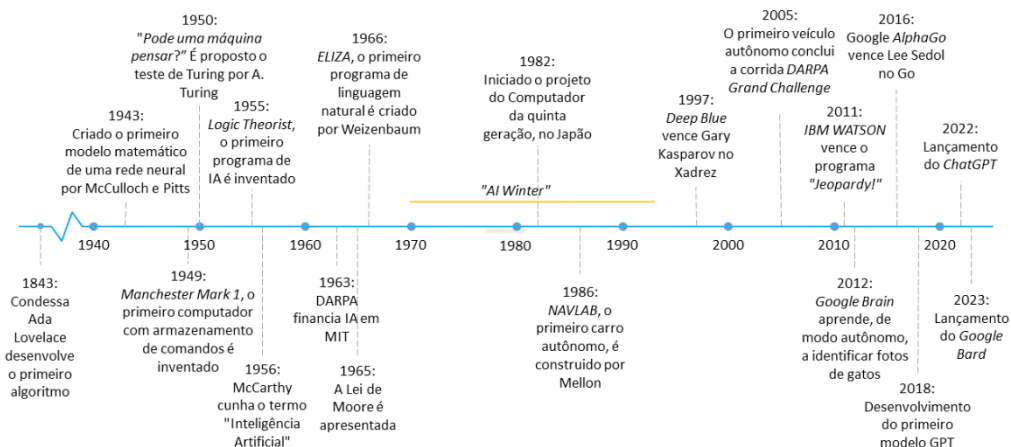
Com a evolução da internet e dos microprocessadores, a inteligência artificial ressurgiu nos anos 2000 como uma força proeminente no cenário tecnológico e científico, transformando-se em um campo fértil para estudos e pesquisas e uma área de grande interesse público e do mercado (Peixoto; Silva, 2019). A partir daí, deu-se início à primavera da IA, período marcado por avanços significativos, tal como a vitória do *Deep Blue*¹² sobre o campeão mundial de xadrez, Garry Kasparov, em 1997, e pelo retorno dos investimentos na área. A partir da década de 2010, o campo da IA passou a experimentar um dos seus verões, marcado e impulsionado pela ascensão “dos

¹¹ Trata-se de artigo acadêmico produzido pelo matemático James Lighthill na forma relatório, cujo objetivo era avaliar a pesquisa acadêmica na área de inteligência artificial. Disponível em: https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf. Acesso em: 18 jun. 2025.

¹² Sistema de computação complexo projetado pela IBM para jogar xadrez. Contava com 256 processadores trabalhando em conjunto, capaz de avaliar cerca de 200 milhões de posições de xadrez por segundo. Vide: <https://www.britannica.com/topic/Deep-Blue>. Acesso em: 24 jun. 2025.

dados (Big Data) e o aumento na capacidade de armazenamento desses, somado ao aperfeiçoamento dos algoritmos, sequência de códigos que automatizam decisões, fatores que proporcionaram o desenvolvimento da aprendizagem profunda” (Junquilha, 2023, p. 23). A figura abaixo representa a evolução da IA.

Figura 1 – Linha do tempo da IA com base em eventos históricos



Fonte: Chahad; Issa, 2024, p. 22.

Atualmente, o campo da IA vive um dos seus períodos mais promissores e de altos investimentos em pesquisas e no desenvolvimento de máquinas e sistemas sofisticados, sem qualquer indício de desaceleração. No entanto, há de se pontuar que, conforme mostra a história, o movimento da IA “não é retilíneo, tendo sofrido invernos e períodos de relativo desprestígio” (Junquilha, 2023, p. 23). É, assim, antes de mais nada, um movimento cíclico e que depende de fatores favoráveis para florescer, tais como aplicações práticas amplas, apoio governamental e regulatório, e investimentos massivos.

Nos próximos capítulos, serão destrinchados os aspectos essenciais do funcionamento da IA, a começar pelo seu principal “combustível”: os dados pessoais.

2.2. Dados pessoais e big data

A origem dos dados pessoais se confunde com a história da humanidade. Muito antes da internet e da automação das decisões, as pessoas já reuniam e avaliavam dados brutos em busca de padrões. Eram os primórdios da estatística relacionada a

temas inerentemente humanos, surgidos da análise detalhada de dados pessoais para revelar comportamentos, realidades e anseios sociais.

Desde sempre, os dados pessoais têm servido para moldar ações governamentais, impulsionar o desempenho de agentes econômicos, verificar a aceitação de produtos e serviços perante a população, avaliar candidatos a vagas de empregos, entre outras aplicações. A evolução humana e tecnológica permitiu refinar a obtenção de dados pessoais e a interpretação das informações neles contidas. Contudo, em essência, seu propósito inicial se mantém inalterado, e eles continuam sendo ferramentas valiosas para a análise de diferentes cenários inferidos a partir de modelos matemáticos (Lindoso, 2019).

Um efeito direto do aprimoramento na coleta de dados é que qualquer informação ligada a uma pessoa específica passa a ser vista como dado pessoal. Isso abrange desde informações como nome, data de nascimento, sexo e raça/etnia, até o tempo gasto em redes sociais, interações online e pesquisas em motores de busca, como o Google. A quantidade de informações que podem ser obtidas na internet e que são consideradas dados pessoais é vasta, à medida que relacionam ações ou características a um indivíduo em particular.

Numa concepção restrita, um dado pessoal pode ser definido como a descrição de ocorrências relacionadas a um indivíduo identificado, ou seja, informações que se referem a alguém que é conhecido e distinguível dentro de um grupo ou da coletividade (Machado, 2023). Nesse caso, a identificação acontece por meio de dados específicos, conhecidos como identificadores, que têm uma ligação bem forte e pessoal com cada indivíduo. Esses identificadores podem ser classificados como diretos e/ou indiretos.

Nessa perspectiva, identificadores diretos são informações que, por si só, permitem a identificação imediata de uma pessoa natural sem a necessidade de dados adicionais. São dados que possuem uma conexão direta e inequívoca com o titular, como o nome completo, número de CPF, RG, número de passaporte, endereço de e-mail, número de telefone associado a um indivíduo específico, entre outros. Já os identificadores indiretos são informações que, isoladamente, não identificam uma pessoa, mas, quando combinadas com outros dados ou dentro de um contexto, podem permitir a sua identificação. É necessário, no entanto, um processo de cruzamento ou análise para estabelecer a conexão com o titular, sendo o caso do endereço IP, da localização geográfica, dos padrões de comportamento online (como

histórico de navegação), das características demográficas (idade, gênero, profissão), ou mesmo identificadores técnicos, como cookies ou IDs de dispositivos.¹³

De outro lado, a concepção ampla se estende para além da simples identificação de uma pessoa natural, englobando também informações de natureza pessoal que dizem respeito a indivíduos identificáveis. Por esse ponto de vista, dados pessoais não se restringem à presença de identificadores diretos que distinguem um indivíduo em particular. Dados que são suscetíveis de levar à identificação de uma pessoa também são considerados pessoais. Essa abordagem abrange não apenas identificadores diretos, mas também identificadores indiretos, desde que seja possível, com meios razoavelmente disponíveis, associar esses dados a uma pessoa específica (Machado, 2023).

É seguro dizer que o ordenamento jurídico brasileiro trilhou um caminho semelhante ao da União Europeia¹⁴ ao incorporar uma definição ampla de dado pessoal. De acordo com o artigo 5º, inciso I, da Lei Geral de Proteção de Dados (LGPD), o conceito de dado pessoal no contexto da proteção de dados no Brasil se refere a “qualquer informação que possa identificar ou tornar identificável uma pessoa natural”. Essa definição legal apenas reforça o que a Lei de Acesso à Informação (Lei n. 12.527/2011) já estabelecia, em seu artigo 4º, inciso IV, ao definir informação pessoal como “aquela que se relaciona a uma pessoa física identificada ou que possa ser identificada”.

O legislador ordinário em um esforço para delimitar a natureza jurídica do dado pessoal, aproximou-a da corrente que defende ser uma extensão da personalidade do indivíduo¹⁵ – e que, por assim ser, relaciona-se diretamente com direitos fundamentais tais como o direito à privacidade e à liberdade –, contrastando com a

¹³ Em consonância com o Considerando 30 do GDPR. União Europeia. Regulamento (UE) nº 2016/679 do Parlamento Europeu e do Conselho, de 23 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Jornal Oficial da União Europeia, Estrasburgo, 04/05/2016. Disponível em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&from=PT>>. Acesso em: 4 fev. 2025.

¹⁴ Nos termos do art. 4º da GDPR, dado pessoal é definido como “informação relativa a uma pessoa singular identificada ou identificável («titular dos dados»); é considerada identificável uma pessoa singular que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrónica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular”.

¹⁵ Vide os arts. 1º e 2º da LGPD que dispõem, respectivamente, que o objetivo da lei consiste em “proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural” e que a disciplina de proteção de dados pessoais tem como fundamentos, entre outros, o respeito à privacidade, a autodeterminação informativa e o livre desenvolvimento da personalidade.

corrente oposta que o compreende como um bem jurídico. Um expoente dessa última corrente é o renomado teórico Lawrence Lessing, que, em 1998, já enunciava que:

Dados são um ativo. São um recurso que se tornou extremamente valioso. E, à medida que se tornaram extremamente valiosos, o comércio tentou explorá-los. Esse uso tem um custo — uma externalidade suportada por aqueles que preferem que esses dados não sejam usados. Portanto, o truque é construir um regime em que aqueles que usariam os dados internalizem esse custo. Um regime que garanta que eles paguem por esse custo. As leis de propriedade são um desses regimes. Se os indivíduos pudessem receber o direito de controlar seus dados, ou mais precisamente, se aqueles que usariam os dados tivessem primeiro que garantir o direito de usá-los, então uma negociação poderia ocorrer sobre se, e em que quantidade, os dados deveriam ser usados. O mercado, isto é, poderia negociar esses direitos, se um mercado para esses direitos pudesse ser construído (Lessing, 1998, p. 17, tradução nossa).

Conforme evidencia Amália Batocchio, a preferência na classificação do dado pessoal enquanto propriedade e, portanto, bem jurídico, é temporal e concentrada sobretudo na literatura dos Estados Unidos, sendo poucos os trabalhos contemporâneos que endossam essa visão (Batocchio, 2023). Dessa forma, ao associarem características e qualidades a um determinado usuário, os dados pessoais delineiam sua essência no ciberespaço e, conseqüentemente, remetem à sua personalidade. O usuário, por sua vez, passa a possuir todos os atributos inerentes à condição humana, protegidos no ordenamento jurídico brasileiro sob o manto dos direitos de personalidade (Lindoso, 2019).

A evolução dessa corrente proporcionou intensos debates¹⁶ que culminaram na constitucionalização do direito à proteção de dados a partir da introdução do art. 5º, inciso LXXIX, à Constituição Federal. Surge daí um direito fundamental que é compreendido como parte integrante da personalidade e, por conseguinte, é intransmissível, irrenunciável e não pode sofrer limitação voluntária em seu exercício. Além disso, não há como negar que a preocupação com a privacidade foi um fator crucial para a criação desse direito fundamental, notadamente porque os dados pessoais são regularmente tratados em contextos de transações comerciais como meras moedas de troca, quando, na realidade, afetam de maneira direta a vida das pessoas.

¹⁶ No Brasil, a história da proteção de dados pessoais remonta, de forma concreta, à consulta pública realizada no ano de 2010 pelo Ministério da Justiça sobre o anteprojeto de lei, a qual culminou na promulgação do Marco Civil da Internet, em 2014, e na promulgação da Lei Geral de Proteção de Dados Pessoais (LGPD), em 2018, com entrada em vigor em 2020.

Vale dizer que, antes mesmo de ter sido reconhecido como direito fundamental com a efetiva inclusão na Constituição Federal através da Emenda Constitucional n. 115/2022, o Supremo Tribunal Federal (STF) atento às discussões no mundo jurídico, como foco especialmente nas implicações sobre o direito à privacidade, à liberdade e ao livre desenvolvimento da personalidade, já acenava nesse sentido. Na Ação Direta de Inconstitucionalidade (ADI) n. 6387 MC-REF/DF – julgada juntamente com as ADI's 6.388, 6.389, 6.390 e 6.393 –, visava-se a declaração de inconstitucionalidade do inteiro teor da Medida Provisória n. 954, de 17 de abril de 2020, a qual dispunha sobre “o compartilhamento de dados por empresas de telecomunicações prestadoras de Serviço Telefônico Fixo Comutado e de Serviço Móvel Pessoal com a Fundação Instituto Brasileiro de Geografia e Estatística, para fins de suporte à produção estatística oficial durante a situação de emergência de saúde pública de importância internacional decorrente do coronavírus (covid19), de que trata a Lei nº 13.979, de 6 de fevereiro de 2020”¹⁷. Oportunamente, ao enfrentar o mérito do caso, o Plenário da Corte se debruçou sobre a questão relativa à proteção de dados pessoais e sua relação com garantias constitucionais, fazendo constar que:

Tais informações, relacionadas à identificação – efetiva ou potencial – de pessoa natural, configuram dados pessoais e integram, nessa medida, o âmbito de proteção das cláusulas constitucionais assecuratórias da liberdade individual (art. 5º, caput), da privacidade e do livre desenvolvimento da personalidade (art. 5º, X e XII). Sua manipulação e tratamento, desse modo, hão de observar, sob pena de lesão a esses direitos, os limites delineados pela proteção constitucional.

Decorrências dos direitos da personalidade, o respeito à privacidade e à autodeterminação informativa foram positivados, no art. 2º, I e II, da Lei nº 13.709/2018 (Lei Geral de Proteção de Dados Pessoais), como fundamentos específicos da disciplina da proteção de dados pessoais.

No clássico artigo *The Right to Privacy*, escrito a quatro mãos pelos juízes da Suprema Corte dos Estados Unidos Samuel D. Warren e Louis D. Brandeis, já se reconhecia que as mudanças políticas, sociais e econômicas demandam incessantemente o reconhecimento de novos direitos, razão pela qual necessário, de tempos em tempos, redefinir a exata natureza e extensão da proteção à privacidade do indivíduo. Independentemente do seu conteúdo, mutável com a evolução tecnológica e social, no entanto, permanece como denominador comum da privacidade e da autodeterminação o entendimento de que a privacidade somente pode ceder diante de justificativa consistente e legítima. Em seus dizeres, “a invasão injustificada da privacidade individual deve ser repreendida e, tanto quanto possível, prevenida”.

¹⁷ Brasil. Medida Provisória nº 954, de 17 de abril de 2020. **Dispõe sobre o compartilhamento de dados por empresas de telecomunicações prestadoras de Serviço Telefônico Fixo [...]**. Brasília, DF, [2020]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/mpv/mpv954.htm. Acesso em: 02 jan. 2025.

O entendimento exarado pelo STF se alinha à conceituação de dado pessoal como uma extensão dos direitos de personalidade.¹⁸ Tal definição objetiva assegurar que a coleta e o tratamento observem as garantias do indivíduo, incluindo o direito à privacidade. Isso é fruto, conforme bem pontuado no julgado acima transcrito, da evolução tecnológica e social e dos desdobramentos do direito à privacidade ao longo da história. Até o século XIX, o direito à privacidade era intrinsecamente vinculado à noção de propriedade privada, marcado por um viés individualista, e a partir da obra seminal de Warren e Brandeis (1890), o seu sentido foi drasticamente alterado durante o século XX para abranger o controle dos usuários sobre suas informações e de sua esfera privada (Rodotà, 2008). Portanto, falar sobre dado pessoal é falar, invariavelmente, sobre privacidade, autodeterminação informativa e direitos de personalidade.

Diante da conceituação de dado pessoal, é fundamental reconhecer que, como consequência, se estabelece uma demarcação a respeito do que não caracteriza um dado pessoal. Se os dados não se referem a uma pessoa natural identificada ou identificável, seja desde a sua origem ou após um tratamento, eles são classificados como dados anônimos ou anonimizados. Não obstante os termos sejam usados com frequência de forma intercambiável, há relevante distinção que importa aos estudos sobre tecnologias digitais. Anônimos são os dados que, desde o momento da coleta, não contêm informações que permitam identificar uma pessoa específica. Ou seja, nunca houve vínculo com dados pessoais identificados ou identificáveis. Dados anonimizados são, por sua vez, aqueles que, inicialmente pessoais, passaram por um processo de tratamento que aplicou diversas técnicas e padrões a fim de eliminar seus identificadores. Consequentemente, esses dados, em tese, perdem sua natureza pessoal.

Tal classificação e distinção é importante porque o exercício de direitos e garantias asseguradas legal e constitucionalmente são reservados aos titulares de dados pessoais, ficando de fora do amparo da legislação sobre proteção de dados, como a LGPD, os dados não pessoais. Essa exclusão impede que os indivíduos adquiram direitos relacionados a esses dados, bem como que os agentes de

¹⁸ Vale enfatizar que, desde a constitucionalização do direito à proteção de dados pessoais, o STF ainda não explorou a questão a fundo e apenas a enfrentou de forma reflexa (Sarlet, Ingo Wolfgang. **O STF e a Proteção de Dados Pessoais**. Disponível em: <https://www.conjur.com.br/2025-mar-03/o-stf-e-a-protecao-dos-dados-pessoais/>. Acesso em: 16 mar. 2024).

tratamento assumam situações jurídicas passivas tais como obrigações e encargos ou que se submetam à regulamentação de proteção de dados (Machado, 2023).

Conforme prescrito na LGPD, os dados anonimizados não devem ser capazes de ser associados a uma pessoa identificada ou identificável de forma permanente e irreversível. Em caso de reversão do processo de anonimização pelo qual os dados foram submetidos, não seria correto referir-se a esses dados como não pessoais. Com o avanço das técnicas e das pesquisas centradas na reversão do processo de anonimização, “pode-se afirmar a existência de um consenso científico sobre a impraticabilidade de um cenário de ausência de risco de reidentificação nas situações de tratamento de dados anonimizados” (ANPD, 2023).¹⁹ Isso se deve ao imenso volume de dados disponíveis publicamente, principalmente na internet, ao avanço da capacidade de processamento de sistemas computacionais e aos refinados algoritmos para reidentificação.

Os grandes volumes de dados, conhecidos como Big Data, são os propulsores do campo relacionado à inteligência artificial, pois possibilitam o tratamento e a unificação de diversos tipos de informações em um sistema unificado. A sua principal vantagem está na agilidade e na eficiência de custos que proporciona ao lidar com esses dados. A coleta de dados brutos visando a identificação de padrões é uma atividade antiga, a qual foi otimizada com o passar do tempo a partir da noção de que toda informação pode ser objeto de análise, assim como quase tudo pode ser transformado em dado, igualmente passível de análise para encontrar padrões. No entanto, a capacidade de interpretar grandes quantidades de dados durante muito tempo foi limitada, restringindo a precisão de estudos estatísticos mais abrangentes (Mayer-Schonberger; Cukier, 2013). Em sua essência,

big data relaciona-se com previsões. Apesar de ser descrito como um ramo da ciência da computação chamado inteligência artificial e, mais especificamente, uma área chamada “aprendizado de máquina”, esta ideia é enganosa. Big data não tem a ver com tentar “ensinar” um computador a “pensar” como ser humano. Ao contrário, trata-se de aplicar a matemática a enormes quantidades de dados a fim de prever probabilidades: a chance de

¹⁹ Vide: Sweeney, Latanya. **Simple Demographics Often Identify People Uniquely**. Carnegie Mellon University, Data Privacy Working Paper 3, Pittsburgh, 2000; Narayanan, Arvind; Shmatikov, Vitaly. **How to break anonymity of the Netflix Prize dataset**. 2007. Disponível em: <https://arxiv.org/pdf/cs/0610105>. Acesso em: 10 mar. 2025; Wondracek, Gilbert et al. **A practical attack to de-anonymize social network users**. Proceedings – IEEE Symposium on Security and Privacy, p. 223–238, 2010; De Montjoye, Yves A. et al. **Unique in the Crowd: The privacy bounds of human mobility**. Scientific Reports, v. 3, p. 1–5, 2013; Rocher, L.; Hendrickx, J. M.; De Montjoye, Y. A. **Estimating the success of reidentifications in incomplete datasets using generative models**. Nature Communications, v. 10, n. 1, 2019.

um e-mail ser um spam; de as letras “msa” na verdade significarem “mas”; de a trajetória e velocidade de uma pessoa que atravessa a rua significarem que ela a atravessará a tempo – o carro com piloto automático precisa reduzir apenas um pouco a velocidade. O segredo é que esses sistemas funcionam porque são alimentados por enormes quantidades de dados, que formam a base das previsões. Além disso, os sistemas são criados para se aperfeiçoarem com o tempo, ao continuamente analisar os melhores sinais e padrões a fim de encontrar mais dados para uso (Mayer-Schonberger; Cukier, 2013, p. 8).

Com essa compreensão, uma parcela da comunidade científica e tecnológica se dedicou a construir métodos para interpretar dados com maior agilidade, buscando transcender a capacidade humana e estabelecer processamento em larga escala. Entre meados e o fim do século XX, o campo científico-tecnológico experienciou avanços expressivos, com o surgimento dos primeiros bancos de dados e um aumento na produção de computadores, que se tornaram extremamente populares nas décadas subsequentes. Naquela época, o mundo passava por uma transição para o meio digital, o que contribuiu para o aumento da produção de dados e incentivou o progresso na análise de dados por meio de computadores. Paralelamente, a internet surgiu, tornando o acesso à comunicação por meio de uma plataforma unificada cada vez mais popular.

Essa revolução digital redefiniu a operação dos mercados, moldando-os como os conhecemos hoje e culminando em um aumento notável na produção de dados. Em outras palavras, além da constatação de que diversos conteúdos poderiam ser submetidos a análises matemáticas, o volume de dados gerados aumentou devido à maior disponibilidade dos dispositivos que os produziam – neste caso, os computadores. Essa mudança foi impulsionada pelo big data, que abriu caminho para a modernização da tecnologia, resultando no desenvolvimento da inteligência artificial e de sistemas automatizados, e para a produção de conhecimento através de processos envolvendo algoritmos.

2.3. Algoritmos e aprendizado de máquina

Os algoritmos são as fórmulas matemáticas que tornam possível a organização da automação de processos e a análise do big data, de forma a criar previsões e, por conseguinte, tomar decisões de forma automatizada. É “comumente descrito como um conjunto de instruções, organizadas de forma sequencial, que determina como algo deve ser feito” (Mendes; Mattiuzzo; Fujimoto, 2021, p. 429) e

remonta à história da matemática e da computação (Silva, 2020), logo, não deve ser apreendido como um conceito que depende do uso do computador moderno, pois “nada mais [é] que uma fórmula na qual tarefas são colocadas em uma ordem específica para atingir determinado objetivo” (Mendes; Matiuzzo, 2019, p. 19).

No entanto, há de se reconhecer que “conceituá-lo como um conjunto de etapas definidas para produzir uma solução específica, é simplificar muito” (Lage, 2020, p. 35). Isso porque, “como a maioria das ferramentas relacionadas à matemática, ela começa bastante simples, mas se torna infinitamente complexa quando expandida. A diferença entre um algoritmo e um programa geralmente é uma questão de nível de detalhe” (Lage, 2020, p. 35).

Tais construções matemáticas são, essencialmente, traduções formais dos anseios e pretensões humanas. Elas se organizam com base em padrões que se repetem, analisando os dados que são introduzidos inicialmente (*inputs*) para gerar um resultado específico (*output*). Os algoritmos, presentes em diversos setores, são, no contexto do estudo de estruturas repetitivas, a engrenagem que possibilita identificar e compreender um padrão a ser seguido. Ao entender a estrutura do modelo, ou seja, seus padrões, o algoritmo produz uma resposta que, teoricamente, representa a melhor forma de executar a tarefa designada.

A análise preditiva e a automatização do processo de tomada de decisão se referem, portanto, à habilidade matemática de examinar um grande volume de dados, identificar semelhanças, criar um modelo e fornecer, rapidamente, uma visão estatisticamente embasada sobre qual resultado tem maior probabilidade de sucesso. É essa a função do algoritmo. Sendo o algoritmo um modelo matemático, é necessário que haja treinamento contínuo – aqui entendido como uma espécie de estímulo – para identificar padrões e interpretar dados relevantes para decisões específicas. Geralmente, cientistas da computação, matemáticos e profissionais de áreas correlatas fazem isso usando técnicas de aprendizado que replicam informações diretamente no algoritmo.

Uma dessas técnicas é o aprendizado de máquina (ou *machine learning*, em inglês), que consiste em um processo sistemático de construção, ajuste e validação de modelos computacionais que aprendem a partir de dados para resolver problemas complexos. Esse processo envolve a integração de dados, escolha de representações, definição de critérios de desempenho e aplicação de técnicas de otimização, com o objetivo final de alcançar generalização. Segundo Caitlin

Mulholland, por meio da técnica de aprendizado de máquina “a IA desenvolve a aptidão para ampliar experiências, aferindo delas conhecimentos, por meio de um ciclo contínuo e crescente de aprendizagem. A IA, portanto, só é plenamente eficiente porque se apropria desses métodos de análise de dados” (Mulholland, 2021, p. 331). Já na definição de Ethem Alpaydin:

Aprendizado de máquina é programar computadores para otimizar um critério de desempenho usando dados de exemplo ou experiência passada. Temos um modelo definido até alguns parâmetros, e aprendizado é a execução de um programa de computador para otimizar os parâmetros do modelo usando os dados de treinamento ou experiência passada. O modelo pode ser preditivo para fazer previsões no futuro, ou descritivo para obter conhecimento a partir de dados, ou ambos.

O aprendizado de máquina usa a teoria da estatística na construção de modelos matemáticos, porque a tarefa principal é fazer inferência a partir de uma amostra. O papel da ciência da computação é duplo: primeiro, no treinamento, precisamos de algoritmos eficientes para resolver o problema de otimização, bem como para armazenar e processar a enorme quantidade de dados que geralmente temos. Segundo, uma vez que um modelo é aprendido, sua representação e solução algorítmica para inferência também precisam ser eficientes. Em certas aplicações, a eficiência do algoritmo de aprendizado ou inferência, ou seja, sua complexidade de espaço e tempo, pode ser tão importante quanto sua precisão preditiva (Alpaydin, 2014, pp. 3-4, tradução nossa)

O aprendizado de máquina opera identificando padrões, o que permite gerar resultados e construir modelos através do processamento automatizado. Em síntese, o AM busca resolver problemas práticos através do seguinte esquema: (1) obtenção de um conjunto de dados; (2) construção algorítmica de um modelo estatístico com base nesse conjunto de dados; (3) treinamento do modelo construído e, por fim, (4) resolução do problema prático (Lage, 2020, p. 66). Para que essa capacidade atinja níveis mais complexos e sofisticados, é crucial combinar dois elementos: um volume massivo de dados – viabilizado por meio do big data – e um poder de computação significativamente expressivo (Kremer, 2021).

Os tipos de aprendizado de máquina se dividem entre o supervisionado, o não supervisionado e o por reforço. O aprendizado supervisionado é a forma mais simples de AM e envolve treinar um sistema inteligente para realizar tarefas específicas usando dados rotulados, ou seja, exemplos onde a resposta correta já é conhecida. Cada exemplo de entrada vem com uma saída esperada (rótulo) e o objetivo é aprender uma função que mapeie entradas para saídas corretas. No caso, por exemplo, da construção de um sistema de IA que classifica imagens (casa, carro, pessoas, etc.): primeiro, coleta-se um grande conjunto de dados (*dataset*) com as

imagens de casas, carros e pessoas, cada uma rotulada com sua categoria correspondente e, durante o treinamento, o sistema recebe uma imagem como entrada e gera uma saída na forma de um vetor de pontuações, com um valor para cada categoria possível, visando que a categoria correta receba a pontuação mais alta (Lecun et al., 2015). Esse tipo de aprendizado “se vale das técnicas de regressão (que inclui predição, árvores de decisão, redes bayesianas, regressão linear, dentre outras) e classificação (que abrange regressão lógica, redes neurais artificiais, dentre outras)” (Junquilha, 2023, p. 26).

De outro lado, no aprendizado não supervisionado o modelo trabalha com dados não rotulados, ou seja, sem saídas predefinidas, e seu objetivo é encontrar padrões, estruturas ou agrupamentos nos dados. Aqui “só são conhecidos os dados de entrada, mas ainda não se sabe quais são as variáveis de saída correspondentes” e, ao contrário aprendizado supervisionado, “não há respostas corretas e não há um treinamento dos dados feito por humanos: os algoritmos são deixados por conta própria para descobrirem e apresentarem a estrutura mais interessante nos dados” (Lage, 2020, p. 71). Essa é justamente a principal vantagem do aprendizado não supervisionado, ele é útil para explorar dados desconhecidos e, conseqüentemente, para identificar padrões desconhecidos nos dados. Esse tipo de aprendizado se vale “de técnicas para segmentação (que compreende técnicas como algoritmos genéticos, por exemplo) e para redução de dimensão (que utiliza técnicas como o aprendizado múltiplo, análise fatorial, projeções aleatórias, dentre outras)” (Junquilha, 2023, p. 27).

Por fim, no aprendizado por reforço o modelo (agente) aprende por tentativa e erro, interagindo com um ambiente e recebendo recompensas – ou penalidades – por suas ações. Nesse tipo, o agente toma decisões, observa os resultados (recompensas) e ajusta sua política de ações com base no feedback do ambiente. O objetivo é maximizar a recompensa acumulada ao longo do tempo. O aprendizado por reforço “é encontrado, por exemplo, em carros não tripulados, que por meio de um complexo sistema de sensores de suporte são capazes de percorrer estradas urbanas e não urbanas, reconhecendo obstáculos, seguindo as placas e muito mais” (Lage, 2020, p. 72).

Ainda no campo do aprendizado de máquina, convém mencionar a existência da aprendizagem profunda (ou *deep learning*, em inglês), subdomínio do AM “que consiste em múltiplas camadas em cascata, modeladas a partir do sistema

nervoso humano (uma prática denominada codificação neural), conhecida como rede neural articular” (Lage, 2020, p. 27) – a intenção aqui é replicar o trabalho do cérebro humano (sistema nervoso) em máquinas usando camadas de neurônios. Yann LeCun, Yoshua Bengio e Geoffrey Hinton explicam que:

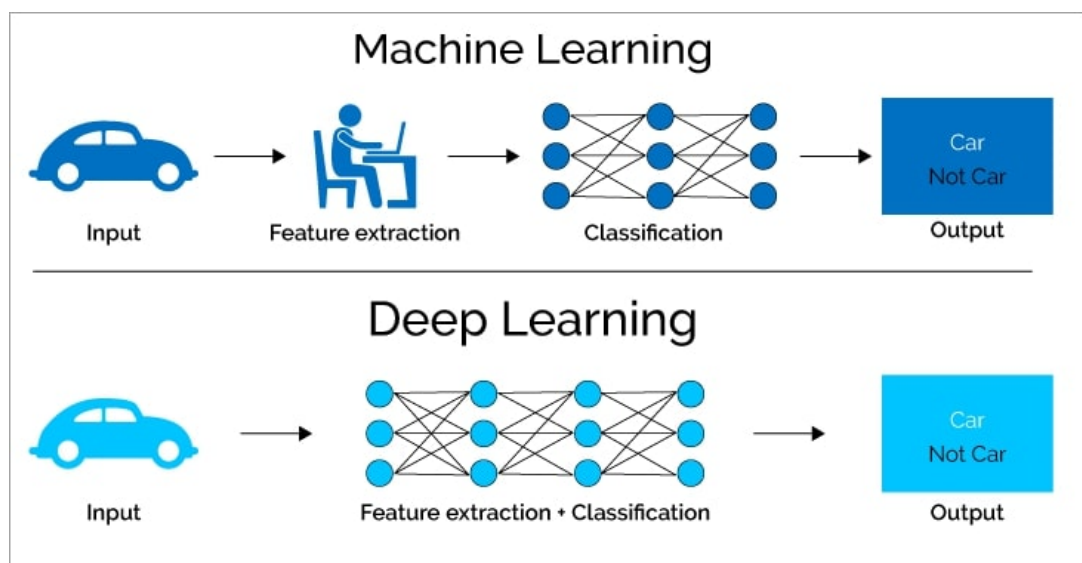
Métodos de aprendizado profundo são métodos de aprendizado de representação com múltiplos níveis de representação, obtidos pela composição de módulos simples, porém não lineares, que transformam cada um a representação em um nível (começando com a entrada bruta) em uma representação em um nível superior, ligeiramente mais abstrato. Com a composição de transformações suficientes, funções muito complexas podem ser aprendidas. Para tarefas de classificação, camadas superiores de representação amplificam aspectos da entrada que são importantes para a discriminação e suprimem variações irrelevantes. Uma imagem, por exemplo, vem na forma de uma matriz de valores de pixels, e as características aprendidas na primeira camada de representação normalmente representam a presença ou ausência de bordas em orientações e locais específicos na imagem. A segunda camada normalmente detecta motivos identificando arranjos específicos de bordas, independentemente de pequenas variações nas posições das bordas. A terceira camada pode reunir motivos em combinações maiores que correspondem a partes de objetos familiares, e as camadas subsequentes detectariam objetos como combinações dessas partes. O aspecto fundamental do aprendizado profundo é que essas camadas de características não são projetadas por engenheiros humanos: elas são aprendidas a partir de dados usando um procedimento de aprendizado de propósito geral (Lecun et al., 2015, p. 436, tradução nossa).

A base e a razão de ser da aprendizagem profunda são as redes neurais artificiais (RNA), as quais operam através das interações de bilhões de neurônios artificiais que estão interconectados. O funcionamento da AP compreende uma rede neural composta por três tipos de camadas: camada de entrada (*input layer*), camadas ocultas (*hidden layers*) e camada de saída (*output layer*). A camada de entrada recebe um conjunto de dados brutos – como pixels de uma imagem, palavras em um texto ou valores numéricos – os quais são passados para as camadas ocultas, que dentro do modelo de RNA são o centro da rede profunda. Cada neurônio em uma camada oculta realiza cálculos simples e, à medida que os dados passam por camadas mais profundas, a rede extrai características cada vez mais abstratas. Por exemplo, em uma imagem, as primeiras camadas detectam bordas ou cores, enquanto camadas mais avançadas identificam formas complexas, como olhos ou rostos inteiros. A camada de saída gera o resultado final, que pode ser uma classificação (ex.: casa, pessoa, pet, etc.) ou uma previsão numérica (ex.: valor de uma casa ou de uma passagem aérea).

As RNA são treinadas para reconhecer padrões ajustando conexões com base em exemplos. Nesse processo, os dados entram na rede como números e passam por camadas, onde pesos (valores que definem a força das conexões) geram uma previsão. No treinamento, a rede compara sua previsão com a resposta correta usando uma função de custo, que mede o erro. Esse erro é usado para ajustar os pesos por meio da otimização por gradiente, que calcula como mudar cada peso para reduzir o erro, como descer uma colina. A taxa de aprendizado controla o tamanho desses ajustes e o processo se repete por várias iterações, refinando os pesos com muitos exemplos até a rede acertar padrões complexos, como formas ou contextos.

A AP é amplamente utilizada em três tipos de tecnologias distintas: reconhecimento de fala, visão computacional e processamento de linguagem natural (PLN). A imagem abaixo reproduzida ilustra de forma bastante elucidativa o funcionamento do aprendizado de máquina e da aprendizagem profunda:

Figura 2 – Estrutura de processamento no *machine learning* e no *deep learning*



Fonte: Sperling, 2018.

O AM e a AP compartilham do mesmo objetivo, que é permitir que sistemas inteligentes aprendam com dados. A diferença entre AM e AP, por sua vez, está na forma como cada um aprende e o nível de autonomia. Enquanto o AM se vale de algoritmos “para analisar dados, aprender com esses dados e tomar decisões informadas com base no que aprendeu”, a AP “estrutura algoritmos em camadas para criar uma rede neural artificial – uma tentativa de reproduzir o trabalho do cérebro

humano em um computador usando camadas de neurônios - que pode aprender e tomar decisões inteligentes por conta própria” (Lage, 2020, p. 75).

A aplicação da técnica de aprendizado de máquina e de aprendizado profundo revela um aspecto relevante dos algoritmos, com especial relevância para este estudo, que é sua capacidade de atuação autônoma. Em conjunto com o big data, e considerando a imensidão de dados analisados, é comum dizer que os algoritmos geram resultados imprevisíveis – e, portanto, fora do controle de seu desenvolvedor, assim como do público –, independentemente das intenções do programador. Nesse cenário, desponta um subcampo da IA que tem gerado, simultaneamente, entusiasmo e preocupação quanto ao seu uso indiscriminado e em larga escala: a inteligência artificial generativa (IAGen).

2.4. Inteligência artificial generativa: da criação ao ineditismo

O atual estágio em que se encontra o campo da inteligência artificial é decorrência direta da revolução tecnológica ora em curso, ou seja, da Quarta Revolução Industrial – ou Indústria 4.0. Diversamente dos eventos ocorridos nas revoluções industriais anteriores — com a chegada das máquinas a vapor no século XVIII, a produção em larga escala e a distribuição da energia elétrica no século XIX e a automação no século XX —, essa nova fase é marcada fortemente pela onipresença de tecnologias da informação como a própria inteligência artificial, internet das coisas (IoT), big data, robótica, blockchain e computação em nuvem. Doneda *et al.* destaca que se trata de uma revolução que “alimentada por dados, está eliminando as fronteiras entre a Física e a Biologia” (2018, p. 10).

Klaus Martin Schwab explica que essa revolução é única por três motivos: ela acontece muito rápido, com inovações surgindo numa velocidade impressionante e nunca antes vista; afeta praticamente tudo, de fábricas a hospitais, escolas e até governos; e muda sistemas inteiros, desde como as empresas funcionam até o jeito que as pessoas trabalham e se relacionam com a tecnologia (Schwab, 2016). O autor faz menção a sistemas ciberfísicos, onde o mundo real e o digital se misturam, como em indústrias superconectadas ou cidades inteligentes.

Fala-se, atualmente, na transição para a Indústria 5.0 ou Quinta Revolução Industrial, ideia esta que, primeiro, foi apresentada por Michael Rada em 2015 e melhor desenvolvida pelo próprio em 2017. Segundo Rada, essa revolução é marcada

pela evolução sustentável e humana da indústria, centrada no *Industrial Upcycling*²⁰ para prevenir resíduos, promover colaboração entre humanos e tecnologia no mundo físico, e garantir benefícios ambientais, sociais e econômicos sem excluir as pessoas do processo. Ele a vê como essencial para um futuro "sem resíduos", onde a eficiência e a sinergia substituem o desperdício e a automação excessiva. Assim, a Indústria 5.0 é definida pela combinação entre tecnologia e potencial humano, com foco no ser humano, ao invés da automação e “vem com a evolução da quarta Revolução Industrial com o propósito de unir o homem e máquina por sistemas inteligentes para aumentar a eficiência de processos no trabalho” (Bittencourt; Okumura, 2020, p. 3).

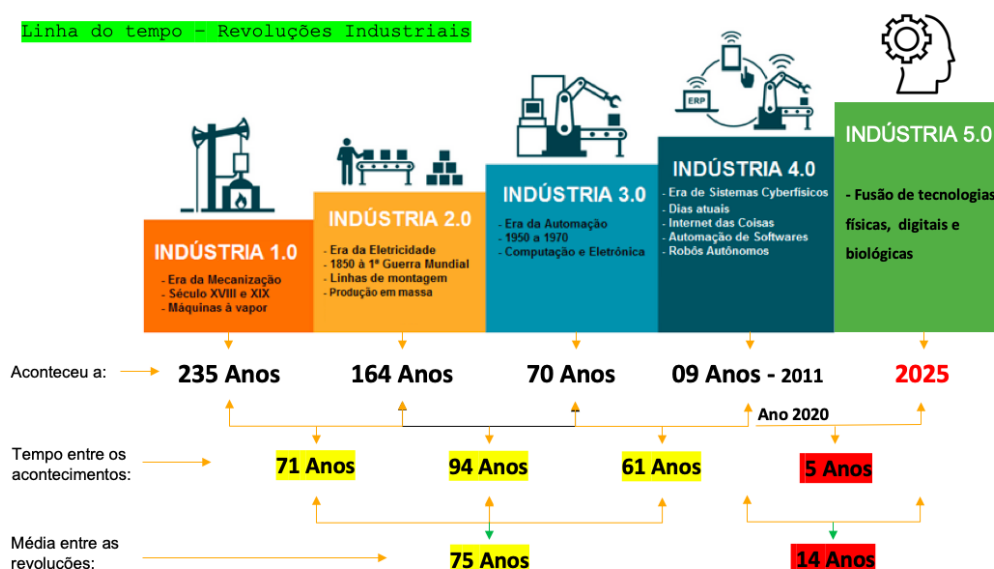
A União Europeia sedimentou o conceito de Indústria 5.0 ao defini-la como “uma visão da indústria que visa além da eficiência e da produtividade como únicos objetivos e reforça o papel e a contribuição da indústria para a sociedade”, que “coloca bem-estar do trabalhador no centro do processo de produção e usa novas tecnologias para proporcionar prosperidade além de empregos e crescimento, respeitando os limites de produção do planeta”.²¹ Desta forma, não se trata da superação da Indústria 4.0 – afinal, não representa propriamente um salto tecnológico –, mas antes de sua complementação. O foco aqui é na personalização em massa, sustentabilidade, centralidade humana, bem-estar dos trabalhadores e, por consequência, nos limites planetários (AKUNDI et al., 2022). A imagem abaixo apresenta uma linha do tempo das revoluções industriais e destaca as principais características de cada e o período de duração:

Figura 3 – Linha do tempo das revoluções industriais

²⁰ Na definição de Rada, é uma metodologia central da Indústria 5.0 que visa a prevenção sistemática de resíduos em processos industriais, transformando materiais, produtos e subprodutos em novos recursos antes que se tornem lixo. Diferentemente da reciclagem, que muitas vezes envolve processos de degradação de materiais, o *upcycling industrial* busca reutilizar e revalorizar materiais em sua forma original ou com maior valor, sem necessidade de grandes investimentos iniciais, ao contrário da economia circular.

²¹ European Commission. **Industry 5.0**. Disponível em: https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en. Acesso em: 4 jul. 2025.

Linha do tempo - Revoluções Industriais



Fonte: Lilla, 2023.

Em meio a tais transformações na indústria e na sociedade, foi lançado ao fim do ano de 2022 para o público geral o *Chat Generative Pre-trained Transformer* (ChatGPT)²², da empresa OpenAI, o que desencadeou inúmeras discussões sobre as aplicações da inteligência artificial, especialmente sobre a modalidade generativa (RAMOS, 2023). A existência de *chatbots* como o ChatGPT não é necessariamente uma novidade, pois desde de meados do século XX já existia o Teste de Turing e o ELIZA, mas seu diferencial reside justamente no uso de tecnologias de ponta para compreensão contextual profunda para geração criativa de texto. Isto é, enquanto os primeiros *chatbots* funcionavam com base em regras pré-definidas e usavam de truques de programação para passar uma ilusão de compreensão, os modelos atuais não são programados com regras, mas sim treinados com quantidades massivas de texto da internet e, por esse motivo, passam uma ideia de compreensão muito mais convincente e natural. Em pesquisa realizada no ano de 2024 pela McKinsey & Company, constatou-se que 65% (sessenta e cinco por cento) das empresas usam IAGen no mundo²³, sendo esta uma forte tendência para o presente e futuro.

²² Muniz, Carolina. **Robô 'ultrainteligente' que usa IA responde a dúvidas e redige textos sozinho.** Folha de S. Paulo, São Paulo, 5 dez. 2022. Disponível em: <https://www1.folha.uol.com.br/blogs/hashtag/2022/12/robo-ultrainteligente-que-usa-ia-responde-a-duvidas-e-redige-textos-sozinho.shtml>. Acesso em: 2 jul. 2025.

²³ Cafferata, Pepe; Dias, Yran Bartolomeu. **65% das empresas usam Gen AI no mundo; líderes apontam caminhos para obter retorno financeiro.** Disponível em: <https://www.mckinsey.com.br/our-insights/all-insights/65-das-empresas-usam-gen-ai-no-mundo>. Acesso em: 12 jul. 2025.

Mas, afinal, o que é IA generativa? Em termos objetivos, “é uma tecnologia de inteligência artificial (IA) que gera conteúdo de forma automática em resposta a comandos escritos em interfaces de conversação em linguagem natural” (UNESCO, 2024, p. 8). De modo oposto aos modelos preditivos de aprendizado de máquina, “centrados em extrair padrões de dados e fazer previsões em tarefas específicas –, [a IAGen] produz conteúdo original a partir de grandes bases de dados, ou seja, usa dados para gerar mais dados, sintetizando texto, imagem, voz, vídeo, códigos (Santanella; Kaufman, 2024, p. 39). Apesar da capacidade de gerar conteúdo novo e inédito, a IAGen “não pode gerar novas ideias ou soluções para desafios do mundo real, pois não compreende objetos ou relações sociais do mundo real que sustentam a linguagem” (UNESCO, 2024, p. 8).

As RNAs, referidas anteriormente, são o alicerce técnico dessas capacidades, pois permitem modelar relações complexas e probabilísticas nos dados. Na IA generativa, as RNAs são projetadas para aprender distribuições de dados e gerar amostras novas que seguem essas distribuições. Elas operam em aprendizagem profunda permitindo a captura de padrões complexos em grandes volumes de dados, como bilhões de palavras ou imagens. As RNAs são organizadas em arquiteturas específicas²⁴, cada uma otimizada para diferentes tipos de dados ou tarefas. As principais são: as redes adversariais generativas (GANs, do inglês *Generative Adversarial Networks*), os modelos baseados em transformador (em inglês, *Transformer*) e os autocodificadores variacionais (VAEs, do inglês *Variational Autoencoders*)

Propostas em 2014 por Ian Goodfellow e outros pesquisadores da Universidade de Montreal no artigo intitulado “*Generative Adversarial Networks*”, as GANs funcionam com duas redes neurais: o gerador, que produz dados falsos (ou sintéticos²⁵) – como imagens ou áudios – a partir de ruído aleatório²⁶, e o discriminador, que avalia se os dados são reais ou falsos, como um crítico. O gerador se esforça para produzir dados que possam enganar o discriminador, fazendo-o acreditar que são reais. A esse processo dá-se o nome de “adversário” e ele perdura até que o gerador consiga criar dados que sejam indistinguíveis dos dados reais do

²⁴ Dentro do universo da IA generativa, arquitetura refere-se à organização dos componentes das redes neurais, isto é, neurônios artificiais, camadas e conexões.

²⁵ Dados sintéticos são dados gerados artificialmente, em contraste com os dados reais que são oriundos da realidade (AEPD, 2023).

²⁶ Refere-se a um conjunto de dados iniciais, geralmente números aleatórios, usados como ponto de partida para o gerador criar conteúdos sintéticos, como imagens, áudios ou outros tipos de dados.

conjunto de treinamento (Goodfellow et al., 2014). Essa dinâmica ajuda ambas as redes a aprimorar suas habilidades – à medida que o gerador refina suas criações para enganar o discriminador, que, por sua vez, se torna mais eficaz em identificar dados falsos –, resultando em dados gerados que são mais realistas e de melhor qualidade. Apesar de eficazes para gerar imagens realistas, como faces sintéticas ou *deepfakes*, Goodfellow (2016) destaca que as GANs podem ser instáveis e produzir resultados repetitivos, um problema chamado *mode collapse*.

Os *Transformers*, introduzidos em 2017 por Ashish Vaswani e outros pesquisadores no artigo intitulado “*Attention Is All You Need*”, são a base de modelos que geram textos, como os modelos de linguagem de grande escala (LLMs, do inglês *Large Language Models*). O nome dado ao artigo faz menção direta ao mecanismo de atenção que analisa todas as partes de uma frase de uma vez, decidindo quais palavras são mais importantes, similar à tarefa manual de um leitor que destaca trechos-chave de um texto. No modelo baseado em transformador, as redes neurais funcionam a partir da assimilação do contexto e do significado através da análise das interações de dados sequenciais (Vaswani et. al, 2017), o que indica que esses modelos se destacam em atividades de processamento de linguagem natural, como tradução automática, geração de perguntas e respostas e modelagem da linguagem. Tais modelos, que são a espinha dorsal das LLMs, têm sido empregados em modelos de linguagem amplamente conhecidos, como o GPT-4 (*Generative Pre-trained Transformer 4*).

Assim como as arquiteturas precedentes, os VAEs foram apresentados por meio de artigo publicado em 2014 e intitulado “*Auto-Encoding Variational Bayes*”, de autoria dos pesquisadores da Universidade de Amsterdã Diederik P. Kingma e Max Welling. Os VAEs usam redes neurais para transformar dados em uma representação compacta, chamada espaço latente, e gerar variações a partir dela. Esses modelos se assemelham aos GANs, pois operam com duas redes neurais distintas (codificadores e decodificadores), todavia, estas não estão competindo e formam uma única rede com duas partes. A primeira, os codificadores, é responsável por comprimir grandes quantidades de dados em uma representação menor na forma de números (como um resumo), e a segunda, os decodificadores, usa esses números para recriar ou inventar dados novos semelhantes aos originais (Kingma; Welling, 2014). Os VAEs são frequentemente aplicados na criação de imagens, vídeos e sons. A tabela 1

abaixo disposta sintetiza as principais técnicas utilizadas em IA generativa e suas aplicações.

Tabela 2 – Técnicas utilizadas em IA generativa

Aprendizado de máquina (ML)		Um tipo de IA que utiliza dados para aprimorar automaticamente seu desempenho.
Rede neural artificial (RNA)		Um tipo de ML inspirado na estrutura e funcionamento do cérebro humano (por exemplo, as conexões sinápticas entre neurônios).
IA generativa de texto	Transformadores de propósito geral	Um tipo de RNA que é capaz de focar em diferentes partes de dados para determinar como elas se relacionam entre si.
	Modelos de linguagem grandes (LLM)	Um tipo de transformador de propósito geral que é treinado com grandes quantidades de dados textuais.
	Transformador generativo pré-treinado (GPT)⁴	Um tipo de modelo de linguagem grande (LLM) que é pré-treinado em quantidades ainda maiores de dados, o que permite ao modelo capturar as nuances da linguagem e gerar textos coerentes e de acordo com o contexto.
IA generativa de imagem	Redes generativas adversariais (GANs)	Tipos de rede neurais usados para geração de imagens.
	Autoencoders variacionais (VAEs)	

Fonte: UNESCO, 2024. p. 8.

Ainda que em estágios iniciais, a IA generativa tem transformado diversos setores com aplicações práticas. Na saúde, a NVIDIA Clara analisa dados para diagnósticos médicos e descobertas farmacêuticas. No marketing, ferramentas como Jasper e albert.ai geram conteúdos personalizados em forma de texto e de conteúdo para publicidade digital. Na arte, o Midjourney cria imagens e modelos 3D. Nas finanças, a IA prevê cenários, gerencia riscos e detecta fraudes. Em pesquisas na internet, o Bing da Microsoft – integrado com tecnologia da OpenIA – entrega respostas em linguagem natural, facilitando a compreensão e interação com os resultados. No campo da acessibilidade, a IA promove inclusão com tecnologias como transcrição de fala e comandos por voz. Essas aplicações destacam o potencial da IA generativa para promover eficiência, inclusão e inovação.

De outro lado, no entanto, são muitos os riscos e efeitos negativos que impactam pessoas, em nível individual e coletivo, sociedades, economias e o meio ambiente. Para citar alguns exemplos, a IAGen é conhecida por gerar conteúdo falso, ofensivo ou manipulador, como *deepfakes* (vídeos e/ou áudios falsos), alucinações (informações inventadas apresentadas como fatos) ou instruções prejudiciais para os usuários. A esse respeito, a própria OpenIA reconhece que o GPT-4 possui limitações similares aos modelos anteriores de GPT e, por essa razão, “ele ainda não é

totalmente confiável ('alucina' fatos e comete erros de raciocínio). Deve-se ter muito cuidado ao usar os resultados do modelo de linguagem, especialmente em contextos de alto risco".²⁷

No campo legal, o uso de dados protegidos por direitos autorais no treinamento de modelos de IA generativa levanta questões concernentes à propriedade intelectual, criando incertezas sobre a propriedade de conteúdos gerados com base em materiais já existentes e protegidos por lei. Há também os riscos associados à privacidade e à proteção de dados pessoais, uma vez que esses modelos podem expor informações contidas em seus dados de treinamento, como dados pessoais identificáveis e dados pessoais sensíveis (e.g., dados de saúde, dados biométricos, e informações que revelam origem racial ou étnica, convicções religiosas ou políticas, etc.).

Do ponto de vista econômico, a automação de tarefas criativas, como redação, design e codificação, ameaça empregos em setores como jornalismo, marketing e artes, gerando desigualdades e exigindo requalificação da força de trabalho. Por sua vez, o desenvolvimento de modelos de IA generativa depende diretamente de uma mão de obra mal remunerada e frequentemente explorada, composta por trabalhadores em países em desenvolvimento que recebem salários baixos e são submetidos a jornadas de trabalho intensas. O impacto ambiental é outro ponto crítico, já que o treinamento e a operação desses modelos necessitam de alta capacidade computacional e, por consequência, consomem grandes quantidades de energia e água para resfriamento de servidores, contribuindo para emissões de carbono e escassez hídrica.

Por fim, a falta de transparência e explicabilidade dos modelos, que muitas vezes operam como caixas-pretas, dificulta a compreensão de suas decisões e a responsabilização dos agentes envolvidos. Além disso, os modelos podem perpetuar vieses presentes nos dados de treinamento, como preconceitos raciais, de gênero ou sociais, gerando saídas discriminatórias que agravam desigualdades em áreas como recrutamento, justiça ou saúde, impactando comunidades vulneráveis e violando direitos fundamentais, conforme será apresentado a seguir.

²⁷ OpenIA. **GPT-4**. Disponível em: <https://openai.com/index/gpt-4-research/>. Acesso em: 18 jul. 2025.

3. RACISMO ALGORÍTMICO E PERFILAMENTO

3.1. Vieses discriminatórios

Na era em que algoritmos moldam desde escolhas cotidianas até decisões de impacto social, a promessa de uma inteligência artificial neutra e eficiente se confronta com uma realidade incômoda: a reprodução de padrões históricos de desigualdade. Ao tratar sobre os efeitos do uso massivo de sistemas de IA no livro “Algoritmos de Destruição em Massa”, a matemática Cathy O’Neil apresenta o conceito do que ela própria denomina de Armas de Destruição Matemática – ADM (do inglês, *Weapons of Math Destruction* – WMD). Segundo a autora, uma AMD é marcada por três características principais: opacidade, escala e dano (O’neil, 2020). Esses elementos convergem para transformar um modelo algorítmico em uma ADM.

A opacidade dos algoritmos de IA e a sua forma de funcionamento pode ser vista como benéfica aos agentes econômicos, à medida que impõe verdadeira dificuldade para o monitoramento dessas estruturas. Isso ocorre porque, em razão da opacidade, os agentes podem proteger suas fórmulas algorítmicas através do manto do segredo de negócios e sob o pretexto de que há confusão entre o código de programação – *inputs* de dados, componentes, propriedades do código, cálculos, variáveis, etc. – e as informações que são protegidas por lei, ou seja, em tese, não seria possível discernir onde termina um e onde começa o outro. Assim, sob a ótica econômica e concorrencial, há inúmeras vantagens na manutenção da confidencialidade de informações sobre como é construído e como funciona o algoritmo/sistema de IA.

Diante da insegurança e das incertezas se intensificaram os debates sobre a importância de transparência e explicabilidade dos sistemas de IA, o que culminou no surgimento do campo de pesquisa denominado *Explainable AI* (XAI). Seu objetivo consiste em “tornar o comportamento desses sistemas compreensível para os humanos, elucidando os mecanismos subjacentes aos seus processos de tomada de decisão” (EDPS, 2023, p. 3, tradução nossa). Para tanto, idealmente, “o XAI deve incluir a capacidade de explicar as competências e entendimentos do sistema, explicar suas ações passadas, processos em andamento e etapas futuras, e divulgar as informações relevantes nas quais suas ações se baseiam” (EDPS, 2023, p. 4, tradução nossa).

A principal justificativa é que os sistemas de IA funcionam como verdadeiras "caixas pretas", das quais os usuários não têm controle algum. Divulgados como sistemas naturalmente incompreensíveis, eles acabam sendo frequentemente manobrados por empresas em busca de ganhos e da consolidação de seu domínio no mercado de dados. Segundo Frank Pasquale, a sociedade está imersa na mencionada "caixa preta", onde os algoritmos ditam os rumos da vida das pessoas, sem que elas entendam completamente como isso acontece (Pasquale, 2015). A "caixa preta" detém um conhecimento vasto e armazena tudo, funcionando como um mecanismo de vigilância ininterrupta. Assim, de modo pouco ou nada perceptível, os algoritmos são responsáveis por moldar identidades e influenciar os desejos e pensamentos humanos.

A ideia de "caixa preta" evoca a falta de informações claras sobre o funcionamento dos algoritmos de IA, sobre seus elementos confidenciais, não revelados, invisíveis ao público ou acessíveis a um grupo restrito. Não há dúvidas de que esses elementos impactam a sociedade, indo de encontro aos interesses dos indivíduos e da coletividade e, às vezes, dos próprios objetivos divulgados publicamente pelas empresas. Por exemplo, uma pessoa pode ser rejeitada em um processo seletivo inclusivo por um algoritmo de recrutamento sem entender o motivo, já que os cálculos e as variáveis envolvidas não são explicados. A ininteligibilidade do sistema algoritmo de IA dificulta a contestação de decisões e a identificação de possíveis vieses, como discriminações raciais embutidas no sistema. Por essa razão, Pasquale questiona até que ponto convém obstar o acesso a esses sistemas sob o argumento de proteger o segredo de negócios e o interesse público, particularmente quando os interesses em questão são de agentes governamentais e econômicos. E vai além, ao indagar de que forma é possível regular o campo tecnológico quando seu funcionamento não é totalmente compreendido pelo ser humano e há esforços contínuos para blindá-lo do escrutínio público.

É notável que as *big techs* – as denominadas gigantes da tecnologia, como Microsoft, Apple, Meta, Amazon, X (antigo Twitter), OpenAI e Google – controlam uma enorme parcela do poder tecnológico global, a partir da concentração de conhecimento técnico e do monopólio de mercado. Essas grandes companhias influenciam na formação do pensamento coletivo a respeito da relação entre tecnologia e sociedade, atuando como intermediárias através da coleta massiva de dados. As vastas quantidades de dados disponibilizados voluntaria ou

involuntariamente pelos usuários são exploradas ativamente pelas empresas com o propósito de obter vantagens e impulsionar a venda de seus produtos e serviços, numa estratégia que visa o controle absoluto do mercado. Como consequência, são tomadas decisões com repercussão sobre toda a sociedade sem qualquer consulta ou supervisão. A professora Shoshana Zuboff chama esse fenômeno de capitalismo de vigilância, que consiste num modelo econômico baseado na coleta, análise e comercialização de dados pessoais em larga escala (Zuboff, 2019). Por meio desse processo, empresas transformam experiências humanas — como comportamentos, preferências, desejos, hábitos e até emoções — em dados brutos, que são então tratados, usados para predições e comercializados como produtos no mercado.

Segundo o professor Pierri Lévy, o poder das *big techs* transcende a esfera econômica, estendendo-se ao campo político, uma vez que diversas funções sociais e de infraestrutura, que tradicionalmente e até então eram exclusivas dos Estados-nações, agora estão sob o controle dessas corporações.²⁸ As *big techs* formam o que ele chama de "Estados-plataforma", pois controlam infraestruturas essenciais, detêm vastas quantidades de dados e influenciam a esfera pública — superando as fronteiras nacionais e a geopolítica, e, por conseguinte, estabelecendo-se como um poder que abrange todo o planeta. Nisso consiste a escala referenciada por O'neil: a capacidade dos modelos algorítmicos de IA, como os desenvolvidos pelas *big techs*, de impactar um grande número de pessoas ou situações de forma simultânea e em níveis distintos — econômico, político, social, etc.

Nessa conjuntura, é fundamental considerar que, desde a concepção, por trás da criação e evolução de sistemas algorítmicos, existem interesses de setores econômicos, mesmo que não evidentes. Os impactos observados nas áreas política e social indicam a formação de um discurso baseado na suposição de que, não obstante as intenções dos agentes econômicos e governamentais, algoritmos de IA são neutros, objetivos e desprovidos de valores e ideologias. A ideia de neutralidade tecnológica emerge da visão de que o conhecimento técnico é produzido e disseminado em apartado da sociedade, sendo desprovido de valores e influências externas, apresentando uma lógica funcional própria. Assim, seriam as tecnologias não apenas neutras, mas também autônomas e isentas de vieses. Contudo, trata-se

²⁸ Fernandes, Daniela. '**Gigantes da web são novo Estado**', diz Pierre Lévy. Valor Econômico. 23 de Out. 2020. Disponível em: https://valor.globo.com/eu-e/noticia/2020/10/23/gigantes-da-web-sao-novo-estado-diz-pierre-levy.ghtml?utm_source=chatgpt.com. Acesso em: 16 abr. 2025.

de um mito fundado na hierarquização do saber científico e na supervalorização do conhecimento quantitativo (Frazão, 2021), que oculta valores e interesses que são, em essência, humanos e sociais.

Andrew Feenberg defende que a tecnologia é um artefato sociocultural que reflete os valores da sociedade industrial, especialmente os das classes dominantes que se manifestam através do domínio da técnica, e, por essa razão, não é neutra (Feenberg, 1991). Em sentido análogo e em crítica ao determinismo tecnológico²⁹, Langdon Winner sustenta que artefatos técnicos — máquinas, infraestruturas, sistemas — possuem qualidades políticas que moldam as relações de poder e de autoridade na sociedade. Por assim serem, tecnologias não são neutras, uma vez que incorporam escolhas sociais que refletem interesses de determinados grupos, influenciando diretamente como as pessoas vivem, trabalham, consomem e interagem (Winner, 1980).

A neutralidade tecnológica é parte do arcabouço ideológico do Vale do Silício, que se alinha a uma visão tecnolibertária e neoliberal³⁰ que difunde a ideia de que os resultados – sejam positivos ou negativos – dependem apenas de como a tecnologia é usada, não de como é projetada ou por quem. No entanto, como visto, essa ideia é fortemente rechaçada pelo campo da Ciência, Tecnologia e Sociedade (CTS), que compreende a tecnologia como um artefato cultural, moldado pelas intenções, valores e preconceitos de seus criadores – neste caso, uma elite predominantemente branca, masculina e privilegiada. A suposta neutralidade é, portanto, e antes de mais nada, uma fachada que protege as empresas de tecnologia das críticas sobre os produtos e serviços distribuídos em larga escala e, por consequência, da responsabilização e da prestação de contas (*accountability*), especialmente quando estes amplificam discriminações ou lucram com elas (Noble; Roberts, 2020).

²⁹ A teoria do determinismo tecnológico estabelece que a tecnologia e seus avanços são os principais motores das transformações históricas e sociais. Parte-se da premissa que a tecnologia não é apenas um elemento essencial para a evolução da humanidade e da sociedade, mas também uma força inovadora e autônoma. Nesse sentido, os aspectos humanos e sociais são vistos como menos importantes, já que a visão predominante sugere que a tecnologia não pode ser controlada; ao contrário, ela influencia os caminhos da história, bem como os comportamentos e os valores humanos. Vide: Corrêa, Raquel Folmer.; Geremias, Bethania Medeiros. **Determinismo tecnológico: elementos para debates em perspectiva educacional**. Revista Tecnologia e Sociedade. v. 9, n. 18, 2013. Disponível em: <https://periodicos.utfpr.edu.br/rts/article/view/2633/1750#>. Acesso em: 23 nov. 2024.

³⁰ Morente, Jaime Caro. **The Silicon (Valley) Doctrine. Las ideologías de las Big-Tech**. El Viejo Topo. Disponível em: <https://www.elviejotopo.com/articulo/the-silicon-valley-doctrine-las-ideologias-de-las-big-tech/>. Acesso em: 3 dez. 2024.

O manifesto contra a diversidade de James Damore³¹, ex-engenheiro de software do Google, exemplifica bem essa lógica. Damore defendia que iniciativas de diversidade eram "discriminatórias" com homens brancos, insinuando que a tecnologia deveria funcionar em um ambiente neutro, sem influências relacionadas a raça ou gênero. Todavia, essa perspectiva se mostra historicamente equivocada, pois desconsidera séculos de exclusão estrutural de minorias, que a tecnologia, ao invés de corrigir, muitas vezes reproduz. A noção de neutralidade, nesse contexto, serve como uma ferramenta retórica para preservar o *status quo*, impedindo que as empresas reconheçam as consequências de suas decisões sobre design e recrutamento.

Algoritmos, plataformas e dispositivos carregam inscrições algorítmicas e genéticas de raça (Noble; Roberts, 2020), ou seja, refletem os vieses de seus criadores, os dados com os quais são treinados e o ambiente ao qual foi exposto, evidenciando que tais vieses não são acidentais, mas estruturais, resultado de uma força de trabalho homogênea e de uma ideologia que nega a influência da raça no design tecnológico. Dentro do campo da IA, os vieses podem ser compreendidos sob três diferentes perspectivas: estatística, legal/jurídica e social/cognitiva. Em sistemas técnicos, o viés tende a ser tratado como um fenômeno estatístico e representa “um efeito que priva um resultado estatístico de representatividade, distorcendo-o sistematicamente, diferentemente de um erro aleatório, que pode distorcer em uma ocasião, mas se equilibra na média” (Schwartz et. al, 2022, p. 3, tradução nossa). Deduz-se daí que os vieses da inteligência artificial são falhas que ocorrem reiteradamente durante o uso de algoritmos de IA, as quais produzem resultados que não correspondem com a realidade ou com o comportamento desejado. Nesse contexto, diz-se que um sistema de IA é tendencioso quando exhibe comportamento sistematicamente impreciso.

Ainda que tecnicamente precisa, essa definição não abrange nem comunica suficientemente todo o espectro de riscos representados pelo viés em sistemas de IA, razão pela qual é complementada pelas outras duas perspectivas. Sob o ponto de vista jurídico,

³¹ Conger, Kate. **Exclusive: Here's The Full 10-Page Anti-Diversity Screed Circulating Internally at Google [Updated]**. Gizmodo, 5 ago. 2017. Disponível em: <<https://gizmodo.com/exclusive-heres-the-full-10-page-anti-diversity-screed-1797564320>>. Acesso em: 21 mar. 2025.

O viés discriminatório inadmissível geralmente é definido pelos tribunais como consistindo em tratamento diferenciado, amplamente definido como uma decisão que trata um indivíduo de forma menos favorável do que indivíduos em situação semelhante devido a uma característica protegida, como raça, sexo ou outra característica, ou como impacto diferenciado, que é amplamente definido como uma política ou prática aparentemente neutra que prejudica desproporcionalmente um grupo com base em uma característica protegida (Schwartz et. al, 2022, p. 4, tradução nossa).

O tratamento desigual, em si, não configura discriminação que é vedada pelo ordenamento jurídico pátrio, uma vez que se admite a chamada discriminação positiva – que nada mais é do que “favorecer de modo diferenciado aqueles que estejam em situações de indevida desvantagem social (os fragilizados, os oprimidos, as “minorias”) ou impor um gravame maior aos que estejam numa situação de exagerada vantagem social” (Rothenburg, 2009, p. 81). Não se admite, no entanto, a dita discriminação ilícita, pois vedada expressamente por lei, e abusiva, pois resulta em restrições indevidas de direitos, tratamento desigual ou impactos negativos desproporcionais aos indivíduos, sem que haja uma proibição expressa e categórica na lei (Mendes; Mattiuzzo; Fujimoto, 2021).

Na perspectiva social, o foco é direcionado às equipes que estão envolvidas no projeto e desenvolvimento de sistemas de IA e que transferem seus vieses cognitivos, tanto individuais quanto de grupo, para o processo. Assim, “o viés é predominante nas suposições sobre quais dados devem ser usados, quais modelos de IA devem ser desenvolvidos, onde o sistema de IA deve ser implantado — ou se a IA é realmente necessária (Schwartz et. al, 2022, p. 5, tradução nossa). Em nível institucional, os vieses sistêmicos influenciam a forma como organizações e equipes são estruturadas e quem controla os processos de tomada de decisão.

Assim, o viés não é apenas um problema isolado de natureza técnica, legal ou social, mas uma interação dinâmica que amplifica danos em múltiplas escalas quando não mitigado. E é esse o último elemento da ADM proposta por O’neil, qual seja, o dano causado pelas decisões algorítmicas que reforçam desigualdades ou produzem injustiças. Em estudo sobre LLMs realizado em 2023, pesquisadores da Universidade de Washington identificaram e quantificaram vieses contra 93 (noventa e três) grupos estigmatizados (e.g., estigmas relacionados a etnia, doenças, deficiências, uso de drogas, saúde mental, religião, sexualidade, status socioeconômico, etc.) (Mei; Fereidooni; Caliskan, 2023). Tal número, que não é exaustivo, indica uma variedade grande de grupos estigmatizados que sofrem com os

vieses presentes em sistemas de IA e, conseqüentemente, sugere uma abordagem multidisciplinar.

3.2. Racismo algorítmico

Em maio de 2025, a nutricionista e pesquisadora Bruna Oliveira, conhecida nas redes sociais como Bruna Crioula, fez uma denúncia pública contra o LinkedIn, apontando a ocorrência de discriminação algorítmica por parte da plataforma.³² Com um currículo sólido, incluindo formação em nutrição, mestrado em ciências sociais e atuação profissional em nutrição ecológica, Bruna fez uso da plataforma com o objetivo de encontrar vaga de emprego em sua área de formação, todavia, passou a receber apenas recomendações para a área de serviços gerais, a qual não guarda qualquer relação com o seu perfil. Assim, não obstante o registro de pedido formal de explicações para o LinkedIn, a jovem relata que recebeu resposta genérica, na qual a empresa limitou-se a justificar a lógica de seus algoritmos – sem, contudo, apresentar informações que ajudassem a entender o porquê do resultado em completo descompasso com o seu perfil.

O algoritmo de recrutamento do LinkedIn trabalha com base na compreensão pormenorizada do perfil do usuário e das vagas disponíveis através da análise de dados feita por um sistema de inteligência artificial. Ainda que os detalhes desse sistema e de seu funcionamento não sejam de conhecimento público, é notório que são consideradas as informações constantes no perfil do usuário, tais como experiência profissional, formação acadêmica, competências, localização geográfica, além das interações prévias na plataforma. O mesmo é feito em relação às vagas disponíveis, de onde são extraídos os dados sobre os requisitos exigidos por quem está contratando, a descrição do cargo, a localização da empresa, a modalidade do emprego (se presencial, híbrido ou remoto), entre outras informações pertinentes. E, assim,

Com base nesses dados, o algoritmo emprega técnicas de aprendizado de máquina para estabelecer correlações entre candidatos e vagas. Por exemplo, o algoritmo pode empregar algoritmos de classificação para determinar a pertinência de uma vaga específica para um determinado candidato, levando em conta suas competências e experiências. Ele pode

³² Crioula, Bruna. **Quando o algoritmo é cruel: uma mulher negra e a violência silenciosa do LinkedIn**. Disponível em: <https://brunacrioula.medium.com/quando-o-algoritmo-%C3%A9-cruel-uma-mulher-negra-e-a-viol%C3%Aancia-silenciosa-do-linkedin-d387a614d3aa>. Acesso em: 28 jun. 2025.

também analisar padrões nos dados para identificar quais competências são mais valorizadas por certos empregadores ou setores da indústria. Ainda, o algoritmo considera o envolvimento do usuário com a plataforma. Se um usuário demonstra interesse em vagas de um setor específico ou interage regularmente com conteúdo relacionado a uma determinada área profissional, o algoritmo pode priorizar vagas nessa área em suas recomendações. (Barros, 2024, pp. 20-21)

A denúncia formalizada por Bruna a respeito da microagressão racial sofrida pelo LinkedIn (Silva, 2020) dialoga diretamente com reportagem feita pela Bloomberg a respeito de imagens geradas pelo Stable Diffusion (da Stability AI) a partir de comandos de texto.³³ Assim como muitos modelos de IA generativa, o conteúdo inédito criado pelo Stable Diffusion parece ser confiável e guardar relação com a realidade, mas foi revelado que ele leva as disparidades raciais a extremos. Na análise feita pela Bloomberg, o Stable Diffusion foi utilizado para criar representações de trabalhadores para 14 empregos — 300 imagens para cada um, representando sete empregos tipicamente considerados "bem remunerados"³⁴ nos EUA e sete considerados "mal remunerados".³⁵

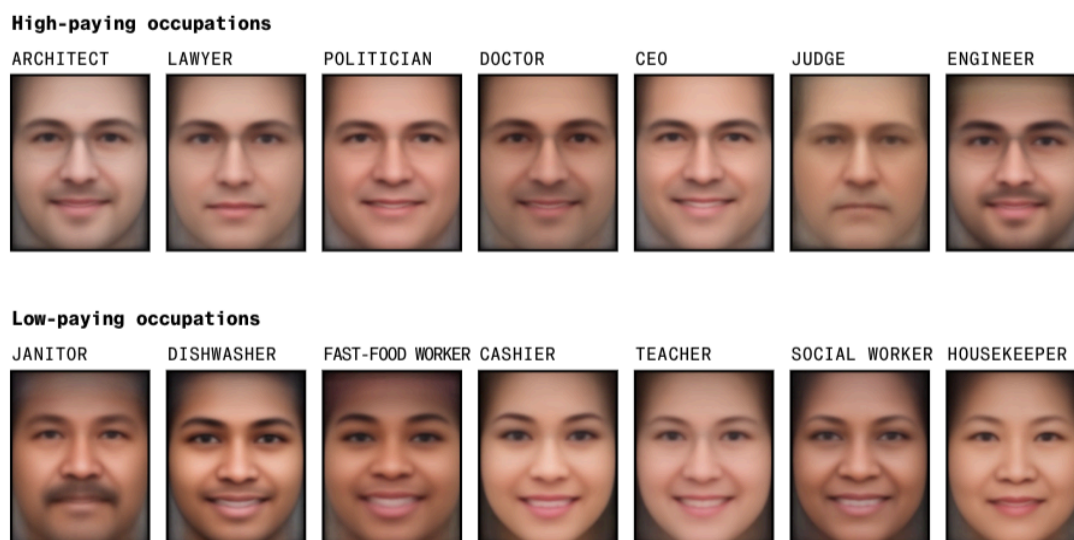
No total foram analisadas 5.100 imagens de pessoas geradas por inteligência artificial e, para cada imagem, foi calculada uma cor média tendo por parâmetro a pele do rosto. Com base nessa cor média, cada rosto foi classificado em uma das seis categorias de pigmentação da pele, conforme definido pela Escala de Pele Fitzpatrick (um sistema usado por médicos dermatologistas e pesquisadores). A pesquisa revelou que os conjuntos de imagens criadas para cada trabalho bem remunerado eram dominados por pessoas de pele clara, enquanto aquelas com pele mais escura eram frequentemente associadas a *prompts* como "trabalhador de fast-food" e "assistente social" (Figura 4).

Figura 4 – Representação gráfica do “rosto” dos trabalhos mais bem e mal remunerados

³³ Nicoletti, Leonardo; Bass, Dina. **Humans Are Biased. Generative AI Is Even Worse.** Bloomberg, 09 Jun. 2023, disponível em <<https://www.bloomberg.com/graphics/2023-generative-ai-bias/>>. Acesso em: 23 abr. de 2025.

³⁴ Entre as sete profissões mais bem remuneradas listadas estão: Arquiteto, advogado, político, médico, diretor executivo/geral (CEO), juiz e engenheiro.

³⁵ Entre as sete profissões mais mal remuneradas listadas estão: zelador, lavador de pratos, trabalhador de fast-food, caixa, professor, assistente social e empregada doméstica.



Fonte: Bloomberg, 2023.³⁶

À primeira vista, seria até possível pensar que as imagens geradas refletem apenas um problema estrutural da sociedade referente à desigualdade racial e de gênero no mercado de trabalho. No entanto, quando cruzados os resultados das imagens criadas com os dados fornecidos pelo Departamento de Estatísticas do Trabalho dos EUA, verificam-se imprecisões na representação do que seria o “rosto” de cada uma das ocupações foco da análise. Por exemplo, o modelo gerou imagens de pessoas com tons de pele mais escuros 70% do tempo para a palavra-chave “funcionário de fast-food”, embora 70% dos funcionários de fast-food nos EUA sejam brancos. Da mesma forma, 68% das imagens geradas de assistentes sociais tinham tons de pele mais escuros, enquanto 65% dos assistentes sociais dos EUA são brancos. Os resultados gerados pelo Stable Diffusion são representações de estereótipos raciais, mas de maneira ainda mais distorcida da realidade.

Casos do gênero não são novos. Em 2016, a ProPublica publicou artigo revelando a existência de vieses raciais no algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*)³⁷, usado nos EUA para prever o risco de reincidência criminal. A análise mostrou que o algoritmo, desenvolvido pela Northpointe (agora Equivant), classificava desproporcionalmente pessoas negras

³⁶ Nicoletti, Leonardo; Bass, Dina. **Humans Are Biased. Generative AI Is Even Worse.** Bloomberg, 9 jun. 2023, disponível em <<https://www.bloomberg.com/graphics/2023-generative-ai-bias/>>. Acesso em: 23 abr. de 2025.

³⁷ Angwin, Julia; Larson, Jeff; Mattu, Surya; Kirchner, Lauren. **Machine Bias.** ProPublica, 23 mai. 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 21 mar. 2025.

como de alto risco de reincidência, mesmo quando não reincidiam, em comparação com pessoas brancas. Especificamente, 45% dos negros que não reincidiram foram erroneamente classificados como alto risco, contra 23% dos brancos. Em resposta, a Northpointe publicou um artigo técnico³⁸ defendendo a sua abordagem e destacando que o COMPAS foi projetado para prever a probabilidade de reincidência em três categorias de risco (baixo, médio e alto), e não para determinar sucesso ou falha de maneira direta, como sugerido pela ProPublica. Além disso, enfatizou que o COMPAS deveria ser usado como uma ferramenta auxiliar, não como o único critério para decisões judiciais, e que os juízes deveriam considerar outros fatores.

Naquele mesmo ano, Joy Buolamwini, cientista da computação do Instituto de Tecnologia de Massachusetts (MIT), denunciou a presença de vieses raciais em algoritmos de reconhecimento facial através de uma TED Talk³⁹ e, posteriormente, no documentário *Coded Bias* (2020), de Shalini Kantayya.⁴⁰ Ela descobriu que sistemas de inteligência artificial comerciais, como os usados para identificar rostos, apresentavam taxas de erro significativamente maiores para pessoas de pele escura, especialmente mulheres negras, devido a conjuntos de dados enviesados e falta de diversidade nos processos de desenvolvimento. Em 2018, Joy publicou em parceria com a cientista de dados Timnit Gebru o estudo *Gender Shades*⁴¹, no qual demonstrou como essas tecnologias perpetuavam preconceitos raciais e de gênero, levando a consequências discriminatórias em áreas como segurança, recrutamento e justiça.

A denúncia de Bruna e de Joy, a reportagem da Bloomberg, o artigo sobre o COMPAS, assim como tantos outros casos reportados nos últimos anos⁴², fazem parte de um universo do que se convencionou chamar de racismo algorítmico, uma

³⁸ Dieterich, William; Mendonza, Christina; Brennan, Tim. **COMPAS Risk Scales: Demonstrating Accuracy, Equity and Predictive Parity**. Northpointe, 2016. Disponível em: <https://embed.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616/>. Acesso em: 21 mar. 2025.

³⁹ Boulamwini, Joy. In: TED Talk 2016. **How I'm fighting bias in algorithms**. (8 min) Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms. Acesso em: 22 mar. 2025.

⁴⁰ **Coded Bias**. Direção: Shalini Kantayya. Estados Unidos, China, Reino Unido: Netflix, 2020. Disponível em: <https://www.netflix.com/br-en/title/81328723>. Acesso em: 22 mar. 2025.

⁴¹ Buolamwini, Joy; Gebru, Timnit. **Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification**. Disponível em: <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>. Acesso em: 22 mar. 2025.

⁴² SILVA, Tarcízio (ed.) et al. **Mapeamento de Danos e Discriminação Algorítmica**. Desvelar, 2025. Disponível em: <https://desvelar.org/casos-de-discriminacao-algoritmica/>. Acesso em: 17 mai. 2025.

espécie de discriminação algorítmica que tem sua razão de ser pautada na raça. Segundo Tarcízio Silva, o racismo algorítmico⁴³ pode ser compreendido como

o modo pelo qual a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados. Tal ordenação pode ser vista como uma camada adicional do racismo estrutural, que, além do mais, molda o futuro e os horizontes de relações de poder, adicionando mais opacidade sobre a exploração e a opressão global que já ocorriam desde o projeto colonial do século XVI (Silva, 2022, p. 69).

Nesse sentido, faz-se necessário apreender o conceito de raça e de racismo, tão essenciais para o presente trabalho. A construção da raça como categoria está intimamente ligada ao desenvolvimento do capitalismo colonial e à estruturação das relações de dominação. Quijano ensina que “a ideia de raça, em seu sentido moderno, não tem história conhecida antes da América” (Quijano, 2025, p. 117) e que, à época das primeiras embarcações rumo ao aludido continente, as relações foram firmadas tendo por referência as diferenças fenotípicas e biológicas entre conquistadores e conquistados. Daí surgiram as identidades hoje consolidadas – negros, indígenas, mestiços – e outras já existentes foram redefinidas. A despeito da ideia de raça ter nascido como uma realidade biológica, existe consenso amplo dentro da comunidade científica de que se trata de uma construção social.

Convém dizer que o conceito de raça e de racismo estão profundamente imbricados e que este último precede o primeiro, “na medida em que a ciência colonial foi instrumentalizada para definir hierarquias de valores aplicadas aos diferentes grupos demográficos do mundo” (Silva, 2025, p. 26), visando a promoção de teorias racialistas e o posicionamento do europeu “como legítimo universal e merecedor das invasões empreendidas e da formatação do mundo, territórios e modos de vida” (Silva, 2025, p. 26). Dentro desse contexto, racismo pode ser compreendido como uma forma “sistemática de discriminação que tem a raça como fundamento, e que se manifesta por meio de práticas conscientes ou inconscientes que culminam em desvantagens ou privilégios para indivíduos, a depender do grupo racial ao qual pertençam” (Almeida, 2019, p. 22). Na definição de Sueli Carneiro:

⁴³ Frequentemente, o termo racismo algorítmico é usado de forma intercambiável com viés racial, viés algorítmico, discriminação algorítmica e discriminação algorítmica racializada. Na literatura estrangeira, há certa predileção pelo uso do termo “bias” (viés), o qual é associado a categorias como raça e gênero em análises sobre discriminação.

O racismo, enquanto pseudociência, busca legitimar a produção de privilégios simbólicos e materiais para a supremacia branca que o engendrou. São esses privilégios que determinam a permanência e a reprodução do racismo enquanto instrumento de dominação, exploração e, mais contemporaneamente, de exclusão social em detrimento de toda evidência científica que invalida qualquer sustentação para o conceito de raça (Carneiro, 2023, p. 20)

Para Abdias do Nascimento, raça não é apenas uma categoria de diferenciação fenotípica, mas um mecanismo político que opera por meio do racismo mascarado, especialmente no Brasil, onde o mito da democracia racial oculta a exclusão sistemática dos negros, negando-lhes identidade, direitos e autonomia cultural (Nascimento, 2016). Já Sueli Carneiro, objetivando explicar as relações raciais advindas da colonização europeia, desenvolve o conceito de dispositivo de racialidade que, segundo a autora, é um sistema de poder que constrói o negro como "outro" inferior, um "não-ser", para sustentar a supremacia branca, que se posiciona como o "ser" normativo. Esse conceito é central para entender o racismo estrutural no Brasil, especialmente em sua forma velada, mascarada pelo mito da democracia racial (Carneiro, 2023).

Intensamente endossado pelo governo, pela elite política e econômica e por teóricos no século XX⁴⁴, o mito da democracia racial construiu a falsa percepção de que no Brasil o racismo teria sido mais brando do que em outros países que, igualmente, passaram por um processo de escravização da população negra, a exemplo dos Estados Unidos. Tarcízio Silva diz essa versão deturpada da realidade brasileira é um dos entraves para se reconhecer, pela ciência social hegemônica, “aproximações entre abordagens estadunidenses e brasileiras sobre opressão racial” (Silva, 2025, p. 25). Todavia, ele lembra que em âmbito institucional, notadamente

Quanto à segregação explícita, políticas públicas, decretos e leis que desenvolveram a abordagem migratória para embranquecimento do Brasil são exemplos, que se somam ao uso efetivamente discriminatório de muitos instrumentos posteriores, como a Lei de Terras, Lei da Vadiagem, Lei de Drogas e outras provam que houve o uso do direito positivo para a segregação (Silva, 2025, p. 25).

⁴⁴ Por meio do livro *Casa Grande e Senzala* e de outras obras, o sociólogo Gilberto Freyre foi um dos principais articuladores da ideia de que a miscigenação no Brasil teria criado uma sociedade mais igualitária e menos racista que em outros países, como os Estados Unidos.

Diante dessa abordagem a partir da qual se apreende que no Brasil pós-abolição houve segregação institucionalizada da população negra, busca-se uma aproximação com por teóricos estadunidenses para compreender esse fenômeno transcontinental que se manifesta através da tecnologia.

Ruha Benjamin defende que a raça é uma forma de tecnologia, um constructo social projetado para "separar, estratificar e santificar as muitas formas de injustiça vivenciadas por membros de grupos racializados, mas que as pessoas rotineiramente reimaginam e redistribuem para seus próprios fins" (Benjamin, 2019, p. 31, tradução nossa). Assim como ferramentas tecnológicas (algoritmos, sistemas de IA, entre outros), a raça é usada para categorizar, hierarquizar e controlar populações, perpetuando sistemas de poder.

Benjamin estrutura sua análise em torno de frameworks conceituais inovadores, tendo o *New Jim Code* como conceito central. O termo remete ao conjunto de leis estaduais e locais que institucionalizou a segregação racial no sul dos Estados Unidos entre os séculos XIX e XX, mas que dentro da realidade do século XXI refere-se ao "uso de novas tecnologias que refletem e reproduzem desigualdades existentes, mas que são promovidas e percebidas como mais objetivas ou progressistas do que os sistemas discriminatórios de uma era anterior" (Benjamin, 2019, p. 13, tradução nossa). Isto é, em nome do progresso tecnológico, são configuradas novas formas de segregação e perpetuação do racismo, que se reinventa por meio de métodos hodiernos e sofisticados (Silva, 2020). O framework elaborado por Benjamin categoriza formas de discriminação tecnológica em quatro dimensões principais: desigualdade projetada (engineered inequity), discriminação por padrão (default discrimination), exposição codificada (coded exposure) e benevolência tecnológica (technological benevolence).

A primeira dimensão refere-se às tecnologias projetadas de forma que intencionalmente (ou não) incorporem vieses que potencializam desigualdades raciais, frequentemente por meio de dados históricos enviesados ou decisões de design que priorizam eficiência sobre equidade. Logo, em vez de eliminar o racismo, essas tecnologias o "engenhnam" em estruturas invisíveis, tornando-o mais difícil de contestar. Um exemplo bastante ilustrativo e citado por Benjamin é de um concurso de beleza conduzido por um sistema de IA⁴⁵ – supostamente imparcial –, no qual 86%

⁴⁵ Vide: <https://beauty.ai/>. Acesso em: 26 mar. 2025.

dos finalistas selecionados foram pessoas brancas, refletindo vieses nos dados de treinamento que associavam beleza a traços eurocêntricos (pele branca, nariz e boca pequenos, cabelo liso, olhos claros, etc.).

Na segunda dimensão, as configurações padrões das tecnologias assumem um usuário idealizado – branco, masculino e de classe média ou alta, geralmente – ignorando ou marginalizando outros grupos. Isso cria uma discriminação implícita, onde a tecnologia "funciona" para alguns, mas falha para outros, sem que haja um viés explícito programado. É o caso dos sistemas de reconhecimento facial que apresentam taxas de erro significativamente maiores para rostos negros ou, ainda, dos dispensadores de sabão automáticos que falham recorrentemente em detectar mãos de pessoas negras, especialmente as que têm o tom mais escuro, mas funcionam perfeitamente com pessoas brancas.

A terceira dimensão abrange tecnologias que aumentam a visibilidade e a vigilância sobre grupos racializados, expondo-os a maior escrutínio e controle, enquanto protegem ou ocultam os grupos privilegiados. Benjamin compara a exposição codificada a uma forma de segregação digital, onde dados e algoritmos criam "profecias autorrealizáveis" de criminalidade ou risco baseadas em padrões raciais. Essa é a lógica dos sistemas de policiamento preditivo, como o PredPol, que utilizam dados históricos de prisões para prever crimes futuros, resultando em maior policiamento em bairros cuja população negra é predominante – não porque haja mais crimes reais, mas porque os dados refletem décadas de policiamento desproporcional.

Por derradeiro, a última dimensão compreende iniciativas tecnológicas apresentadas como supostamente bem-intencionadas, tais como os modelos de inteligência artificial que pretendem solucionar os problemas da sociedade ou de grupos específicos. Na prática, esse tipo de tecnologia reforça desigualdades ao mascarar problemas estruturais ou impor soluções que beneficiam as elites. Isso quer dizer que a tecnologia não tem como solucionar problemas estruturais sem que antes se operem mudanças sociais profundas. Nesse contexto, a tal benevolência se mostra insidiosa porque usa retórica progressista para justificar opressão, desviando atenção de reformas sistêmicas. Benjamin menciona os aplicativos de saúde que prometem personalização, mas ignoram contextos raciais, como disparidades em acesso médico.

As quatro dimensões do *New Jim Code* evidenciam que a discriminação tecnológica não é um acidente, mas um produto de sociedades racistas que moldam

o design, os dados e o uso de tecnologias, perpetuando hierarquias raciais de forma cada vez mais invisível e eficiente. As dimensões se interconectam de modo cíclico: uma desigualdade engenhada pode levar a discriminação padrão, que, por sua vez, amplifica a exposição codificada, tudo sob o véu de benevolência. Inspirada em tradições abolicionistas históricas contra a escravidão e o encarceramento em massa, Benjamin enfatiza a urgência de uma transformação radical, propondo ferramentas abolicionistas que vão além da mera correção de vieses para desafiar e dismantelar os sistemas subjacentes.⁴⁶

Consoante pontuado, não se descuida que a teoria formulada por Benjamin se insere dentro de um contexto específico dos EUA, contudo, é particularmente relevante para compreender o racismo algorítmico e suas manifestações em tecnologias digitais, especialmente em modelos de inteligência artificial. E, igualmente, importante por ressoar com a história brasileira de colonialismo, escravidão e exclusão racial.

3.3. Perfis e Perfilamento

Em 1876, o antropólogo e psiquiatra Cesare Lombroso publicou a obra *L'Uomo Delinquente* (ou “O Homem Delinquente”, em português) e nela postulou que certos indivíduos possuem uma predisposição inata para a prática de crimes, decorrente de características biológicas herdadas que os situam em um estágio evolutivo inferior, próximo ao de ancestrais primitivos (Lombroso, 2010). Esse conceito, conhecido como atavismo, sugere que o criminoso nato seria uma regressão da raça humana a traços selvagens, manifestados tanto em aspectos físicos quanto psicológicos. Nessa perspectiva, os criminosos natos poderiam ser identificados por meio de características anatômicas específicas, denominadas estigmas. Entre esses traços, destacam-se assimetrias cranianas, mandíbulas proeminentes, orelhas malformadas, sobrancelhas espessas e olhos pequenos, que, segundo Lombroso, indicariam uma propensão ao comportamento delituoso. Somam-se a tais

⁴⁶ “Ruha Benjamin fala de um ‘toolkit abolicionista’, partindo do uso de dados como instrumentos de liberação desde as visualizações de Du Bois e estatísticas de Ida Wells-Barnett no século XIX até iniciativas recentes através de diversos projetos que geram mais transparência, ativismo social a favor da regulação e ferramentas narrativas, que permitam contar as histórias de mais pessoas envolvidas e afetadas pelas tecnologias” (Silva, Tarcízio. **Raça Depois da Tecnologia: ferramentas abolicionistas contra os novos racismos**. Disponível em: <https://tarciziosilva.com.br/blog/raca-depois-da-tecnologia-ferramentas-abolicionistas-contra-os-novos-racismos/>. Acesso em: 27 mar. 2025).

características físicas, traços psicológicos como impulsividade, insensibilidade à dor, ausência de empatia e deficiência moral, que, em tese, tornariam os indivíduos incapazes de se conformar às normas sociais.

A teoria desenvolvida por Lombroso representa uma das primeiras tentativas sistemáticas de explicar o comportamento criminoso por meio de uma perspectiva científica, ancorada sobretudo em pressupostos biológicos. Não obstante tenha sido refutada por carecer de embasamento empírico e por reforçar sistemas de opressão como o racismo, “trouxe pela primeira vez um enfoque sistemático à ideia de traçar características dos autores de crimes, embora hoje suas conclusões sejam amplamente criticadas por falta de rigor científico” (Siena, 2025). Antes de Lombroso, nas sociedades pré-modernas, a identificação de criminosos também carecia de rigor científico, sendo frequentemente fundamentada em explicações místicas ou religiosas. Durante a Antiguidade e a Idade Média, crimes eram atribuídos a forças sobrenaturais ou desvios morais, sem uma análise sistemática do comportamento do indivíduo. Desprovidas de embasamento empírico, tais práticas refletiam as limitações epistemológicas da época, nas quais o crime era interpretado como uma transgressão espiritual, e não como um fenômeno passível de investigação comportamental.

A literatura científica mostra que as primeiras abordagens sobre a prática de perfilamento advieram do campo criminal diante da necessidade de se desenvolver técnicas investigativas eficazes, focadas nas características psicológicas, comportamentais e sociodemográficas dos autores de crimes. No início dos anos 1970, Howard Teten, instrutor da Academia do *Federal Bureau of Investigation* (FBI), introduziu ao departamento conceitos de criminologia aplicada, desenvolvendo perfis para casos não resolvidos (casos arquivados ou, em inglês, *cold cases*). A abordagem de Teten, centrada na análise comportamental, marcou o início da institucionalização do perfilamento no âmbito das instituições públicas, que se consolidou com a criação do *Psychological Profiling Program* em 1978 (Douglas; Olshaker, 1997). A metodologia do FBI foi fortalecida na década seguinte, quando a *Behavioral Science Unit* (BSU) obteve financiamento do *National Institute of Justice*, o que possibilitou a criação de um banco de dados com entrevistas gravadas de homicidas condenados, ampliando o conhecimento sobre o comportamento criminal. O referido banco de dados tornou-se um pilar para a validação empírica do perfilamento, permitindo a sistematização de padrões e a aplicação da técnica em casos reais.

Em uma abordagem embrionária, vinculada ao contexto de investigação criminal da década de 1980, Gary T. Marx e Nancy Reichman descrevem o perfilamento como uma das principais técnicas⁴⁷ de busca sistemática de dados empregadas por organizações e/ou instituições para detectar infrações, especialmente aquelas de “baixa visibilidade”, como fraudes em programas de assistência social, sonegação fiscal e crimes do colarinho branco. Conforme explicam os autores:

O perfilamento permite que os investigadores correlacionem uma série de itens de dados distintos para avaliar o quão próximo uma pessoa ou evento se aproxima de uma caracterização ou modelo de infração predeterminado. As características modais e os padrões de comportamento de violações ou infratores conhecidos são determinados em relação às características de outros presumivelmente não infratores. Indicadores de possíveis violações são desenvolvidos a partir dessa comparação (Marx; Reichman, 1984, p. 3, tradução nossa)

O perfilamento, nessa concepção, é tido como um método indutivo que se distingue de abordagens dedutivas tradicionais, ao buscar correlacionar múltiplos itens de dados para construir modelos estatísticos ou descritivos de possíveis infratores, aumentando a probabilidade de identificação de infrações em comparação com buscas aleatórias. Precisamente por causa de sua natureza preditiva, o perfilamento assume um viés proativo e preventivo, logo, no campo criminal, é frequentemente usado para localizar potenciais transgressores e evitar infrações futuras, refletindo a racionalização do controle social (Marx; Reichman, 1984) – que é impulsionada pelo processo de informatização e pela necessidade de gerenciar infrações em sociedades burocráticas complexas.

Sob ótica distinta, Roger Clarke, ex-Comissário de Privacidade da Austrália, compreende o perfilamento como uma técnica de vigilância de dados (*dataveillance*)⁴⁸

⁴⁷ A outra técnica mencionada é o *matching*, ou cruzamento de dados, que envolve a comparação de informações provenientes de duas ou mais fontes de dados distintas, com o objetivo de verificar consistências, identificar discrepâncias ou detectar múltiplas ocorrências que sugiram infrações. Diferentemente do perfilamento, que opera de forma indutiva com indicadores probabilísticos, o *matching* é uma abordagem direta que busca evidências concretas de violações, como a presença de um mesmo indivíduo em listas mutuamente exclusivas (e.g., beneficiários de assistência social com contas bancárias acima do limite permitido) ou a inconsistência entre registros (e.g., pagamentos a pessoas falecidas).

⁴⁸ Clarke define *dataveillance* como a utilização sistemática de bancos de dados para monitorar ou investigar as ações e comunicações de indivíduos ou grupos, destacando-se como uma forma de vigilância mais eficiente e econômica em comparação com métodos convencionais. A prática se divide em *dataveillance* pessoal, focada no acompanhamento de indivíduos identificados, e *dataveillance* em massa, que visa detectar suspeitos ou potenciais clientes em grandes populações por meio de técnicas como perfilamento e cruzamento de dados.

que infere características de uma classe específica de pessoas a partir de experiências passadas e busca, em grandes bases de dados, indivíduos que correspondam a essas características (Clarke, 1993). Por essa definição, que não se restringe ao âmbito criminal, o perfilamento abrange o processo de criação e aplicação de perfis, com foco no indivíduo. Ao contrário de Marx e Reichman, que têm uma abordagem mais prática e centrada na aplicação do perfilamento para a descoberta de infrações e/ou infratores em contextos de *enforcement*, ou seja, uma abordagem mais setorializada, Clarke enfatiza o processo técnico, com etapas bem definidas e dependente de recursos computacionais e estatísticos.

A visão de Clarke, no entanto, é contraposta pelo jurista Lee Bygrave que, no início do século XXI, criticava os estudos sobre perfilamento que consideravam a criação de perfis como um processo direcionado exclusivamente a indivíduos, sem se reportar de forma significativa à dimensão coletiva (Bygrave, 2002). Pela perspectiva de Bygrave, o processo de perfilamento remete, além de indivíduos, a entidades coletivas (organizadas ou não organizadas), de modo que um perfil pode se relacionar, simultaneamente, a qualquer uma das três categorias. Nessa acepção, “perfilamento é a inferência de um conjunto de características (perfil) sobre uma pessoa individual ou entidade coletiva e o tratamento subsequente dessa pessoa/entidade ou de outras pessoas/entidades à luz dessas características” (Bygrave, 2002, p. 301, tradução nossa).

Nas décadas seguintes, a técnica de perfilamento evoluiu significativamente, tornando-se uma prática automatizada e algorítmica.⁴⁹ A mudança paradigmática foi resultado dos avanços tecnológicos alcançados no campo da inteligência artificial e é impulsionada, principalmente, pelo big data, pelo desenvolvimento de técnicas de mineração de dados e algoritmos de aprendizado de máquina. Mireille Hildebrandt foi uma das principais teóricas a repensar o perfilamento frente à evolução tecnológica, descrevendo-o como processo de descoberta de correlações em bancos de dados, que podem ser usadas para identificar e representar sujeitos (humanos ou não,

⁴⁹ Mireille Hildebrandt lembra que “muito antes de os computadores se tornarem parte da vida cotidiana, investigadores criminais elaboravam perfis de suspeitos desconhecidos, psicólogos compilavam perfis de pessoas com transtornos de personalidade específicos, gerentes de marketing elaboravam perfis de diferentes tipos de clientes em potencial e gerentes perfilavam o potencial de seus funcionários para funções específicas. O perfilamento adequado parece ter sido uma competência crucial da ocupação profissional e das empresas desde o seu surgimento, talvez mais visível hoje em dia no marketing e na investigação criminal” (Hildebrandt, Mireille. **Defining Profiling: A New Type of Knowledge?** In: Gutwirth, Serge; Hildebrandt, Mireille (Coords.). *Profiling the European Citizen: Cross-Disciplinary Perspectives*. New York: Springer, 2008. p. 17-45, pp. 23-24, tradução nossa).

individuais ou coletivos) ou para aplicar perfis (conjuntos de dados correlacionados) a fim de categorizar ou prever comportamentos (Hildebrandt, 2008). O perfilamento é, por esse prisma, um processo indutivo que gera hipóteses a partir de padrões de dados, testadas posteriormente na aplicação dos perfis. Assim, mostra-se acertada a equivalência feita entre perfis e hipóteses, uma vez que a construção destes não se baseia necessariamente “em marcos teóricos estabelecidos e causas definidas, mas na probabilidade de que as inferências derivadas do tratamento de dados também não de se confirmar no futuro” (Machado, 2022, p. 119).

Atualmente, o perfilamento é melhor apreendido pela definição conferida por Francesca Bosco que, a partir da contribuição dos autores retromencionados, o entende como uma técnica de processamento (total ou parcialmente) automatizado de dados (pessoais ou não), destinada a produzir conhecimento por meio da inferência de correlações, gerando perfis individuais ou coletivos que servem como base para a tomada de decisões ou, consoante as exatas palavras da autora:

Perfilamento é uma técnica de processamento (parcialmente) automatizado de dados pessoais e/ou não pessoais, com o objetivo de produzir conhecimento por meio da inferência de correlações a partir de dados na forma de perfis que podem ser posteriormente aplicados como base para a tomada de decisões.

Um perfil é um conjunto de dados correlacionados que representa um sujeito (individual ou coletivo).

A construção de perfis é o processo de descobrir padrões desconhecidos entre dados em grandes conjuntos de dados que podem ser usados para criar perfis.

A aplicação de perfis é o processo de identificar e representar um indivíduo ou grupo específico como se enquadrando em um perfil e de tomar alguma forma de decisão com base nessa identificação e representação (Bosco et. al, 2017, p. 9, tradução nossa).

É perceptível que se trata de categoria em constante evolução e daí justamente a dificuldade dos diplomas legais em captar a essência de uma técnica mutável por meio de uma conceituação precisa e suficiente. Assim como a maioria dos países que possuem legislação própria sobre proteção de dados pessoais, o Brasil é desprovido de uma definição legal – seja sobre perfil, seja sobre perfilamento – e, a nível global, o GDPR é um dos poucos documentos de cunho oficial a apresentar uma definição legal expressa, delimitando em seu art. 4º, 4, o perfilamento (ou *profiling*, em inglês) como

Qualquer forma de tratamento automatizado de dados pessoais que consista em utilizar esses dados pessoais para avaliar certos aspetos pessoais de uma pessoa singular, nomeadamente para analisar ou prever aspetos relacionados com o seu desempenho profissional, a sua situação económica, saúde, preferências pessoais, interesses, fiabilidade, comportamento, localização ou deslocações.

Da leitura do dispositivo, é possível deduzir que a incidência da norma alcança tão somente indivíduos, na medida em que o perfilamento alude ao tratamento automatizado de “dados pessoais para avaliar certos aspetos pessoais de uma pessoa singular”, passando ao largo da dimensão coletiva. Em sentido similar já apontava a Recomendação CM/Rec (2010)13, a respeito da proteção de indivíduos em relação ao tratamento automatizado de dados pessoais no contexto de perfilamento. Na aludida recomendação, perfil é compreendido como “um conjunto de dados que caracteriza uma categoria de indivíduos e que se destina a ser aplicado a um indivíduo” (Conselho da Europa, 2010, p. 9, tradução nossa), enquanto perfilamento concerne a “uma técnica de processamento automático de dados que consiste em aplicar um ‘perfil’ a um indivíduo, nomeadamente para tomar decisões sobre ele ou para analisar ou prever as suas preferências, comportamentos e atitudes pessoais” (Conselho da Europa, 2011, p. 9, tradução nossa). O novo regulamento de inteligência artificial da União Europeia (AI Act) não inova no sentido de conferir uma nova definição, seja para restringir ou ampliar a existente, limitando-se a remeter ao dispositivo do GDPR.

Embora as definições legais existentes não abarquem a dimensão coletiva do perfilamento, cabe fazer a devida distinção com a individual. A primeira categoria compreende a análise de dados referentes a um indivíduo, que tem por objetivo a criação de um perfil específico o qual descreve características, comportamentos, preferências ou riscos associados ao sujeito. Esse tipo de perfilamento é frequentemente utilizado para identificar ou caracterizar um indivíduo de maneira única dentro de um contexto ou para prever suas ações futuras com base em dados coletados exclusivamente sobre ele. Hildebrandt destaca que o perfilamento individual é particularmente relevante em tecnologias como a biometria comportamental, que analisa, por exemplo, padrões de digitação para autenticar a identidade de uma pessoa ou monitorar seu comportamento online.

A segunda categoria, por outro lado, envolve a análise de dados agregados de múltiplos sujeitos para identificar padrões ou correlações que definam características compartilhadas por um grupo. Dois são os contextos a que o

perfilamento de grupo é aplicado: comunidades preexistentes e categorias emergentes. No caso de comunidades preexistentes, o perfilamento busca descobrir atributos comuns entre membros de um grupo já definido, como os frequentadores de uma igreja local ou os estudantes de um dormitório. Através da mineração de dados, por exemplo, é possível descobrir que os membros de uma comunidade religiosa compartilham certos hábitos de consumo ou padrões de deslocamento, permitindo a criação de um perfil coletivo que pode ser usado para fins comerciais ou administrativos.

Já no caso de categorias emergentes, o perfilamento de grupo identifica grupos que não existiam previamente como uma comunidade consciente de si mesma, mas que surgem a partir de correlações encontradas nos dados (e.g., a identificação de um grupo de “mulheres de olhos azuis” que, por meio de mineração de dados, mostram uma probabilidade estatística de desenvolver uma doença específica, como o câncer de mama). Nesse caso, o grupo é constituído pelo processo de perfilamento, e seus membros podem não ter qualquer senso de pertencimento coletivo antes da categorização.

Outro aspecto essencial a respeito do perfilamento de grupo refere-se à distinção entre perfis distributivos e não distributivos. Os perfis distributivos são aqueles em que todos os membros do grupo compartilham integralmente os atributos do perfil (e.g., grupo de pessoas solteiras, o qual é definido pelo atributo universal de não serem casadas). Por natureza, é um tipo de perfil que é raro e geralmente tautológico, mas tem implicações sociais e legais diretas, pois pode ser aplicado a todos os membros sem exceção. Por outro lado, os perfis não distributivos, que são a norma na maioria dos casos, caracterizam grupos em que apenas uma parte dos membros exibe os atributos do perfil – por essa razão, são de generalização não universal. Um exemplo é um perfil baseado em renda e escolaridade, em que uma pessoa com ensino superior completo não ganha alto salário e, por outro lado, uma pessoa com ensino básico tem renda acima da média. Esses perfis são probabilísticos, logo, estão sujeitos a erros quando aplicados individualmente, o que pode levar a generalizações injustas ou discriminatórias.

Tais definições são importantes para compreender o funcionamento da técnica de perfilamento, tanto em sua modalidade individual quanto coletiva, e servem para delinear o panorama quanto ao seu uso, que é extenso e múltiplo. O perfilamento é manejado em tantas áreas quanto é possível se imaginar, não se limitando à

criminologia. Dentro do contexto das tecnologias digitais, o perfilamento racial se manifesta como uma das formas de exclusão e discriminação da população negra e pode assim ser definido:

O perfilamento racial, junto com a reprodução de vieses raciais, é um dos mecanismos de utilização de dados pessoais para a produção de discriminação. O perfilamento, em si, remete a coleta de dados pessoais para produzir um perfil e criar um alvo. Entende-se perfilamento racial (*racial profiling*) como o ato de classificar determinado corpo/território a partir de estereótipos racistas sobre o grupo/espço, tomar decisões sobre ele, sobre consumo, acesso, abordagem policial, informação, oportunidades, etc; e criar um alvo (Monagreda, 2024, p. 118).

Sob essa ótica, pode-se compreender que o perfilamento racial realizado por meio de sistemas de inteligência artificial resulta no que Silva (2022) denomina racismo algorítmico. Esse fenômeno representa uma atualização do racismo estrutural⁵⁰ e se difunde por meio das Tecnologias da Informação e Comunicação (TICs) de diversas formas, que vão desde microagressões – que se manifestam em microinsultos, microinvalidações, desinformação e deseducação (Silva, 2020) – até na vigilância excessiva e na criminalização (Monagreda, 2024).

No Brasil, um dos episódios mais emblemáticos de discriminação baseada em perfilamento racial ocorreu em 2018. A empresa Decolar.com, uma das maiores agências de viagens online da América Latina, foi investigada e multada pelo Departamento de Proteção e Defesa do Consumidor (DPDC) por adotar políticas de diferenciação de preços (*geopricing*) e restrição de serviços e conteúdos (*geoblocking*) com base na localização geográfica.⁵¹ Na prática, a empresa criava perfis territoriais e étnicos com base na localização geográfica, obtida por meio do endereço IP, e dependendo da área geográfica de onde a pesquisa fosse realizada, a empresa tinha a possibilidade de apresentar preços diferentes para as ofertas ou até mesmo restringir ou impedir o acesso a determinadas vagas. O objetivo era incentivar o consumo por indivíduos de certos estados, países e etnias, ao mesmo tempo em que desestimulava a interação de consumidores de localidades indesejadas.

Nos Estados Unidos, uma ferramenta chamada *ShotSpotter* (recentemente renomeada para *SoundThinking*) realiza a detecção de disparos de tiros e identifica a

⁵⁰CEE/FIOCRUZ. Tarcízio Silva: “O racismo algorítmico é uma espécie de atualização do racismo estrutural”. Rio de Janeiro, 2023. Disponível em: <https://cee.fiocruz.br/?q=Tarcizio-Silva-O-racismo-algoritmico-e-uma-especie-de-atualizacao-do-racismo-estrutural>. Acesso em: 14 mai. 2025.

⁵¹ Brasil. Ministério da Justiça e Segurança Pública. Decolar.com é multada por prática de **geo pricing** e **geo blocking**. Brasília, 18 jun. 2018. Disponível em: <https://www.gov.br/mj/pt-br/assuntos/noticias/collective-nitf-content-51>. Acesso em: 14 abr. 2025.

localização geográfica em segundos, permitindo uma resposta rápida das forças de segurança. O sistema funciona a partir de microfones e sensores distribuídos de forma oculta em vias públicas e, quando um barulho alto é identificado, um computador processa o som e o categoriza como um tiro de arma de fogo ou algo diferente – por exemplo, fogos de artifício. Em seguida, um agente humano revisa o resultado gerado pelo computador e, se for o caso, aciona a polícia para o local onde o barulho foi escutado.

Não obstante a ferramenta seja amplamente utilizada na área de segurança pública, estudos mostram a incidência de alertas falsos em bairros predominantemente negros e latinos⁵², que levam a polícia a inúmeras incursões em comunidades sem motivo e em alerta máximo, na expectativa de enfrentar uma situação potencialmente perigosa. Em documentos obtidos pelo *Wired*, constatou-se que os sensores que geram os alertas são posicionados estrategicamente em comunidades predominantemente negras e latinas – enquanto o *ShotSpotter* não está presente em bairros de maioria branca.⁵³ A tecnologia já foi exportada para países como Colômbia e Uruguai e, no Brasil, o Município de Niterói anunciou seu lançamento no ano de 2025.⁵⁴

Ainda no ensejo dos exemplos, Martins complementa:

No contexto brasileiro, os pesquisadores Ramon Vilarino e Renato Vicente demonstram como, na fase de testes de um sistema de pontuação de crédito alimentado com uma base de dados brasileira, o uso dos primeiros 3 dígitos do CEP enquanto variável de entrada funciona como um proxy para gerar vieses raciais. Os autores sugerem, inclusive, que o uso da variável de “proporção de branquitude” equivalha ao uso do CEP. Os pesquisadores usaram os mesmos dados (idade, histórico financeiro e de crédito, hábitos de pagamento), alterando apenas os 3 dígitos do CEP, “movendo” pessoas de São Paulo, em que 36% das pessoas se autodeclararam não-brancas, para a Bahia, em que 78% das pessoas se autodeclararam não-brancas. Em 99,8% dos casos houve uma diminuição absoluta na pontuação de crédito da pessoa.

O estudo ajuda a deixar ainda mais evidente as discriminações que podem ser perpetuadas pelo uso do profiling. Um ponto importante é que não é necessário usar dados de raça para que discriminações raciais sejam

⁵² Gwidt, David. **ShotSpotter Leak Shows That Surveillance Tech is Used to Overpolice Black and Brown Communities.** ACLU, Wisconsin, 6 mar. 2024. Disponível em: <https://www.aclu-wi.org/en/news/shotspotter-leak-shows-surveillance-tech-used-overpolice-black-and-brown-communities>. Acesso em: 25 jun. 2025.

⁵³ Mehrotra, Dhruv; Scott, Joey. **Here Are the Secret Locations of ShotSpotter Gunfire Sensors.** *Wired*, 22 fev. 2024. Disponível em: <https://www.wired.com/story/shotspotter-secret-sensor-locations-leak/>. Acesso em: 25 jun. 2025.

⁵⁴ Prefeitura de Niterói. **Niterói lança o único sistema de detecção de tiros e geolocalização em tempo real do Brasil.** Portal da Prefeitura, Niterói, 30 mai. 2025. Disponível em: <https://niteroi.rj.gov.br/niteroi-investe-em-seguranca-publica-e-lanca-o-unico-sistema-de-deteccao-de-tiros-e-geolocalizacao-em-tempo-real-do-brasil/>. Acesso em: 25 jun. 2025.

geradas. Embora possíveis soluções para o problema ainda sejam incipientes, os motivos e causas de vieses em algoritmos de machine-learning são muito bem mapeados, o que torna evidente que isso não se trata de um problema que afeta a vários indivíduos, mas que se manifesta de forma eminentemente coletiva e difusa (Martins, 2021, pp. 126-127).

A prática de inferir informações a partir de perfis é corriqueira e reverbera em grupos raciais minorizados, que são alvo de discriminações inúmeras originadas de sistemas de IA. O uso de algoritmos dentro desse contexto culmina no que Cathy O'neil denomina *feedback loop*, que são ciclos autorreforçantes nos quais os resultados de algoritmos baseados em dados históricos influenciam as decisões futuras, perpetuando ou amplificando vieses e desigualdades (O'neil, 2020). Os feedback loops são particularmente perigosos porque criam um ciclo vicioso: os dados iniciais, muitas vezes enviesados, geram decisões que reforçam esses mesmos vieses, criando um sistema que parece justo (por ser matemático) mas que, na realidade, perpetua injustiças.

É o caso, por exemplo, de pontuação de crédito que leva em consideração dado de região de nascimento ou raça/etnia do indivíduo, que são características que não tem como ser alteradas. Por consequência, o score do indivíduo sempre será baixo e ele nunca poderá obter crédito e terá dificuldades para ascender socialmente. Essa lógica se aplica à formação e aplicação de perfis, visto que é dessa maneira que os vieses raciais são reforçados por sistemas de IA. A solução para esse problema em prol da proteção da população negra envolve a transparência e auditoria dos algoritmos, a incorporação de feedback humano, a adoção de princípios éticos e, principalmente, a regulação da inteligência artificial.

4. REGULAÇÃO DO PERFILAMENTO NO ORDENAMENTO JURÍDICO BRASILEIRO

4.1. A regulação sob a ótica da proteção de dados

A relação entre dados pessoais e inteligência artificial é marcada pela profunda interdependência entre um e outro, que formam a base para o funcionamento e evolução dos sistemas inteligentes baseados em aprendizado de máquina. Os dados pessoais, que abrangem desde nome e data de nascimento até informações como hábitos, preferências e comportamentos, são o combustível que alimenta os algoritmos de IA permitindo a personalização de serviços, a tomada de decisões automatizadas e o desenvolvimento de soluções inovadoras. Trata-se de uma relação simbiótica que potencializa o progresso tecnológico, de um lado, mas que suscita desafios éticos e legais, de outro.

No que concerne à regulação da técnica de perfilamento, notadamente em sua vertente automatizada – impulsionada por *machine learning* e *deep learning* –, existe no ordenamento jurídico brasileiro normas que tratam diretamente sobre a matéria – seja em relação à fase de formação de perfil, de aplicação de perfil ou, ainda, de ambas as fases – a exemplo do Código de Defesa do Consumidor (Lei n. 8.078/1990), Lei do Cadastro Positivo (Lei n. 12.414/2011), Marco Civil da Internet (Lei n. 12.965/2014) e a Lei Geral de Proteção de Dados (Lei n. 13.709/2018). A despeito do expressivo arcabouço normativo, que abarca tanto leis setoriais quanto leis gerais, o foco do presente capítulo será direcionado para a LGPD.

Por se tratar de uma lei geral, o legislador ordinário optou por conferir à LGPD natureza principiológica – são ao todo 10 princípios dispostos no art. 6º –, mas também conceitual (Mulholland, 2018). Isso significa que, além de estabelecer princípios para nortear a atividade de tratamento de dados, a própria lei se encarregou de apresentar o conceito de categorias que permeiam todo o diploma legal, a começar pela própria definição de dado pessoal e, em contraposição, de dado não pessoal, assim como de dado pessoal sensível:

Art. 5º Para os fins desta Lei, considera-se:

I - dado pessoal: informação relacionada a pessoa natural identificada ou identificável;

II - dado pessoal sensível: dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural;

III - dado anonimizado: dado relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento;

Conforme adiantado no primeiro capítulo, em consonância com o GDPR, a LGPD adota um conceito amplo de dado pessoal, o qual se refere tanto à pessoa identificada quanto à pessoa identificável. Quanto a esse aspecto da norma, Martins faz uma ressalva ao frisar que as regulações de proteção de dados contemporâneas “enfrentarão dificuldades para lidarem com os desafios gerados pelo tratamento automatizado de grandes quantidades de dados se continuarem restritas à identificabilidade enquanto parâmetro normativo, ainda que de forma expansionista” (Martins, 2021, p. 137). Alinha-se à visão consequencialista defendida por Bruno Bioni, segundo a qual, a regulação deve recair sobre os efeitos do tratamento de dados, ou seja, sobre os resultados propriamente ditos para que seja eficaz:

Se a premissa da causa regulatória da proteção de dados pessoais é tutelar o cidadão, que é cada vez mais exposto a tais tipos de práticas que afetam a sua vida, então, uma compartimentalização “dura” entre dados pessoais e dados anonimizados deixaria de fazer sentido. Em especial, quando está em questão a formação de perfis comportamentais que tem por objetivo precípua influenciar de alguma forma a vida de uma pessoa, que está atrás de um dispositivo e pouco importa ser ela identificável ou não (Bioni, 2020, p. 71).

A despeito das vozes dissidentes e contrárias ao modelo atual, há se reconhecer que privilegiou-se uma abordagem dualista, a partir da qual o que não é dado pessoal não está salvaguardado pela LGPD e sob esse prisma devem ser analisados seus institutos. É nesse plano que se coloca a prática de perfilamento, a qual se estrutura em três fases distintas e complementares: coleta de dados, construção do perfil (ou modelo) e tomada de decisão – a qual advém da aplicação do perfil. Tratando-se de perfilamento automatizado, os dados coletados são obtidos de duas formas, a primeira através do fornecimento voluntário pelos usuários (e.g., preenchimento de formulários e questionários), a segunda por meio da extração de dados comportamentais em grandes bases de dados, que, por vezes, são complementados por informações de terceiros, como os corretores de dados (*data brokers*)⁵⁵, que agregam dados de fontes públicas e privadas.

⁵⁵ São “empresas ou unidades de negócios que obtêm sua receita primária fornecendo dados ou inferências sobre pessoas, coletados principalmente de fontes que não os próprios indivíduos aos quais os dados se referem” (Rieke, Aaron et al., 2016, p. 4, tradução nossa).

Nessa etapa inicial, de pré-formação do modelo, incide a LGPD em relação aos dados de pessoa natural identificada e identificável e, por consequência, todos seus dispositivos e princípios devem ser observados. Isso que dizer que, antes mesmo da formação do perfil, os agentes de tratamento (seja pessoa pública ou privada) devem lastrear sua atividade com os dados pessoais – coleta, armazenamento, uso, compartilhamento, etc. – em uma das bases legais previstas nos arts. 7º, 11 e 23. Para que o tratamento de dados seja considerado lícito deve existir legitimidade, a qual é conferida ao agente de tratamento diretamente pelo titular via consentimento ou por força da lei quanto às demais hipóteses (cumprimento de obrigação legal ou regulatória, execução de contrato, execução de políticas públicas, exercício regular de direitos, entre outros). Se a atividade não se fundamenta em base normativa, está-se diante de tratamento ilícito, não sendo possível qualquer processamento dos dados coletados para fins de perfilamento.

Havendo legitimidade, a LGPD estabelece que deve ser informada a “finalidade específica do tratamento” (art. 9º, I) e a “forma e duração do tratamento” (art. 9º, II). Contudo, deixa de fixar critérios objetivos acerca do conteúdo da informação na hipótese de tratamento para perfilamento e, nesse ponto, se afasta de seu referencial internacional, visto que o regulamento europeu prevê o fornecimento de “informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento para o titular dos dados” (arts. 13, 2, f, e 14, 2, g do GDPR). Nada obstante, uma vez ciente e devidamente informado sobre o tratamento de dados para fins de formação de perfil, o titular de dados pode revogar seu consentimento (art. 18, IX) ou exercer seu direito de oposição (art. 18, § 2º). Quanto ao exercício de tais direito de objeção, Bioni ensina que:

Diferentemente da revogação do consentimento, que se apresenta como um direito potestativo e sem limitações, *a priori* estabelecidas, a LGPD condicionou o exercício do direito de oposição desde que haja uma violação a uma das suas normas. Uma primeira interpretação mais apressada levaria à conclusão de que o exercício desses dois direitos de objeção teria alcances distintos, já que o último não dependeria única e exclusivamente da vontade do titular em exercê-lo.

Se assim fosse, como resultado haveria uma indesejada assimetria normativa entre tais bases legais, a qual, como visto, procurou ser equalizada pelo legislador ao alocar sob o mesmo nível hierárquico todas as bases legais. Ao fim e ao cabo, deve-se buscar, sempre que possível, uma interpretação que busque colocar em pé de igualdade as hipóteses de legitimação para o tratamento de dados pessoais. Especialmente, o consentimento frente ao legítimo interesse (Bioni, 2020, p. 250).

Enquanto a revogação de consentimento se mostra uma forma de objeção mais direta e sem tantos obstáculos por se ancorar em um direito potestativo, a oposição

está acompanhada de uma condicionante, que é o descumprimento ao disposto na LGPD, e por esse motivo, “não é um direito absoluto, que deverá ser atendido em qualquer circunstância”. “Isso porque, embora o titular possa decidir se opor ao tratamento, demanda-se uma análise contextual que leve em consideração todos os interesses em jogo” (Bioni; Rielli; Kitayama, 2021, p. 233). É o caso, por exemplo, de uma instituição financeira que mantém e trata dados de clientes, como histórico de transações, para identificar padrões suspeitos de fraude. Mesmo que o titular se oponha, a instituição pode continuar o tratamento dos dados, pois isso é fundamentado no interesse legítimo de proteger a segurança das operações financeiras e evitar atividades fraudulentas.

Quanto à fase de construção do perfil, Machado (2022) propõe uma diferenciação entre os dois tipos de perfis, tão crucial para verificar se o estatuto protetivo é aplicável ao modelo gerado de forma automatizada. A premissa inaugural é de que para a construção dos perfis personalizados (individuais) demanda-se, invariavelmente, o tratamento de dados que estão, direta ou indiretamente, relacionados a uma pessoa natural (e.g. Sistemas de *credit scoring* e o tratamento de dados do usuário para obtenção de empréstimo), de modo que dados pessoais tendem a ser processados durante a criação de perfis e, por conseguinte, atraem a aplicação da LGPD.

Em relação aos perfis de grupo, a situação é relativamente distinta. A formação desses perfis pode envolver processamento de dados que se qualificam como não-pessoais, como informações agregadas ou anonimizadas por técnicas como mineração de dados (classificação e clusterização), agregação de dados ou sintetização de dados. Tais técnicas preservam propriedades estatísticas sem identificar indivíduos, tornando os dados anonimizados (não relativos à pessoa natural identificada ou identificável). Sendo esse o cenário, a norma não tem como ser invocada em favor do titular de dados por ausência de previsão legal.

Se houver tratamento de dados pessoais antes da anonimização, que é uma operação de tratamento em si, a LGPD aplica-se temporariamente⁵⁶, exigindo base legal para legitimar a atividade, compatibilidade de finalidade, informação ao titular e o exercício de direitos como revogação ou oposição. No entanto, uma vez

⁵⁶ Trata-se de situação transitória em que os dados, antes de serem anonimizados, ainda são considerados pessoais e, portanto, sujeitos à LGPD. Ou seja, é o interregno entre o início e o fim do processo de anonimização.

anonimizados, os dados não estão mais sob o regime da LGPD, e a criação de perfis de grupo com esses dados não enfrenta as mesmas restrições legais. Por exemplo, técnicas como agregação de dados ou sintetização (geração de dados sintéticos a partir de dados reais) permitem criar perfis sem identificar indivíduos, como em estudos de mobilidade urbana ou análise de tendências populacionais.

O art. 12, §2º, da LGPD é o único dispositivo que trata expressamente da criação de perfis, mas limita sua aplicação a perfis comportamentais de pessoas naturais identificadas ao prescrever que “poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada”. O *caput* do art. 12 exclui dados anonimizados do escopo da LGPD, mas o § 2º cria uma exceção: se os dados anonimizados forem usados para criar um perfil comportamental de pessoa natural identificada, eles são tratados como dados pessoais. Isso implica que, em regra, a LGPD se aplica a perfis individuais, mas não a perfis de grupo, que não se referem a indivíduos identificados ou identificáveis.

No que concerne a essa conclusão, cabe algumas considerações que dialogam diretamente com o tema deste trabalho. Consoante assentado nos capítulos precedentes, o racismo compreende sistema de opressão que opera por meio de instituições sociais, políticas e econômicas, e que está profundamente ligado à lógica colonial e capitalista. Nesse sentido, não é apenas um problema de atitudes e comportamentos individuais, mas uma estrutura que sustenta desigualdades e promove exclusões sociais. Sendo assim, deve ser entendido como fenômeno multidimensional (Silva, 2025), na medida em que, além da discriminação racial, “as vítimas podem sofrer múltiplas ou agravadas formas de discriminação calcadas em outros aspectos correlatos como sexo, língua, religião, opinião política ou de qualquer outro tipo, origem social, propriedade, nascimento e outros” (ONU, 2001, p. 9). Incorporado às tecnologias digitais, a exemplo do perfilamento automatizado, transforma-se em opressão algorítmica (Noble, 2018).

Percebe-se de imediato uma limitação quanto ao alcance da norma para tratar de questões como o perfilamento racial automatizado, que atinge a população negra não somente em seu nível individual, mas principalmente coletivo. Passa-se ao largo do problema maior, que tem dimensões coletivas, e adota-se uma abordagem individualista, centrada na pessoa natural. Quanto a esse ponto não há dúvidas, pois a parte final do § 2º do art. 12 é categórica ao sublinhar que apenas pessoas naturais,

se identificadas, podem se valer dessa exceção. Essa abordagem, no entanto, se mostra insuficiente e inadequada para enfrentar o perfilamento enquanto fenômeno supraindividual, especialmente por desconsiderar que a formação de perfis de grupos é a regra, não a exceção (Martins, 2021). Na prática, a escolha legislativa possibilita que os agentes de tratamento se evadam em relação aos preceitos da lei geral de proteção de dados sob o pretexto de que o tratamento ocorre apenas com dados anonimizados.

É a partir dessa perspectiva, por meio da qual se deduz o caráter pessoal ou não dos perfis individuais e de grupo gerados por perfilamento automatizado, é que deve ser avaliado o arcabouço principiológico da LGPD. A regulação do perfilamento automatizado repousa sobre um conjunto de princípios fundamentais inscritos no art. 6º, os quais servem como vetores normativos para disciplinar as etapas de formação e aplicação de perfis, integrando-se à lógica do perfilamento como um todo. Tais princípios reforçam a função instrumental do direito à proteção de dados pessoais na tutela de direitos e liberdades fundamentais, especialmente frente aos desafios do processamento algorítmico de dados comportamentais.

O princípio da finalidade (art. 6º, I) constitui a base da lógica de perfilamento, exigindo que o tratamento seja para “propósitos legítimos, específicos, explícitos e informados ao titular”. Na etapa de formação de perfis, a coleta e o processamento de dados devem guardar estrita relação com o que foi especificado e informado ao titular de dados e, por consequência, o requisito de legitimidade deve ser cumprido através de uma base legal explícita (e.g., consentimento ou legítimo interesse). A norma veda o uso dos dados coletados para fins incompatíveis com a finalidade originalmente informada, todavia, admite-se o uso secundário na hipótese em que o novo propósito for compatível contextualmente com o que foi estabelecido inicialmente, em prestígio ao princípio da adequação (art. 6º, II). Visa-se assim não apenas garantir a previsibilidade e confiança quanto ao tratamento, mas também mitigar os riscos de desvios na utilização de dados em modelos algorítmicos.

O princípio da necessidade (art. 6º, III) impõe a pertinência, proporcionalidade e não excessividade em relação às finalidades do tratamento, funcionando como restrição na formação de perfis ao incentivar a minimização de dados pessoais identificáveis e o uso de técnicas de anonimização, especialmente em perfis de grupo, onde a LGPD pode não incidir devido à natureza não pessoal dos dados. Na aplicação, ele assegura que decisões automatizadas, como em sistemas de

recomendação ou *credit scoring*, utilizem apenas os dados estritamente necessários, evitando excessos que tenham o potencial de reproduzir vieses.

No que diz respeito aos dados em si, a LGPD traz o princípio da qualidade dos dados (art. 6º, V), que prevê a necessidade de exatidão, clareza, relevância e atualização dos dados. Na formação de perfis, combate falhas em sistemas algorítmicos que gerem modelos imprecisos ou discriminatórios; na aplicação, assegura a confiabilidade de inferências comportamentais, impactando no exercício de direitos como retificação e bloqueio, reforçando sua convergência com a não discriminação. É um dos princípios que dialogam diretamente com o tema concernente ao racismo algorítmico, tendo em vista ser notório que uma das razões para perpetuação de vieses raciais decorre da baixa qualidade dos dados coletados e, posteriormente, processados por sistemas de IA. Dados de alta qualidade garantem que as análises e modelos derivados sejam precisos, consistentes e representativos da realidade. Por outro lado, quando os dados são imprecisos, incompletos ou desatualizados, ou seja, de baixa qualidade, as análises e decisões baseadas neles ficam comprometidas, especialmente as que envolvem grupos minoritários.

Em relação à aplicação ética da IA (Junquilho, 2023), três princípios distintos se interligam: transparência, não discriminação e responsabilização e prestação de contas. De acordo com o princípio da transparência (art. 6º, VI), é assegurado aos titulares o fornecimento “de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial”. Em outras palavras, a transparência

refere-se à capacidade de um modelo específico ser compreendido. No sentido mais estrito, um modelo é transparente se uma pessoa puder contemplar todo o modelo de uma só vez. A transparência pode ser considerada no nível do modelo como um todo, no nível de componentes individuais (por exemplo, parâmetros) e no nível de um algoritmo de treinamento específico. Uma segunda noção, menos estrita, de transparência pode ser que cada parte do modelo (por exemplo, cada entrada, parâmetro e computação) admite uma explicação intuitiva (EDPS, 2023, p. 4, tradução nossa).

A transparência está intimamente conectada a três conceitos centrais, um antagônico e dois complementares: opacidade, interpretabilidade e explicabilidade. A opacidade, conforme já visto, é a dificuldade ou impossibilidade de entender como os algoritmos de IA tomam decisões, quais dados utiliza, ou como chegam a determinados resultados, remetendo à ideia de caixa preta. Em contraposição, a

interpretabilidade diz respeito ao grau de compreensão humana de um determinado modelo ou decisão, enquanto a explicabilidade refere-se à habilidade de descrever com precisão o mecanismo ou o processo que levou ao resultado (*output*) de um algoritmo de IA.

Na etapa de formação de perfil, a transparência obriga os controladores a informar os titulares sobre a finalidade do tratamento, a forma como os dados serão processados e a lógica dos sistemas algorítmicos utilizados na criação de perfis. Por exemplo, em sistemas como assistentes virtuais ou plataformas de recomendação, o titular deve ser informado se seus dados comportamentais (e.g, histórico de navegação ou preferências) serão usados para inferir modelos preditivos. Isso facilita o exercício de direitos como o acesso (art. 9º) e a oposição (art. 18, §2º), permitindo que o titular questione a legitimidade do tratamento antes mesmo da criação do perfil. Na fase de aplicação, a transparência se materializa no dever de fornecer explicações compreensíveis sobre decisões automatizadas, conforme o artigo 20, §1º, da LGPD, que prescreve o fornecimento de “informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada”. Tal aspecto da norma é particularmente relevante no contexto do perfilamento automatizado, onde o titular precisa entender as razões de uma decisão que afeta seus interesses.

Com base no direito fundamental à igualdade, o princípio da não discriminação (art. 6º, IX) proíbe tratamentos discriminatórios ilícitos ou abusivos, diretos ou indiretos, integrando-se ao perfilamento ao vedar vieses na formação, como na definição de metodologias ou bases de dados, e na aplicação, combatendo discriminações algorítmicas, através de mecanismos como restrição de dados sensíveis. O racismo, assim como qualquer forma de discriminação ou segregação baseado em etnia ou raça, é proibida pela legislação brasileira, que aborda o assunto em diversos diplomas legais. Em primeiro lugar, a Constituição Federal assevera que "os objetivos fundamentais da República Federativa do Brasil incluem promover o bem de todos, sem preconceitos de origem, raça, sexo, cor, idade e outras formas de discriminação" (art. 3º, IV) e que o Brasil é regido pelo princípio do repúdio ao racismo (art. 4º, VIII). Além disso, a Constituição estipula que "todos são iguais perante a lei, sem distinção de qualquer tipo, garantindo aos brasileiros e estrangeiros que residem no país o direito inviolável à vida, à liberdade, à igualdade, à segurança e à propriedade" (art. 5º, caput) e criminaliza a prática de racismo ao estabelecer "a lei punirá toda discriminação que atente contra os direitos e liberdades fundamentais"

(art. 5º, XLI) e que "a prática do racismo é considerada um crime inafiançável e imprescritível, sujeito à pena de reclusão, conforme a legislação" (art. 5º, XLII).

A matéria atinente ao racismo e à discriminação racial é ainda regulada pela Convenção Interamericana contra o Racismo, a Discriminação Racial e Formas Correlatas de Intolerância, internalizada no ordenamento jurídico pátrio no ano de 2022 com status de emenda constitucional, e no âmbito infraconstitucional pela Lei n. 7.716/1989, que define crimes resultantes de preconceito de raça ou cor, e a Lei n. 12.288/2010, que institui o Estatuto da Igualdade Racial.

Haja vista o contexto social que levou à formação desigual da população brasileira, foi graças ao esforço organizado de movimentos de grupos racialmente minorizados que o Estado começou a oferecer proteção explícita e institucionalizada à população negra. Portanto, ao tratar sobre racismo e suas manifestações, fala-se sobre práticas não admitidas pela legislação brasileira e que, por assim serem, são antijurídicas. É importante destacar que, até recentemente, não havia na legislação brasileira uma definição ou conceito claro sobre o que seria racismo, de modo que se recorria a termos que representam as consequências do racismo – como “preconceito racial” e “discriminação racial” – na tentativa de explicá-lo. As explicações, no entanto, eram insuficientes e não davam conta de apreender todos os aspectos de sua dimensão coletiva (estrutural e sistêmica). Com a promulgação da Convenção Interamericana contra o Racismo, a Discriminação Racial e Formas Correlatas de Intolerância, passou-se a ter uma definição:

4. Racismo consiste em qualquer teoria, doutrina, ideologia ou conjunto de ideias que enunciam um vínculo causal entre as características fenotípicas ou genotípicas de indivíduos ou grupos e seus traços intelectuais, culturais e de personalidade, inclusive o falso conceito de superioridade racial. O racismo ocasiona desigualdades raciais e a noção de que as relações discriminatórias entre grupos são moral e cientificamente justificadas. Toda teoria, doutrina, ideologia e conjunto de ideias racistas descritas neste Artigo são cientificamente falsas, moralmente censuráveis, socialmente injustas e contrárias aos princípios fundamentais do Direito Internacional e, portanto, perturbam gravemente a paz e a segurança internacional, sendo, dessa maneira, condenadas pelos Estados Partes.

É de se ver que o racismo tem abrangência maior que a discriminação racial, todavia, esta última é essencial para a manutenção do primeiro. Logo, o combate à discriminação, conforme previsto no princípio esculpido no art. 6º, IX, constitui um dos meios de eliminação do racismo.

Durante a formação de perfil, o princípio da não discriminação incide na etapa de construção de modelos algorítmicos, exigindo que metodologias, bases de

treinamento e processos de validação sejam projetados para evitar a inscrição de vieses que perpetuem desigualdades estruturais. Na fase de aplicação de perfil, o princípio serve ao combate de discriminações decorrentes de decisões automatizadas, todavia, enfrenta obstáculos significativos, especialmente porque a identificação de discriminações indiretas é complexa, pois estas emergem de interações sutis entre dados, algoritmos e contextos sociais, frequentemente opacas devido à natureza “caixa-preta” de modelos de IA avançados.

Por fim, o princípio da responsabilização e prestação de contas (art. 6º, X) impõe a prospecção de riscos e demonstração de conformidade, manifestando-se em obrigações como relatórios de impacto e registros de operações, aplicáveis a ambas as etapas para mitigar impactos em direitos fundamentais, alinhando-se a uma abordagem baseada em riscos. Nos moldes como proposta, a responsabilização promove uma abordagem proativa, exigindo que controladores implementem políticas internas de governança, como programas de conformidade e treinamentos, e adotem tecnologias que facilitem a auditoria de sistemas algorítmicos, como ferramentas de explicabilidade de IA.

4.2. Perfilamento e direito de revisão

Feitas as considerações a respeito da parte geral da lei geral de proteção de dados e sua incidência sobre a prática de perfilamento, convém agora direcionar o olhar para o único dispositivo que disciplina em alguma medida. O artigo 20 emerge como um pilar regulatório essencial no âmbito do perfilamento automatizado e sua atual redação assim prescreve:

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

Não existe dentro do ordenamento jurídico brasileiro, e precisamente na lei geral de proteção de dados, definição sobre o que seria perfil e perfilamento, razão

pela qual, para fins do presente trabalho, busca-se a definição apresentada pelo GDPR (art. 4º, 4) e pela doutrina (Bosco et al., 2017; Monagreda, 2024) para amparar as análises feitas. O perfilamento aqui tratado abarca qualquer forma de processamento automatizado de dados pessoais para formação e aplicação de perfil com o objetivo de analisar ou prever aspectos relativos a comportamentos, preferências, interesses, trabalho, saúde, situação econômica, entre outros aspectos relacionados aos titulares de dados. Já perfil consiste em um conjunto de dados correlacionados que representa um sujeito, seja individual ou coletivo.

Os pressupostos de incidência do artigo 20 configuram-se como critérios cumulativos que delimitam seu âmbito de aplicação no contexto do perfilamento. Primeiramente, a decisão deve ser exclusivamente automatizada, sem intervenção humana relevante — entendida não como uma presença nominal ou procedimental vazia, mas como uma influência substantiva capaz de alterar o resultado final (Mulholland; Frajhof, 2020). Se o papel do ser humano se resume à simples ratificação da decisão algorítmica, “sem possibilidade de criticá-la ou descartá-la, então a decisão é automatizada, apesar de eventualmente ser assinada por um ser humano. Neste último caso, ocorre aquilo que se se chama de *rubber-stamping*, ou seja, um mero carimbo do ser humano” (Reis; Furtado, 2022, p. 13). Por outro lado, “se a máquina apenas assiste o humano, fornecendo-lhe elementos para avaliar as melhores alternativas, cabendo a escolha do resultado ao humano, isso não pode ser definido como decisão automatizada” (Reis; Furtado, 2022, p. 13). Em outras palavras, se há juízo discricionário decisivo por parte de agentes humanos, a decisão não tem como ser considerada automatizada, logo, não é acobertada pelo art. 20 da LGPD.

Em segundo lugar, o tratamento de dados deve envolver dados pessoais, definidos no artigo 5º, I, como informações relativas a uma pessoa natural identificada ou identificável, ou seja, dados que permitam, direta ou indiretamente, a associação a um indivíduo específico, como nome, CPF, endereço, histórico de navegação ou preferências comportamentais, desde que vinculados a uma identidade. Delimita-se daí a aplicação da norma a processos de perfilamento que gerem perfis individuais ou a aplicações de perfis que, embora baseados em padrões grupais, resultem em inferências individuais identificáveis, como quando um modelo de segmentação de mercado é aplicado para direcionar publicidade a um usuário específico com base em dados demográficos. Essa exigência exclui do escopo do artigo 20 os processos que se restringem ao uso de dados anonimizados — aqueles submetidos a técnicas que

impossibilitam a identificação do titular, conforme artigo 5º, XI, da LGPD — ou dados agregados não-pessoais, como estatísticas populacionais ou análises macroeconômicas que não permitem vinculação a indivíduos. Cabe pontuar que perfis de grupos baseados em dados não pessoais, embora fora do alcance direto do artigo 20, podem gerar impactos discriminatórios coletivos, como em sistemas de vigilância preditiva que estigmatizam comunidades marginalizadas – como a população negra – sem identificar indivíduos, evidenciando uma lacuna regulatória que demanda abordagens éticas complementares.

Em terceiro lugar, a decisão deve afetar interesses juridicamente relevantes do titular, englobando um espectro amplo de situações jurídicas subjetivas, como direitos subjetivos (e.g., direito à privacidade ou à igualdade), direitos potestativos (e.g., capacidade de exigir a revisão de uma decisão), poderes jurídicos (e.g., direitos de contratar ou acessar serviços) ou interesses legítimos (e.g., expectativas razoáveis de tratamento equitativo em relações comerciais ou sociais). O artigo 20 apresenta um rol exemplificativo, mencionando “decisões destinadas a definir perfil pessoal, profissional, de consumo, de crédito ou aspectos de sua personalidade”, mas sua redação aberta permite uma interpretação expansiva, abarcando impactos em outras esferas não previstas expressamente onde os efeitos das decisões automatizadas comprometam direitos fundamentais ou interesses protegidos pelo ordenamento jurídico. No contexto de perfilamento, a flexibilidade interpretativa é crucial, notadamente porque as diversas tecnologias baseadas em inteligência artificial geram impactos que transcendem os exemplos expressos, afetando esferas como a autonomia individual, a equidade social e a participação democrática. Contudo, a ausência de critérios claros na LGPD para delimitar o que constitui um interesse juridicamente relevante levanta desafios interpretativos, exigindo regulamentações complementares pela ANPD ou decisões judiciais que balizem a aplicação do dispositivo, especialmente em cenários de discriminação algorítmica ou manipulação comportamental.

Dentro da seara de proteção de dados pessoais, o artigo 20 da LGPD se posiciona como um dos principais, senão o principal, dispositivos de enfrentamento às decisões automatizadas e ganha forma mediante o exercício de dois direitos distintos e complementares: direito à revisão e direito à explicação. O primeiro, explícito no *caput*, opera como um mecanismo de contestabilidade, permitindo ao titular impugnar decisões automatizadas que incidam sobre sua esfera jurídica,

promovendo uma redistribuição mais equânime de poder decisório no âmbito digital. O último, apreendido da redação do § 1º, está intimamente ligado à transparência e vai além da mera divulgação de dados brutos, abrangendo a lógica e os critérios utilizados nos algoritmos, na medida do possível, sem comprometer segredos industriais ou comerciais.

Para que o titular alcance o sucesso esperado em sua revisão, é imprescindível uma explicação inteligível acerca do funcionamento do sistema de IA, sem a qual fica comprometido o direito de revisão. Nesse contexto, a transparência não implica a revelação da fórmula algorítmica em si, mas sim a comunicação de informações úteis e significativas (*meaningful information*) sobre a lógica envolvida (Selbst; Powles, 2017), como sugerido pelos arts. 13, 2, f, 14, 2, g e 15, 1, h do GDPR. A norma não garante a reversão automática de decisões automatizadas, mas assegura a suspensão temporária de seus efeitos durante o processo de revisão (Machado, 2022), possibilitando ajustes metodológicos substanciais, como a reformulação de algoritmos ou bases de dados, correções de vieses algorítmicos, como a eliminação de discriminações embutidas em modelos preditivos, ou até o descarte de dados processados indevidamente, como aqueles coletados sem base legal ou em violação ao princípio da não discriminação.

O direito à revisão transcende a mera formalidade processual, configurando-se como um mecanismo dinâmico que se integra organicamente aos princípios da LGPD, notadamente à responsabilização e prestação de contas (art. 6º, X), ao exigir que controladores demonstrem conformidade por meio de registros detalhados das operações de tratamento (art. 37), como logs de processos algorítmicos que permitam rastrear decisões, e relatórios de impacto à proteção de dados pessoais (art. 38), que prospectem riscos associados a sistemas de alto impacto, como plataformas de vigilância preditiva ou recomendação personalizada.

Muitos são os desafios à efetividade da norma, entre os quais é possível citar a opacidade de sistemas de IA, cuja complexidade cada vez maior dificulta explicações compreensíveis por parte dos agentes envolvidos no desenvolvimento do modelo; a tensão com segredos comerciais e industriais, muitas vezes utilizados como pretextos para se evitar o fornecimento de informações necessárias à compreensão e, conseqüentemente, revisão de decisões automatizadas; a ausência especificação de como e por quem a revisão deve ser feita, abrindo margem para que máquinas venham a ser encarregadas dessa tarefa; e a capacidade limitada da ANPD para

fiscalizar sistemas complexos. Além disso, a norma enfrenta o desafio de impactos coletivos de perfis de grupo, que, embora fora do escopo direto do artigo 20 por não envolverem dados pessoais, podem gerar discriminações em desfavor de sujeitos coletivos, como em sistemas de vigilância que estigmatizam comunidades marginalizadas, exigindo regulamentações complementares ou interpretações expansivas.

É importante ressaltar que o texto final do artigo 20 sofreu alterações significativas em relação à redação original. Inicialmente, havia previsão expressa de que a revisão seria realizada por natural, mas a Medida Provisória n. 869/2018 e a Lei n. 13.853/2019 suprimiram essa exigência, vetando o § 3º que previa intervenção humana sob regulamentação da autoridade nacional. Atualmente, a revisão pode ser automatizada, criando um ciclo potencialmente infinito de decisões viciadas, o que compromete a proteção dos titulares de dados. Passos e Oliveira (2024) destacam a incoerência lógica e técnica dessa abordagem devido às limitações matemáticas e a incapacidade de algoritmos de realizar juízos valorativos, reforçando a necessidade de intervenção humana.

Na Nota Técnica n. 12/2025, elaborada pela ANPD com base em Tomada de Subsídios sobre inteligência artificial e revisão de decisões automatizadas, registrou-se convergência nas manifestações colhidas sobre a necessidade de supervisão humana para a mitigação de riscos de discriminação ilícita ou abusiva.⁵⁷ Um contraponto a essa perspectiva é que, ainda com a supervisão humana, o racismo algorítmico pode se manifestar em razão dos vieses sociais/cognitivos dos agentes envolvidos na revisão. Conforme visto, os vieses se apresentam de diversas formas (tanto por humanos quanto por máquinas), de modo que deve se avaliar a alternativa que melhor serve ao propósito do direito de revisão. Frente à atual redação do artigo 20, predomina a visão de que a intervenção humana na revisão é essencial para eliminar vieses discriminatórios dos sistemas de IA.

Embora a estrutura normativa do artigo 20 reflita a influência direta do regulamento europeu, notadamente de seu artigo 22, há diferenças quanto a esse ponto e em relação ao próprio direito de revisão. O GDPR proíbe decisões tomadas exclusivamente com base em tratamento automatizado que gerem efeitos jurídicos ou

⁵⁷ ANPD. **ANPD apresenta resultados da Tomada de Subsídios sobre Inteligência Artificial e Revisão de Decisões Automatizadas**, 19 mai. 2025. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-apresenta-resultados-da-tomada-de-subsidios-sobre-tratamento-automatizado-de-dados-pessoais>. Acesso em: 5 ago. 2025.

impactem significativamente o titular, exceto em casos de consentimento explícito, contrato ou lei, com garantias como intervenção humana e explicação (Considerando 71). De maneira oposta, a LGPD atribui ao titular o direito de solicitar revisão de decisões automatizadas que afetem seus interesses e exige informações claras sobre critérios, com auditoria opcional pela autoridade nacional se negadas sob a justificativa de segredo comercial. Ambas demandam transparência, mas o GDPR foca em prevenção, enquanto a LGPD prioriza uma abordagem corretiva.

Em síntese, a previsão do art. 20 é “imprecisa porque não explicita como pode ser realizada e satisfeita essa revisão, nem tampouco especifica riscos das aplicações que merecem revisão” (Junquilha, 2023, p. 57). Por esse aspecto e outros explicitados, ela se mostra insuficiente e pouco apropriada para regular o perfilamento automatizado, principalmente em sua dimensão coletiva.

4.3. Perspectivas regulatórias e o Projeto de Lei n. 2.338/2023

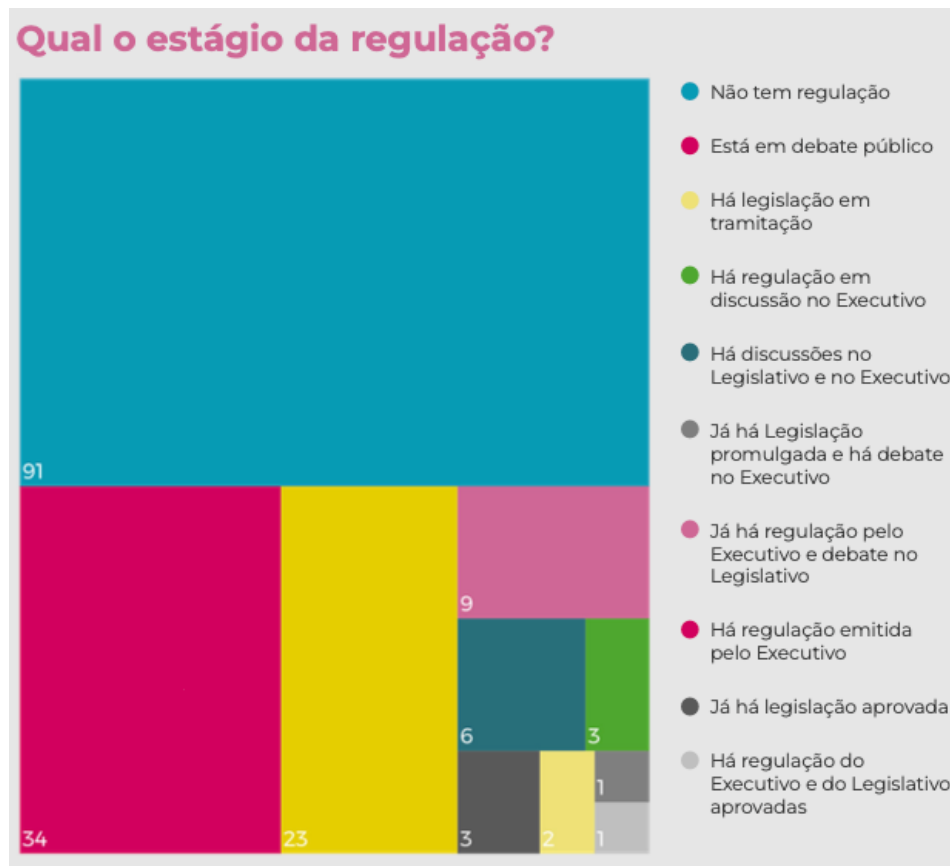
Os avanços tecnológicos observados nas últimas décadas na área da inteligência artificial apontam para um longo “verão” que não tem previsão para terminar. A cada dia que passa o uso IA se intensifica e seu potencial – tanto benéfico quanto nocivo – se revela perante a sociedade, tornando inevitável não falar sobre uma possível regulamentação a nível nacional. As razões para tanto são variadas, mas cabe citar os riscos associados a direitos fundamentais, como o direito à privacidade e à igualdade, e a insuficiência de sua regulamentação sob a ótica da proteção de dados.

Em uma pesquisa conduzida entre o segundo semestre de 2023 e o início de 2024, o TechLab⁵⁸ realizou levantamento a respeito da regulação da inteligência artificial ao redor do mundo.⁵⁹ Constatou-se que, dos 174 países mapeados, 46% (82 países) estavam em algum estágio regulatório, seja com legislação ou regulação pelo executivo aprovada ou em debate, seja com legislação em trâmite ou regulação em discussão no executivo, ou, ainda, em estado incipiente de debates públicos. Em contrapartida, 54% dos países analisados não tinham regulação a respeito da matéria, conforme se infere da Figura 5.

⁵⁸ Grupo de pesquisa da Faculdade de Direito da Universidade de São Paulo.

⁵⁹ TechLab. **Relatório de pesquisa: regulação da inteligência artificial ao redor do mundo.** Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4803041. Acesso em: 10 ago. 2025.

Figura 5 – Estágio da regulação da inteligência artificial ao redor do mundo



Fonte: Techlab, 2024, p. 6.

A pesquisa constatou que, nos países com regulação em curso ou já aprovada, a maioria das normas é emitida pelo Poder Executivo (51,76%), enquanto uma minoria envolve discussões sobre possível legislação da inteligência artificial no âmbito do legislativo (22,25%). Em 71% dos casos de regulação via Poder Executivo, o desenvolvimento tecnológico é priorizado, provavelmente tratado como política governamental desses países.

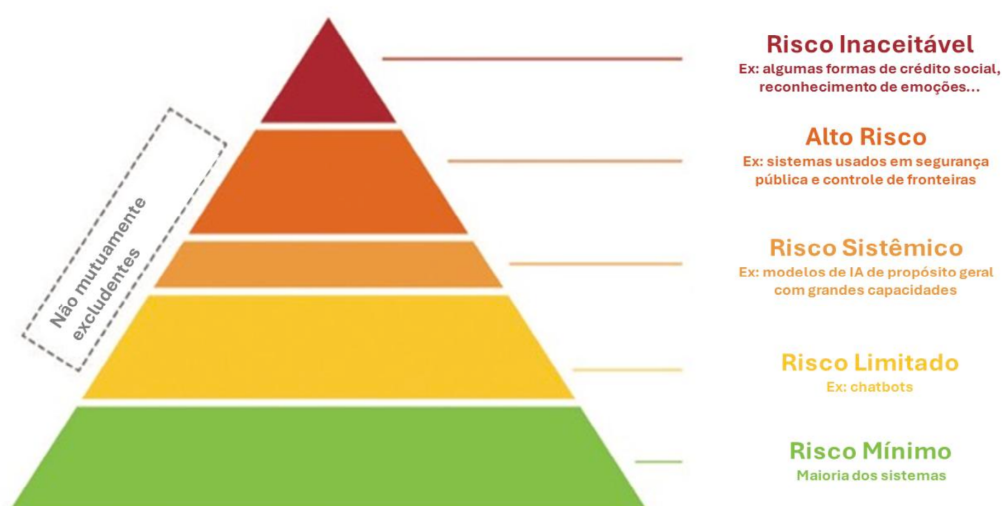
Verificou-se que 42% das regulações encontradas têm como foco principal a garantia de direitos, enquanto aquelas que priorizam pouco ou nada esse aspecto somam 18,25%. Ressalta-se que, entre as regulações voltadas à garantia de direitos, a maioria se concentra na proteção de dados (86%) e na prevenção de discriminação algorítmica (72%). Além disso, constatou-se que, em 42% das regulações, o desenvolvimento tecnológico prevalece sobre a garantia de direitos; em 32%, há um equilíbrio entre um e outro; e em 25%, a garantia de direitos é predominante.

Os números mostram uma tendência mundial à regulação da inteligência tanto por parte dos países do norte global, quanto por parte do sul global. É possível atribuir

a onda de iniciativas e discussões nesse sentido ao “Efeito Bruxelas” (Bradford, 2020)⁶⁰, notadamente em decorrência do *AI Act*, primeira regulamentação de inteligência artificial do mundo aprovada em 2024 na União Europeia. Alinhado aos preceitos da Indústria 5.0 e centrado no ser humano, o *AI Act* adota uma abordagem baseada em risco (inaceitável, alto, sistêmico, limitado e mínimo), classificando os sistemas de IA conforme seu potencial de impacto, de modo a aplicar obrigações proporcionais que variam desde proibições totais até requisitos mínimos.

Destacam-se nessa classificação os sistemas de risco inaceitável, notadamente aqueles que são usados para fins de classificação social, reconhecimento facial em tempo real em espaços públicos, manipulação cognitivo-comportamental e reconhecimento de emoções no local de trabalho e nos estabelecimentos de ensino, que são proibidos; de alto risco, como os usados em biometria, infraestrutura crítica, educação, emprego ou aplicação da lei, que estão sujeitos a requisitos rigorosos, incluindo gestão contínua de riscos, governança de dados para evitar vieses, documentação técnica detalhada, transparência, supervisão humana e robustez contra ataques cibernéticos; e de risco sistêmico, associado a modelos de IA de finalidade geral que apresentam capacidades de alto impacto, a exemplo do ChatGPT, que se submetem a avaliações rigorosas, transparência, governança robusta e supervisão contínua, conforme ilustrado na Figura 6.

Figura 6 – Tipos de sistemas de IA correspondentes a diferentes níveis de risco.



⁶⁰ É o nome dado por Anu Bradford ao fenômeno em que as regulamentações da União Europeia influenciam padrões globais, sendo adotadas por empresas e países fora da Europa, mesmo sem obrigação legal.

O modelo europeu, focado em proteger direitos fundamentais como privacidade, dignidade e igualdade, com o Estado atuando como guardião contra abusos das *big techs*, é descrito por Anu Bradford (2023) como modelo regulatório baseado em direitos (*Rights-Driven Regulatory Mode*). Paralelamente, Bradford sugere que existem outros dois modelos regulatórios dominantes: o modelo regulatório orientado pelo mercado (*Market-Driven Regulatory Model*) adotado pelos EUA e o modelo regulatório orientado pelo Estado (*State-Driven Regulatory Model*) em voga na China.

No caso do primeiro, prioriza-se a liberdade econômica, a inovação, o desenvolvimento tecnológico e a liberdade de expressão, com mínima intervenção governamental. Nos EUA, a regulamentação da IA apresenta um cenário fragmentado e descentralizado, ocorrendo principalmente por meio de iniciativas estaduais, ações executivas e aplicação de leis existentes. Entre as iniciativas a nível federal que tratavam sobre os riscos da IA, destaca-se a Ordem Executiva 14110 do Governo Biden, que visava o desenvolvimento e uso seguro, confiável e ético da inteligência artificial. No entanto, essa ordem foi revogada em janeiro de 2025 pela administração Trump⁶¹, que tem se aproximado abertamente das *big techs*⁶² e adotado uma abordagem desregulatória para fomentar a inovação e manter a competitividade global, evitando restrições excessivas.

No modelo que tem a China como principal representante, enfatiza-se o controle estatal para promover crescimento econômico, estabilidade social e soberania digital, subordinando a tecnologia aos interesses do Estado. Apesar de ser uma realidade que a China trata o progresso tecnológico como uma política estatal, existem interesses que transcendem o Estado, como o desenvolvimento econômico do país e a garantia de direitos, os quais são expressamente previstos em atos como o Plano de Desenvolvimento da Próxima Geração de Inteligência Artificial, publicado em 2017 (Techlab, 2024).

⁶¹ Shepardson, David. **Trump revokes Biden executive order on addressing AI risks**. Reuters, Washington, 20 jan. 2025. Disponível em: <https://www.reuters.com/technology/artificial-intelligence/trump-revokes-biden-executive-order-addressing-ai-risks-2025-01-21/>. Acesso em: 4 ago. 2025.

⁶² Lopes, André. **Os novos amigos do presidente: as razões que uniram as big techs a Trump**. Exame, 11 jan. 2025. Disponível em: <https://exame.com/tecnologia/os-novos-amigos-do-presidente-as-razoes-que-uniram-as-big-techs-a-trump/>. Acesso em: 4 ago. 2025.

Outras abordagens sobre possíveis modelos regulatórios são discutidas, especialmente para dar conta de um cenário mais amplo e global, e de realidades que fogem das principais potências mundiais. Petit e Cooman (2020) propõem a regulação da IA a partir de variados modelos, que abrange desde uma regulamentação baseada princípios éticos para informar o desenvolvimento e a aplicação da IA, oferecendo flexibilidade para evoluções tecnológicas e mantendo a regulação funcional por mais tempo (modelo ético) até uma que antecipa a evolução tecnológica e sugere a criação de novas normas específicas para disciplinar a IA (modelo emergente).

E onde o Brasil se posiciona dentro desse cenário? No caso da maior potência econômica da América Latina⁶³, as discussões sobre a regulamentação da inteligência artificial estão concentradas no Projeto de Lei n. 2.338/2023⁶⁴, de autoria do senador Rodrigo Pacheco. Inspirado no regulamento europeu, mas com adaptações feitas para se adequar à realidade local, o modelo proposto combina elementos de regulação baseada em risco, como a classificação de sistemas de IA em níveis de risco (excessivo e alto), com princípios éticos que priorizam a centralidade da pessoa humana, transparência e proteção de direitos fundamentais.

O projeto de lei é fruto da Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil (CJSUBIA) que, no ano de 2022, elaborou anteprojeto para subsidiar as discussões no Congresso Nacional, visto que os principais projetos que tramitavam anteriormente (PL n. 5.051/2019⁶⁵, proposto pelo senador Styvenson Valentim; PL n. 21/2020⁶⁶, proposto pelo deputado federal Eduardo Bismarck; e o PL n. 872/2021⁶⁷, proposto pelo senador Veneziano

⁶³ Caveiro, Juliana. **Brasil ocupa primeira posição entre as maiores economias da América Latina, diz FMI; confira a lista.** Money Times, 19 fev. 2024. Disponível em: <https://www.moneytimes.com.br/brasil-ocupa-primeira-posicao-entre-as-maiores-economias-da-america-latina-diz-fmi-confira-a-lista/>. Acesso em: 10 jun. 2025.

⁶⁴ Brasil. **Projeto de Lei nº 2.338, de 2023.** Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>. Acesso em: 6 ago. 2025.

⁶⁵ Brasil. **Projeto de Lei nº 5.051, de 2019.** Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Brasília, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 6 ago. 2025.

⁶⁶ Brasil. **Projeto de Lei nº 21, de 2020.** Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília, 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236340>. Acesso em: 6 ago. 2025.

⁶⁷ Brasil. **Projeto de Lei nº 872, de 2021.** Dispõe sobre o uso da Inteligência Artificial. Brasília, 2021. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/147434>. Acesso em: 28 maio 2023.

Vital do Rego) eram demasiadamente simples e insuficientes para regular a IA de forma eficiente.

As contribuições e discursos em torno da Comissão de Juristas e do substitutivo posteriormente consolidado é objeto de aprofundada análise feita por Silva (2025). Ele aponta que, à luz do Estatuto da Igualdade Racial e da Convenção Interamericana Contra o Racismo, a Discriminação Racial e Formas Correlatas de Intolerância, o Estado brasileiro assumiu compromisso no combate ao racismo em todas as suas formas e na promoção de políticas públicas e institucionais em favor de grupos racialmente marginalizados. Nada obstante,

os discursos hegemônicos sobre a relação entre inteligência artificial e racismo em torno da produção do PL 2338 ignoraram o acúmulo apresentado sobre o tema, assim como ignoraram também as propostas de mitigação e compromissos do Estado brasileiro na superação das desigualdades raciais (Silva, 2025, p. 100).

Sobre a versão do projeto consolidada e submetida à revisão da Câmara dos Deputados, não existe disciplina específica quanto à prática de perfilamento automatizado, mas tão somente indicação de que a identificação e definição de perfis comportamentais em contexto de investigação ou de repressão representa uma atividade de alto risco e, por consequência, está sujeita aos termos do regulamento. Por outro lado, há de se reconhecer como pontos importantes a categorização de riscos, o reconhecimento de direitos de grupos afetados por sistema de IA, previsão de avaliações de impacto, reconhecimento da discriminação indireta e instrumentos de participação social.

Quanto às categorias já existentes em normas correlatas, a exemplo da LGPD, o PL n. 2.338/2023 reforça a importância de princípios centrais como transparência, explicabilidade, não discriminação, prestação de contas e responsabilização, assim como dos direitos à contestação e à revisão com a possibilidade de supervisão humana. Deve-se ponderar, no entanto, que a implementação dessas medidas em sistemas complexos pode representar desafios técnicos significativos, requerendo padrões mais elaborados.

Entre o projeto atual aprovado em dezembro de 2024 e a minuta entregue pela Comissão de Juristas em 2022, Silva (2025) destaca os seguintes pontos:

- a) Negacionismo sobre os impactos do reconhecimento facial no espaço público e ampliação de seus usos. Em relatório preliminar, a comissão já

havia afirmado, contra as evidências, que a “utilização massiva [de identificação biométrica à distância] tem demonstrado que, em regra, esses sistemas não têm potencial para causar dano significativo”. Na versão final do projeto, as exceções para o uso da tecnologia para fins de segurança pública não diferem do que já está em realização no Brasil hoje apesar de estar oficialmente inclusa como “risco excessivo”;

b) Delimitação estreita de discriminação indireta. O termo relacionado a discriminação indireta foi mantido, apesar de um estreitamento do que é considerado: “a capacidade de acarretar desvantagem para pessoa ou grupos afetados, ou os coloquem em desvantagem, desde que essa normativa, prática ou critério seja abusivo ou ilícito”;

c) Desobrigação de realização de avaliação preliminar. A minuta estabelece que o “agente de IA poderá realizar avaliação preliminar para determinar o grau de risco do sistema”, atividade considerada apenas como indicador de boas práticas. Assim, o projeto de lei coloca sob responsabilidade do próprio agente desenvolvedor a avaliação do nível de risco da tecnologia, abrindo a possibilidade que obrigações posteriores sejam evadidas.

d) Exclusão de itens sobre orientação à diversidade. Entre as medidas de gestão e governança indicadas aos sistemas de alto risco, o projeto anterior incluía, ainda que de forma vaga, a recomendação de “composição de equipe inclusiva responsável pela concepção e desenvolvimento do sistema”, trecho removido da versão da CTIA.

e) Retirada de mecanismos de participação social. O projeto final da CTIA retirou mecanismos de participação social explícitos nas entidades supervisoras e nas avaliações de impacto (Silva, 2025, p. 152).

É inegável que existe um esforço tanto para tratar tanto do perfilamento quanto do combate a práticas discriminatória, o que abrange, por óbvio, o racismo nos sistemas de IA. Convém mencionar, por exemplo, a previsão de discriminação indireta (art. 4º, inciso XII), tão importante no contexto de racismo perpetuado por sistemas de IA, e a classificação de alto risco para o sistema de IA que é utilizado no âmbito de investigação por autoridades administrativas.

Embora a raça e o racismo tenham sido considerados para o desenho do modelo regulatório ora em trâmite na Câmara dos Deputados, sua abordagem é através do guarda-chuva da “diversidade” e do princípio da não discriminação. Ainda que essa abordagem não esteja no todo errada, não é capaz de compreender a dimensão estrutural e sistêmica do problema, além de ignorar os compromissos firmados pelo Brasil. Inegável que a regulação da IA é um dos meios mais eficazes para mitigação do racismo algorítmico, especialmente de práticas como o perfilamento automatizado, todavia, um regulamento deficiente em aspectos essenciais é a receita perfeita para adversidades futuras.

5. CONSIDERAÇÕES FINAIS

A emergência da inteligência artificial alterou de maneira definitiva o curso do avanço científico e tecnológico nas últimas décadas, trazendo alterações significativas na dinâmica social e na maneira como os indivíduos se desenvolvem. Atualmente, essa tecnologia, que é tanto fluida quanto intrincada, está presente em uma ampla gama de plataformas, dispositivos e sistemas digitais, sendo utilizada em diversas atividades do dia a dia, que vão desde o uso de redes sociais — um espaço onde costuma-se mapear nosso perfil de preferências para a realização de previsões e criação de anúncios personalizados — até a implementação de métodos de reconhecimento facial por parte do Estado como apoio à segurança pública.

Por um lado, a implementação de sistemas de IA nas atividades de empresas, órgãos governamentais e na sociedade proporcionou o aperfeiçoamento e a simplificação de processos que antes eram extremamente complexos. Por outro lado, fez surgir inquietações e questionamentos devido aos impactos negativos sobre as pessoas. O uso indiscriminado de sistemas inteligentes, que se “alimentam” de quantidades massivas de dados pessoais, trouxe à tona os potenciais nocivos da IA e os riscos a direitos fundamentais, como a igualdade, a privacidade e o livre desenvolvimento da personalidade, além de ter desvelado a face de um problema estrutural que se perpetua no ambiente digital: o racismo. É nesse contexto que se insere o presente trabalho, cuja pretensão é analisar a prática de perfilamento automatizado como manifestação do racismo algorítmico.

O primeiro capítulo estabelece uma base técnica essencial, explorando conceitos como inteligência artificial – a partir de suas formas e sentidos diversos –, dados pessoais, big data e algoritmos, além de abordar o aprendizado de máquina e a inteligência artificial generativa, fornecendo uma compreensão inicial dos fundamentos tecnológicos que sustentam os sistemas discutidos ao longo da dissertação. Destaca-se a evolução e a autonomia dos sistemas inteligentes, que estão cada vez mais complexos e imbricados em atividades do cotidiano. Nesse cenário, a compreensão do funcionamento dos algoritmos de IA foi crucial para abordar o problema objeto deste trabalho.

O segundo capítulo focou na análise do racismo algorítmico e do perfilamento automatizado, utilizando autores que produzem na área de direito, tecnologia e raça

como lente interpretativa. Discutiu-se como os vieses discriminatórios, decorrentes da não neutralidade tecnológica, podem perpetuar desigualdades, especialmente contra a população negra. A construção do conceito de racismo algorítmico foi detalhada, evidenciando sua manifestação em tecnologias digitais, com ênfase no perfilamento automatizado. Esse capítulo trouxe uma reflexão sobre as raízes históricas e estruturais do racismo, conectando-as aos desdobramentos contemporâneos no uso de algoritmos, o que reforça a necessidade de abordagens regulatórias sensíveis a essas dinâmicas discriminatórias.

O terceiro capítulo foca na regulação do perfilamento automatizado no contexto jurídico brasileiro, avaliando mecanismos existentes e propostos para proteger a população negra contra os riscos discriminatórios. A LGPD, embora constitua um marco na proteção de dados no Brasil, apresenta lacunas no que diz respeito a disposições específicas que abordem diretamente os riscos discriminatórios associados ao perfilamento automatizado, notadamente em sua dimensão coletiva. Reconhece-se que as ferramentas oferecidas pela LGPD são importantes para regulação da IA sob a ótica da proteção de dados, todavia, são limitadas e insuficientes para mitigação do racismo algorítmico. O direito de revisão de decisões automatizadas enfrenta barreiras, como seu foco predominantemente individual e associado à pessoa identificada ou identificável; obstáculos à explicabilidade como uma etapa essencial à transparência; e a ausência de previsão de supervisão humana. De outro lado, o projeto de lei em tramitação, apesar de promissor, enfrenta desafios relacionados à ausência de enfrentamento do racismo enquanto um fator chave para se pensar em políticas institucionais e, especificamente, no desenho do modelo regulatório a ser adotado no Brasil.

O combate ao racismo algorítmico exige uma atuação antirracista e uma estratégia contínua e adaptável, que transcenda a simples regulamentação para promover uma transformação cultural e tecnológica mais ampla. A integração de princípios de justiça racial nas etapas de design, desenvolvimento e uso de sistemas de IA pode ser um caminho viável para reduzir vieses e garantir equidade, exigindo a colaboração entre o setor público, empresas de tecnologia e sociedade. Apesar dos progressos observados, o caminho à frente permanece longo e complexo, demandando esforços conjuntos para monitorar a eficácia das medidas existentes, ajustar legislações conforme necessário e desenvolver novas soluções tecnológicas que priorizem a inclusão racial.

O presente trabalho reforça a ideia de que a regulação da IA deve ser um processo dinâmico, sensível às especificidades culturais e históricas do Brasil, e comprometido com a redução das disparidades sociais. A proteção da população negra contra os riscos do perfilamento automatizado não é apenas uma questão técnica ou jurídica, mas um imperativo ético que reflete o compromisso com uma sociedade mais justa. Assim, este estudo contribui para um debate mais informado e propositivo, incentivando a continuidade da pesquisa e a mobilização de recursos para enfrentar os desafios impostos pela interseção entre tecnologia e racismo no século XXI.

6. REFERÊNCIAS

AEPD. Agencia Española Protección Datos. **Synthetic data and data protection**. 2023. Disponível em <https://www.aepd.es/en/prensa-y-comunicacion/blog/synthetic-data-and-data-protection>. Acesso em: 14 jul. 2025.

AKUNDI, Aditya; EURESTI, Daniel; LUNA, Sergio; ANKOBIAH, Wilma; LOPES, Amit; EDINBAROUGH, Immanuel. **State of industry 5.0 – analysis and identification of current research trends**. Applied System Innovation, v. 5, n. 1, p. 27, 2022. Disponível em: <https://www.mdpi.com/2571-5577/5/1/27>. Acesso em: 3 jul. 2025.

ALMEIDA, Silvio. **O que é racismo estrutural**. Coleção Feminismos Plurais. Djamila Ribeiro (Org.) São Paulo: Sueli Carneiro; Pólen, 2019.

ALPAYDIN, Ethem. **Introduction to machine learning**. 5 ed. MIT Press, 2014.

BARROS, Atila. **Desafios éticos na utilização de algoritmos de recrutamento do LinkedIn e seu impacto na equidade de oportunidades de emprego**. Revista Tópicos, 2(7). Disponível em: <https://doi.org/10.5281/zenodo.10790850>. Acesso: 18 jun. 2025.

BATOCCHIO, Amália. **Natureza jurídica dos dados pessoais e suas implicações**. Dissertação (Mestrado) – Universidade de São Paulo, São Paulo, 2023. Disponível em: <https://www.teses.usp.br/teses/disponiveis/2/2139/tde-07032024-092810/>. Acesso em: 10 mar. 2025

BENJAMIN, Ruha. **Race after technology: abolitionist tools for the new jim code**. Social Forces, 2019.

BIONI, Bruno. **Proteção de dados pessoais: a função e os limites do consentimento**. 2ª. ed. Rio de Janeiro: Forense, 2020.

BIONI, Bruno; RIELLI, Mariana; KITAYAMA, Marina. **Colocando em movimento o legítimo interesse**. In: BIONI, Bruno (coord.). Proteção de dados: contexto, narrativas e elementos fundantes [livro eletrônico]. São Paulo: B. R. Bioni Sociedade Individual de Advocacia, 2021, pp. 200-243.

BITENCOURT, Rosimeire Sedrez; OKUMURA, Maria Lucia Miyake. **Um panorama da indústria 5.0: o resgate do fator humano**. In: XX Congresso Brasileiro de Ergonomia - Virtual 2020. Anais. Lorena (SP) ABERGO, 2020. Disponível em: <https://www.even3.com.br/anais/abergo2020/295493-UM-PANORAMA-DA-INDUSTRIA-50--O-RESGATE-DO-FATOR-HUMANO>. Acesso em: 22/08/2025

BOSCO, Francesca *et. al.* **Profiling technologies and fundamental rights: an introduction**. In: CREEMERS, Niklas *et. al.* Profiling Technologies in Practice: Applications and Impact on Fundamental Rights and Values. Oisterwijk: Wolf Legal Publishers, 2017.

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

_____. **Digital empires: the global battle to regulate technology**. 1ª ed. New York: Oxford University Press. 2023.

BRASIL. Autoridade Nacional de Proteção de Dados. **Estudo Técnico Sobre a Anonimização de Dados na LGPD: Análise Jurídica**. Disponível em: https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/estudo_tecnico_sobre_anonimizacao_de_dados_na_lgpd_analise_juridica.pdf. Acesso em: 25 fev. 2025.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 10 mar. 2025.

BRASIL. Medida Provisória nº 954, de 17 de abril de 2020. **Dispõe sobre o compartilhamento de dados por empresas de telecomunicações prestadoras de Serviço Telefônico Fixo [...]**. Brasília, DF, [2020]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/mpv/mpv954.htm. Acesso em: 02 jan. 2025.

BRASIL. **Projeto de Lei nº 5.051, de 2019**. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Brasília, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 6 ago. 2025.

BRASIL. **Projeto de Lei nº 21, de 2020**. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Brasília, 2020. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236340>. Acesso em: 6 ago. 2025.

BRASIL. **Projeto de Lei nº 872, de 2021**. Dispõe sobre o uso da Inteligência Artificial. Brasília, 2021. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/147434>. Acesso em: 28 maio 2023.

BRASIL. **Projeto de Lei nº 2.338, de 2023**. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, 2023. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>. Acesso em: 6 ago. 2025.

BYGRAVE, Lee A. **Data protection law: approaching its rationale, logic and limits**. The Hague: Kluwer Law International, 2002.

CARNEIRO, Sueli. **Dispositivo de racialidade: a construção do outro como não ser como fundamento do ser**. Rio de Janeiro: Zahar, 2023.

CHAHAD, José Paulo Zeetano; ISSA, Tuffy Licciardi. **O estágio atual da inteligência artificial: história, impactos no mercado de trabalho, aspectos sociais e alternativas de regulação (parte i – aspectos históricos e impactos no mercado de trabalho)**. INFORMAÇÕES FIPE, v. 520, p. 15-40, 2024.

CHRISTAKIS, Theodore; KARATHANASIS, Theodoros. **Tools for navigating the eu ai act: (2) visualisation pyramid**. AI Regulation, 7 Mar. 2024, disponível em <<https://ai-regulation.com/visualisation-pyramid/>>, acesso em 9 ago. 2025.

CLARKE, Roger. **Profiling: a hidden challenge to the regulation of data surveillance**. Journal of Law and Information Science, v. 4, n. 2, p. 403-419, 1993.

CONSELHO DA EUROPA. **The protection of individuals with regard to automatic processing of personal data in the context of profiling: recommendation cm/rec(2010)13 and explanatory memorandum**. Estrasburgo: Council of Europe Publishing, 2011.

CRAWFORD, Kate. **Atlas of AI: power, politics, and the planetary costs of artificial intelligence**. New Haven: Yale University Press, 2021.

CRIOULA, Bruna. **Quando o algoritmo é cruel: uma mulher negra e a violência silenciosa do LinkedIn**. Disponível em: <https://brunacrioula.medium.com/quando-o-algoritmo-%C3%A9-cruel-uma-mulher-negra-e-a-viol%C3%Aancia-silenciosa-do-linkedin-d387a614d3aa>. Acesso em: 28 jun. 2025.

DONEDA, Danilo; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes de. **Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal**. Fortaleza: Revista Pensar, vol. 23, no 4, p. 1-17, out-dez 2018.

DOUGLAS, John; OLSHAKER, Mark. **Mindhunter: inside the FBI's elite serial crime unit**. New York: Scribner, 1997.

EUROPEAN DATA PROTECTION SUPERVISOR (EDPS). **TechDispatch #2/2023 - explainable artificial intelligence**. Disponível em: https://www.edps.europa.eu/system/files/2023-11/23-11-16_techdispatch_xai_en.pdf. Acesso em: 24 jul. 2025.

FEENBERG, A. **Critical theory of technology**. New York: Oxford University Press, 1991.

FRAZÃO, Ana. **Discriminação algorítmica: os algoritmos como a última fronteira da “matematização” do mundo**. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-3-30062021>. Acesso em: 10 nov. 2024.

GOODFELLOW, Ian; POUGET-ABADIE, Jean; MIRZA, Mehdi; XU, Bing; WARDEFARLEY, David; OZAIR, Sherjil; COURVILLE, Aaron; BENGIO, Yoshua (2014). **Generative Adversarial Networks**. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf. Acesso em: 15 jul. 2025.

HILDEBRANDT, Mireille. **Defining profiling: a new type of knowledge?** In: Gutwirth, Serge; Hildebrandt, Mireille (Coords.). **Profiling the European Citizen: Cross-Disciplinary Perspectives**. New York: Springer, 2008. p. 17-45,

JUNQUILHO, Tainá Aguiar. **Aplicação de inteligência artificial ao direito: análise de parâmetros da literatura, da regulação e da percepção de atores sobre limites éticos**. 2023. 167 f., il. Tese (Doutorado em Direito) - Universidade de Brasília, Brasília, 2023.

KAUFMAN, Dora. **A inteligência artificial irá suplantará a inteligência humana?** Barueri: Estação das Letras e Cores, 2019.

_____. **Desmitificando a inteligência artificial**. Belo Horizonte: Autêntica, 2022.

KREMER, Bianca. **Direito e tecnologia em perspectiva africana: autonomia, algoritmos e vieses raciais**. Tese de doutorado em Direito na Pontifícia Universidade Católica do Rio de Janeiro, 2021.

LAGE, Fernanda de Carvalho. **A inteligência artificial na repercussão geral: análise e proposições da vanguarda de inovação tecnológica no Poder Judiciário brasileiro**. 2020. 279 f., il. Tese (Doutorado em Direito) — Universidade de Brasília, Brasília, 2020.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. **Deep learning**. Nature, v. 521, p. 436–444, 2015.

LESSING, Lawrence. **The Architecture of Privacy**. 1998.

LILLA, Marcos. **A Revolução digital: redefinindo organizações e gerando valor**. Disponível em: <https://www.linkedin.com/pulse/revolu%C3%A7%C3%A3o-digital-redefinindo-organiza%C3%A7%C3%B5es-e-gerando-valor-lilla?originalSubdomain=pt>. Acesso em: 18 jun. 2025.

LINDOSO, Maria Cristine Branco. **Discriminação de gênero em processos decisórios automatizados**. Dissertação de mestrado em Direito na Universidade de Brasília. 2019. Disponível em: <https://repositorio.unb.br/bitstream/10482/38524/1/2019_MariaCristineBrancoLindoso.pdf> Acesso em: 10 mar. 2025.

LOMBROSO, Cesare. **O homem delinquente**. Tradução: Sebastian José Roque. 1. Reimpressão. São Paulo: Ícone, 2010.

MACHADO, Diego Carvalho. **A regulação das tecnologias de perfilamento no direito brasileiro: articulando direito e tecnologia para a promoção da proteção de dados desde a concepção**. 2022. 292 f. Tese (Doutorado em Direito) - Faculdade de Direito, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2022.

_____. **Considerações iniciais sobre o conceito de dado pessoal no ordenamento jurídico brasileiro**. Civilistica.com, Rio de Janeiro, v. 12, n. 1, p. 1–34, 2023, pp. 6-7. Disponível em: <https://civilistica.emnuvens.com.br/redc/article/view/843>. Acesso em: 21 fev. 2025.

MARQUES, André Ferreira. **Inteligência artificial: regulação ética e responsabilidade civil**. Dissertação (Mestrado) – Universidade de Marília, Marília,

2020. Disponível em: <https://dspace.unimar.br/handle/123456789/939>. Acesso em: 11 junho de 2025.

MARTINS, Pedro Basto Lobo. **A regulação do profiling na lei geral de proteção de dados: o livre desenvolvimento da personalidade em face da governamentalidade algorítmica**. Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Belo Horizonte, 2021. Disponível em: <https://repositorio.ufmg.br/handle/1843/43900>. Acesso em: 14 jul. 2025.

MARX, Gary T.; REICHMAN, Nancy. **Routinizing the discovery of secrets: computers as informants**. The American Behavioral Scientist, v. 27, n. 4, p. 423–452, 1984

MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth. **Big data. Como extrair volume, variedade, velocidade e valor da avalanche de informação cotidiana**. 1ª. ed., Rio de Janeiro: Elsevier, 2013.

MCCARTHY, John, *et al.* **A proposal for the Dartmouth summer research project on artificial intelligence**. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 21 fev. 2025

MCCULLOCH, Warren S.; PITS, Walter. **A logical calculus of the ideas immanent in nervous activity**. The bulletin of mathematical biophysics, 5(4): 115–133, 1943.

MEI, Katelyn X.; FEREIDOONI, Sonia; CALISKAN, Aylin. **Bias against 93 stigmatized groups in masked language models and downstream sentiment classification tasks (2023)**. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/3593013.3594109>. Acesso em: 23 jul. 2025.

MENDES, Laura Schertel; MATIUZZO, Marcela. **Discriminação algorítmica: conceito, fundamento legal e tipologia**. RDU, Porto Alegre, Volume 16, n. 90, 2019, 39-64, nov-dez 2019.

MENDES, Laura Schertel; MATTIUZZO, Marcela; FUJIMOTO, Mônica Tiemy. **Discriminação algorítmica à luz da Lei Geral de Proteção de Dados**. In: DONEDA, Danilo (coord.); SARLET, Ingo Wolfgang (coord.); MENDES, Laura Schertel (coord.); RODRIGUES JUNIOR, Otavio Luiz (coord.) BIONI, Bruno Ricardo (coord.). Tratado de Proteção de Dados Pessoais. Rio de Janeiro: Forense, 2021, p. 429-454.

MINSKY, Marvin. **Communication with alien intelligence**. Cambridge, Cambridge University Press, 1985

MONAGREDA, Johanna K. **Por que falar de raça quando falamos de dados pessoais, inteligência artificial e algoritmos?** In: GONÇALVES, Adriana; TORRE, Luísa; MELO, Paulo Victor (eds.) Inteligência artificial e algoritmos. Covilhã: Editora LabCom, 2024.

MULHOLLAND, Caitlin. **Dados pessoais sensíveis e a tutela de direitos fundamentais: uma análise à luz da Lei Geral de Proteção de Dados (Lei 13.709/18)**. Revista de direitos e Garantias Fundamentais, v. 19, nº 3, 29 dez. 2018, p. 159-180.

_____. **Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (IA): autonomia, imputabilidade e responsabilidade.** In: FRAZÃO, Ana de Oliveira; MULHOLLAND, Caitlin (Coords.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. 2. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020. pp. 327-350.

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z. **Inteligência artificial e Lei Geral de Proteção de Dados Pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de machine learning.** In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coords.). *Inteligência artificial e Direito: ética, regulação e responsabilidade*. 2. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020. pp. 183-202.

NASCIMENTO, Abdias. **O genocídio do negro brasileiro: processo de um racismo mascarado.** São Paulo: Editora Perspectiva SA, 2016.

NOBLE, Safiya Umoja. **Algorithms of oppression: how search engines reinforce racism.** NYU Press, 2018.

NOBLE, Safiya; ROBERTS, Sarah T (2020). **Elites tecnológicas, meritocracia e mitos pós raciais no Vale do Silício.** *Fronteiras - estudos midiáticos*, 22(1), 36-46. Disponível em: <<https://revistas.unisinos.br/index.php/fronteiras/article/view/fem.2020.221.04>>. Acesso em: 16 dez. 2024.

O'NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia.** Tradução Rafael Abrahan. Santo André, SP: Editora Rua do Sabão, 2020.

ONU. **Declaração e Programação de Ação da III Conferência Mundial Contra o Racismo, Discriminação Racial, Xenofobia e Intolerâncias Correlatas.** África do Sul, 2001. Disponível em https://brasil.un.org/sites/default/files/2021-10/declaracao_durban.pdf. Acesso em: 27 jul. 2025.

PASQUALE, Frank. **The black box society. The secret algorithms that control money and information.** Cambridge: Harvard University Press, 2015.

PASSOS, Iago Moreno de Lucas; OLIVEIRA, Khadja Vanessa Brito de. **Uma análise sobre discriminação algorítmica no Brasil à luz das soluções apresentadas pela Lei Geral de Proteção de Dados (LGPD).** In: BARBOSA, Bia; TRESCA, Laura; RONCARATI, Luanna; TENÓRIO, Mozart; MIELLI, Renata; LAUSCHNER, Tanara (Coords.). *TIC, Governança da Internet, Gênero, Raça e Diversidade: tendências e desafios*. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2024. pp. 105-124.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. **Inteligência artificial e direito.** Curitiba: Alteridade, 2019.

PETIT, Nicolas; COOMAN, Jerome De. **Models of law and regulation for AI.** EUI Working Papers, RSCAS n. 63, 2020. Disponível em:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3706771. Acesso em: 28 jul. 2025.

QUIJANO, Anibal. **Colonialidade do poder, Eurocentrismo e América Latina**. Disponível em: https://biblioteca.clacso.edu.ar/clacso/sur-sur/20100624103322/12_Quijano.pdf. Acesso em: 6 mai. 2025.

RAMOS, Anátalia Saraiva Martins. **Inteligência artificial generativa baseada em grandes modelos de linguagem-ferramentas de uso na pesquisa acadêmica**. Pré-impressões SciELO; 2023. DOI: 10.1590/scielopreprints.6105. Disponível em: https://www.researchgate.net/publication/371026954_Inteligencia_Artificial_Generativa_baseada_em_grandes_modelos_de_linguagem_-_ferramentas_de_uso_na_pesquisa_academica. Acesso em: 16 jul. 2025.

REIS, Nazareno César Moreira; FURTADO, Gabriel Rocha. **Decisões automatizadas: definição, benefícios e riscos**. Civilistica.com, Rio de Janeiro, v. 11, n. 2, p. 1–44, 2022. Disponível em: <https://civilistica.emnuvens.com.br/redc/article/view/763>. Acesso em: 29 jul. 2025.

RIEKE, Aaron; YU, Harlan; Robinson, DAVID; HOBOKEN, Joris von. **Data brokers in an open society**. 2016. Disponível em: <https://www.opensocietyfoundations.org/uploads/42d529c7-a351-412e-a065-53770cf1d35e/data-brokers-in-an-open-society-20161121.pdf>. Acesso em: 19 jul. 2025.

RODOTÀ, Stefano. **A vida na sociedade da vigilância: a privacidade hoje**. Tradução Danilo Doneda e Luciana Cabral Doneda. São Paulo: Renovar, 2008.

ROTHENBURG, Walter Claudius. **Igualdade material e discriminação positiva: o princípio da isonomia**. Novos Estudos Jurídicos, Itajaí- (SC), v. 13, n. 2, p. 77–92, 2009. DOI: 10.14210/nej.v13n2.p77-92. Disponível em: <https://periodicos.univali.br/index.php/nej/article/view/1441>. Acesso em: 21 jul. 2025.

RUSSELL, Stuart; NORVIG, Peter. **Artificial intelligence: a modern approach**. Pearson, 2013.

RUSSELL, Stuart. **Human Compatible: Artificial Intelligence and the Problem of Control**. New York: Viking; Penguin, 2019.

SANTAELLA, Lucia; KAUFMAN, Dora. **A Inteligência artificial generativa como quarta ferida narcísica do humano**. MATRIZES, São Paulo, Brasil, v. 18, n. 1, p. 37–53, 2024. DOI: 10.11606/issn.1982-8160.v18i1p37-53. Disponível em: <https://www.revistas.usp.br/matrizes/article/view/210834>. Acesso em: 17 jul. 2025.

SCHWAB, Klaus. **A quarta revolução industrial**. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

SCHWARTZ, Reva; VASSILEV, Apostol; GREENE, Kristen; PERINE, Lori; BRUT, Andrew; HALL, Patrick. **NIST special publication 1270: towards a standard for identifying and managing bias in artificial intelligence**. Disponível em:

<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf>. Acesso em: 17 jul. 2025.

SEARLE, John R. **Mente, linguagem e sociedade: filosofia no mundo real**. Tradução: F. Rangel. Rio de Janeiro: Rocco, 2000.

SELBST, Andrew D.; POWLES, Julia. **Meaningful information and the right to explanation**. In: International Data Privacy Law, 233, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125. Acesso em: 4 ago. 2025.

SIENA, David Pimentel Barbosa de. **Qual é a natureza jurídica do perfilamento criminal?**. Consultor Jurídico. Disponível em: <https://www.conjur.com.br/2025-fev-06/qual-e-a-natureza-juridica-do-perfilamento-criminal/>. Acesso em: 21 mar. 2025.

SILVA, Nilton Correia da. **Inteligência artificial**. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coords.). Inteligência artificial e Direito: ética, regulação e responsabilidade. 2. ed. São Paulo: Revista dos Tribunais, 2020. pp. 1-17.

SILVA, Tarcízio. **Racismo algorítmico em plataformas digitais: microagressões e discriminação em código**. In: Comunidades, algoritmos e ativismos digitais. Tarcizio Silva (Org.) São Paulo: LiteraRua, 2020.

_____. **Racismo algorítmico: inteligência artificial e discriminação nas redes digitais**. 1ª Edição. São Paulo: Edições Sesc SP, 2022.

_____. **Racismo algorítmico e regulação de inteligência artificial: o contrato racial na produção do PL 2338/2023**. Tese (Doutorado) – Universidade Federal do ABC, Programa de Pós-Graduação em Ciências Humanas e Sociais, São Bernardo do Campo, 2025.

SPERLING, Ed. **Deep learning spreads**. Disponível em: <https://semiengineering.com/deep-learning-spreads/>. Acesso em: 22 jun. 2025.

STEIBEL, Fabro; VICENTE, Victor Freitas; DE JESUS, Diego Santos Vieira. **Possibilidades e potenciais da utilização da inteligência artificial**. In: FRAZÃO, Ana de Oliveira; MULHOLLAND, Caitlin (Coords.). Inteligência artificial e direito: ética, regulação e responsabilidade. 2. ed. São Paulo: Thomson Reuters Revista dos Tribunais, 2020. pp. 18-38.

TECHLAB. **Relatório de pesquisa: regulação da inteligência artificial ao redor do mundo**. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4803041. Acesso em: 10 ago. 2025.

TURING, Alan M. **Computing machinery and intelligence**. In: Mind, v. 59, p. 433-460, 1950.

UNESCO. **Guia para a IA generativa na educação e na pesquisa**. França: UNESCO, 2024. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000390241>. Acesso em: 14 jun. 2025.

UNIÃO EUROPEIA. **Regulamento (UE) nº 2016/679 do Parlamento Europeu e do Conselho, de 23 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados)**. Jornal Oficial da União Europeia, Estrasburgo, 04/05/2016. Disponível em: <<https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&from=PT>>. Acesso em: 4 fev. 2025.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N; KAISER, Łukasz; POLOSUKHIN, Illia (2017). **Attention is all you need**. Disponível em: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>. Acesso em: 15 jul. 2025.

WARREN, Samuel D.; BRANDEIS, Louis D. **The right to privacy**. Harvard Law Review, v. IV, n. 5, December, 1890. Disponível em: <<https://faculty.uml.edu/sgallagher/Brandeisprivacy.htm>>. Acesso em: 26 out. 2024.

WINNER, Langdon. **Do artifacts have politics?**. Daedalus, v. 109, n. 1, 1980, pp. 121-36

ZUBOFF, Shoshana. **The age of surveillance capitalism. The fight for a human future at the new frontier of power**. New York: Public Affairs, 2019.