Dados sensíveis como fator central na discriminação algorítmica: análise crítica das implicações sociais

Sensitive data as a central factor in algorithmic discrimination: critical analysis of social implications

Recebido em 31.10.2025 | Aprovado em 19.04.2026
DOI: https://doi.org/10.63601/bcesmpu.2026.n67.e-67tc01

Thiago Duarte Mesquita

http://lattes.cnpq.br/1812833805742004

Mestre em Direito pela Universidade Católica de Brasília (UCB). Especialista em Governança da Tecnologia da Informação pela Escola Superior do Ministério Público da União (ESMPU). Servidor do Ministério Público do Distrito Federal e Territórios (MPDFT). Assessor Jurídico na Unidade Especial de Proteção de Dados Pessoais no Conselho Nacional do Ministério Público (CNMP).

Resumo: O presente artigo investiga o papel dos dados sensíveis como vetores centrais de discriminação em sistemas de inteligência artificial. O problema de pesquisa consiste em verificar em que medida a discriminação algorítmica decorre do tratamento de dados sensíveis, sejam eles coletados diretamente, inferidos, sejam correlacionados, e quais parâmetros normativos, especialmente na Constituição Federal (CF/1988), na Lei Geral de Proteção de Dados (LGPD) e no Regulamento Geral de Proteção de Dados da União Europeia (GDPR), podem orientar a prevenção e a contenção desses efeitos. A metodologia adotada é jurídico-dogmática, baseada em pesquisa bibliográfica, documental e em análise comparada entre a LGPD e o GDPR. Demonstra-se que a aparente neutralidade dos algoritmos é insuficiente para afastar riscos discriminatórios, pois a discriminação indireta ocorre frequentemente por meio de variáveis proxy, de dados inferidos e por ciclos de retroalimentação (feedback loops). Conclui-se que a proteção de dados

deve transitar de uma visão estática baseada em categorias formais para uma análise funcional de risco, exigindo transparência, explicabilidade e auditoria de sistemas automatizados, com implicações diretas para a atuação do Ministério Público na defesa de direitos fundamentais.

Palavras-chave: discriminação algorítmica; dados sensíveis; LGPD; não discriminação; decisões automatizadas.

Abstract: *This article investigates the role of sensitive data as central vectors of discrimination in artificial intelligence systems. The research problem consists in verifying to what extent algorithmic discrimination stems from the processing of sensitive data, whether directly collected, inferred, or correlated, and which normative parameters, especially in the Brazilian Federal Constitution, the LGPD, and the GDPR, can guide the prevention and containment of these effects. The methodology adopted is legal-dogmatic, based on bibliographic and documentary research and comparative analysis between the LGPD and the GDPR. It is demonstrated that the apparent neutrality of algorithms is insufficient to eliminate discriminatory risks, as indirect discrimination frequently occurs through proxy variables, inferred data, and feedback loops. It is concluded that data protection must transition from a static view based on formal categories to a functional risk analysis, requiring transparency, explainability, and auditing of automated systems, with direct implications for the Public Prosecutor's Office in the defense of fundamental rights.*

Keywords: *algorithmic discrimination; sensitive data; LGPD; non-discrimination; automated decision-making.*

1 Introdução

O avanço tecnológico e a crescente utilização de algoritmos em processos decisórios transformaram profundamente a dinâmica social e institucional, ampliando o uso de sistemas automatizados em setores como crédito, saúde, segurança pública, recrutamento e justiça. Embora frequentemente apresentados como instrumentos de eficiência e de objetividade, tais sistemas podem reproduzir e intensificar preconceitos históricos, sobretudo quando operam com dados sensíveis ou com variáveis capazes de inferir atributos protegidos. A relevância do tema é ainda maior, porque esses processos decisórios incidem sobre esferas diretamente vinculadas à dignidade da pessoa humana, à igualdade e à proteção contra discriminações indevidas.

Nesse contexto, o presente artigo delimita seu objeto na relação entre discriminação algorítmica e tratamento de dados pessoais sensíveis, com especial atenção aos riscos jurídicos decorrentes do uso direto, inferido ou correlacionado dessas informações em decisões automatizadas. O problema de pesquisa pode ser formulado nos seguintes termos: em que medida a discriminação algorítmica decorre do tratamento de dados sensíveis e quais instrumentos normativos existentes, notadamente a Constituição Federal (CF/1988), a Lei Geral de Proteção de Dados (LGPD) e o *General Data Protection Regulation* ou Regulamento Geral sobre a Proteção de Dados (GDPR – União Europeia) são aptos a prevenir, a limitar e a controlar tais práticas? Parte-se da hipótese de que a discriminação algorítmica não se restringe ao tratamento explícito de categorias sensíveis, mas pode surgir também por inferências estatísticas, variáveis *proxy* e ciclos de retroalimentação, o que exige leitura materialmente protetiva do princípio da não discriminação e do regime jurídico das decisões automatizadas.

O objetivo geral consiste em analisar criticamente a centralidade dos dados sensíveis na discriminação algorítmica e examinar a suficiência dos parâmetros jurídicos disponíveis para a tutela dos direitos fundamentais nesse contexto. Como objetivos específicos, o artigo busca demonstrar a insuficiência da ideia de neutralidade algorítmica; identificar os mecanismos técnicos e sociais que favorecem a produção de discriminações diretas e indiretas; comparar a disciplina da LGPD com elementos relevantes do GDPR; e apontar medidas regulatórias e éticas voltadas à mitigação de riscos e à promoção de maior equidade no desenvolvimento e na aplicação dessas tecnologias.

A justificativa da pesquisa reside no fato de que a crescente delegação de escolhas socialmente relevantes a sistemas automatizados amplia os riscos de invisibilização de vieses, dificultando a contestação de decisões injustas e a responsabilização dos agentes envolvidos. Além disso, o uso de algoritmos em contextos sensíveis pode reforçar desigualdades estruturais, o que torna indispensável uma abordagem jurídica capaz de articular proteção de dados, igualdade material, transparência, explicabilidade e *accountability*.

Metodologicamente, adota-se abordagem jurídico-dogmática, com pesquisa bibliográfica e documental de caráter exploratório e analítico. A revisão da literatura partiu de obras de referência nacional (Doneda,

Bioni, Konder, Frazão) e estrangeira (Solove, Pasquale, Barocas, Citron, Browne, Winner, Feenberg, Srnicek, Kitchin, Mann, Matzner, Noble, Zarsky, O'Neil, Ajunwa) sobre discriminação algorítmica, proteção de dados e teoria crítica da tecnologia, permitindo construir o marco teórico sobre a insuficiência da neutralidade algorítmica, o determinismo tecnológico e econômico, o tratamento de dados sensíveis sob uma perspectiva funcional e os mecanismos discriminatórios (variáveis *proxy*, dados inferidos, feedback loops).

A análise normativa concentrou-se na comparação entre LGPD (arts. 5º, II; 6º, VI e IX; 11, II, g; e 20) e GDPR (art. 5º, 9º e 22, além dos Considerandos 51 e 71), em que foram identificadas convergências e divergências no regime de dados sensíveis e decisões automatizadas, com exame complementar da legislação antidiscriminatória norte-americana (*Equal Credit Opportunity Act*, *Fair Housing Act*) e de orientações regulatórias recentes (Consumer Financial Protection Circular 2022-03) sobre aplicação da *disparate impact doctrine* a algoritmos.

Os casos paradigmáticos foram selecionados por comprovação empírica, diversidade setorial (recrutamento, crédito, saúde, justiça criminal, segurança pública) e capacidade de ilustrar diferentes formas de discriminação: Amazon (viés de gênero em recrutamento), *Apple Card* (discriminação de gênero em concessão de crédito), sistema Compas (viés racial em justiça criminal), algoritmo de saúde dos Estados Unidos (subtratamento de pacientes negros) e reconhecimento facial no Rio de Janeiro (erro identificatório contra pessoa negra – caso Daiane dos Santos). Adicionalmente, analisou-se o uso de software de policiamento preditivo (*PredPol*) como exemplo de ciclo de retroalimentação discriminatória.

A articulação entre teoria, norma e casos buscou identificar riscos jurídicos e mecanismos de contenção disponíveis no ordenamento brasileiro, com especial atenção para os instrumentos de transparência, explicabilidade, auditoria e *accountability* previstos na LGPD e em projetos regulatórios em tramitação (PL n. 2.338/2023).

De modo geral, este artigo contempla a introdução, desenvolve-se em duas seções principais e nas considerações finais. A primeira seção examina a expansão dos algoritmos na tomada de decisão, a crítica à sua pretensa neutralidade e a necessidade de substituição desse paradigma

por padrões de transparência, contestabilidade e responsabilidade. A segunda analisa a centralidade dos dados sensíveis na discriminação algorítmica, com destaque para dados inferidos, variáveis *proxy*, feedback loops, casos concretos e os instrumentos normativos de tutela previstos na LGPD e no GDPR.

2 Algoritmos e sua expansão na tomada de decisão

Atualmente, os algoritmos estão profundamente integrados em praticamente todas as esferas da vida cotidiana, desde as decisões de crédito e processos de recrutamento até a definição de políticas públicas e o fornecimento de serviços de saúde. Seu uso massivo é impulsionado pela capacidade de processar grandes volumes de dados com rapidez e uma suposta precisão, oferecendo soluções aparentemente objetivas para problemas complexos. No entanto, essa confiança crescente nos algoritmos também levanta preocupações sobre a falta de transparência, sobre o potencial para a perpetuação de preconceitos e sobre a possível erosão de direitos fundamentais. À medida que os algoritmos se tornam onipresentes, é preciso avaliar criticamente como são projetados e implementados, especialmente quando envolvem dados sensíveis, para garantir que não se tornem ferramentas de discriminação e de exclusão.

Paralelamente, o aumento da presença de sistemas algorítmicos em diversas esferas da vida humana e social torna a questão do poder cada vez mais central. Em um cenário marcado pela crença de que "a tecnologia fala por si" e seguiria um curso inevitável, corre-se o risco de assumir uma postura de determinismo tecnológico, isto é, de tratar os algoritmos como forças autônomas que, por si sós, definem rumos sociais e institucionais. Na prática, porém, esse determinismo tecnológico frequentemente funciona como fachada para um determinismo econômico, pois são interesses de mercado, estratégias empresariais e assimetrias de poder informacional que orientam quais tecnologias serão desenvolvidas, que dados serão explorados e quais grupos serão mais vigiados ou segmentados.

Como sustenta Winner (1986), as tecnologias não são meras ferramentas neutras, mas possuem propriedades políticas inerentes, sendo capazes de incorporar formas específicas de poder e autoridade em

sua própria estrutura física ou lógica. Um exemplo trazido pelo autor, à época, foram as pontes de Long Island, projetadas com altura baixa propositalmente para impedir a passagem de ônibus e, consequentemente, barrar o acesso de minorias e pessoas pobres (que dependiam de transporte público) às praias da região.

Ademais, a compreensão da inteligência artificial como fenômeno sociotécnico afasta a ideia de que vieses discriminatórios decorrem exclusivamente de falhas acidentais nos bancos de dados. Em muitos casos, o problema encontra-se na própria definição das métricas de desempenho e das funções objetivo utilizadas pelos modelos algorítmicos. Todo sistema de aprendizado de máquina opera mediante processos de otimização orientados por critérios previamente definidos: maximizar lucro, reduzir inadimplência, ampliar engajamento ou prever comportamentos futuros. Entretanto, a seleção desses parâmetros não constitui decisão meramente técnica, mas escolha normativa que privilegia determinados valores em detrimento de outros. Quando plataformas de crédito adotam exclusivamente índices de rentabilidade e risco financeiro como métricas de sucesso, relegando critérios de equidade racial ou de gênero a posição secundária, não estão reproduzindo neutralidade matemática, mas prioridades econômicas específicas.

Essa perspectiva aproxima-se da teoria crítica da tecnologia desenvolvida por Feenberg (2002), para quem o desenho técnico jamais é neutro, refletindo disputas sociais sobre quais interesses serão incorporados ao funcionamento dos sistemas. Através do conceito de "código técnico", o autor demonstra como os interesses econômicos e as hegemonias políticas são traduzidos em requisitos técnicos, moldando o funcionamento dos sistemas para reproduzir as assimetrias de poder já existentes na sociedade.

Essa perspectiva se aprofunda quando se observa que a relação entre tecnologia e racionalidade econômica torna-se ainda mais evidente no contexto do chamado capitalismo de plataforma. Para Srnicek (2017) – cujo contexto se insere na crise imobiliária de 2008, nos Estados Unidos –, plataformas digitais estruturam-se a partir da captura massiva de dados e da capacidade de convertê-los em vantagem competitiva e em previsão comportamental. Nesse cenário, a inteligência artificial deixa de ser apenas ferramenta técnica e passa a integrar a própria lógica de acumulação econômica. A coleta, a classificação e a análise automatizada

de dados pessoais tornam-se mecanismos centrais de extração de valor. Isso reforça incentivos para sistemas de decisão opacos, escaláveis e orientados prioritariamente à eficiência econômica. A consequência é a consolidação de modelos algorítmicos que tendem a privilegiar maximização de desempenho mercadológico, ainda que isso produza impactos discriminatórios ou aprofundamento de desigualdades sociais.

O design algorítmico constitui, portanto, espaço de exercício de poder. A escolha das variáveis relevantes, dos dados considerados legítimos e dos critérios utilizados para aferição de desempenho expressa prioridades institucionais e econômicas que influenciam diretamente a distribuição de oportunidades e de riscos na sociedade digital. Sob tal perspectiva, cientistas de dados, engenheiros e organizações empresariais não podem ser tratados como agentes puramente imparciais, pois participam ativamente da construção normativa dos sistemas automatizados.

Plataformas de redes sociais desenvolvem algoritmos de recomendação não porque a tecnologia "evolui naturalmente", mas porque modelos de negócio baseados em publicidade direcionada exigem maximização do tempo de permanência e refinamento contínuo de perfis comportamentais, convertendo experiências humanas em dados para predição de comportamento. Similarmente, sistemas automatizados de *credit scoring* incorporam variáveis comportamentais e geolocalização não por imperativo técnico, mas porque a segmentação granular de risco permite precificação diferenciada e expansão de mercado, mesmo que isso implique discriminação indireta por meio de variáveis *proxy* (Frazão, 2021a).

Nesse sentido, não é adequado tratar a criação de algoritmos como mera questão de engenharia neutra, tal como muitas vezes sugerem os próprios programadores, como se se limitassem a resolver problemas técnicos previamente dados. Frazão (2021c), citando Ben Green, argumenta que os cientistas de dados devem reconhecer-se como atores políticos envolvidos em construções normativas da sociedade, motivo pelo qual precisam analisar e avaliar seu trabalho com base no impacto que ele terá na vida das pessoas.

Acrescenta que não se pode aceitar a ideia de que a programação de algoritmos seja simplesmente uma questão técnica de engenharia, haja vista os numerosos exemplos de sistemas que moldam comportamentos, redistribuem poder e influenciam diferentes aspectos da sociedade. A

falta de entendimento sobre o verdadeiro alcance do trabalho dos cientistas de dados impede que esses profissionais compreendam e assumam as responsabilidades éticas e sociais que acompanham suas atividades.

Em relação a essa suposta neutralidade, a autora argumenta que ela é um objetivo inatingível, uma vez que é impossível participar da ciência e da política sem ser influenciado por contextos, valores e interesses pessoais. Além disso, o esforço na sua busca não representa uma postura politicamente imparcial, mas sim uma posição essencialmente conservadora, que acaba por defender o *status quo*.

Assim, fica claro que a programação de algoritmos e a busca pela neutralidade na ciência e na política são questões muito mais complexas e influenciadas por fatores subjetivos do que se costuma pensar. Tratar a programação de algoritmos apenas como uma questão técnica de engenharia ignora o profundo impacto que esses sistemas exercem na formação de comportamentos, na redistribuição de poder e na configuração das dinâmicas sociais.

Nessa mesma linha intelectiva, Kitchin (2017) critica fortemente a noção de neutralidade dos algoritmos. O autor sustenta que eles não são desenvolvimentos imparciais ou livres de influências políticas. Longe de serem neutros, os algoritmos são criados para gerar valor econômico, direcionar comportamentos e moldar preferências de maneiras específicas. Além disso, eles desempenham funções como identificar, selecionar e categorizar indivíduos.

Prossegue o autor estabelecendo que os algoritmos realizam uma série de tarefas, incluindo a pesquisa, o agrupamento, a classificação, a categorização, a combinação, a análise, a criação de perfis, a modelagem, a simulação, a visualização e a regulação de pessoas, processos e lugares. Ao desempenharem essas funções por meio de software, os algoritmos "moldam nossa percepção do mundo e exercem um impacto profundo na realidade".

É fundamental reconhecer, portanto, que algoritmos não são neutros: eles podem incorporar vieses que refletem a cultura interna da organização, bem como os preconceitos, as escolhas e as omissões de quem os desenvolve. No contexto corporativo, em especial quando prevalecem incentivos voltados à maximização de lucros no curto prazo, tende a predominar uma lógica de "lançar primeiro e corrigir depois", que desloca os riscos e

os custos de eventuais erros para os usuários e para a coletividade. Uma das manifestações mais significativas desse fenômeno é o uso de perfis algorítmicos como instrumento de classificação e de segmentação social.

Nesse sentido, Mann e Matzner (2019) mostram que um dos instrumentos mais utilizados na contemporaneidade para discriminar pessoas se chama "perfil algorítmico", e por meio dele é feita uma classificação social. Eles demonstram que esses perfis tendem a incidir de forma desproporcional sobre grupos já marginalizados, como minorias raciais, pessoas em situação de vulnerabilidade socioeconômica e mulheres.

Por sua vez, Brown (2015) argumenta que o perfil algorítmico contribui para perpetuar hierarquias baseadas em um emaranhado de marcadores de identidade, reforçando posições de privilégio e de subordinação. As práticas discriminatórias tornam-se, em certa medida, autoaplicáveis: ciclos de feedback são alimentados por bases de dados que concentram, de forma desproporcional, informações sobre determinados grupos, o que resulta em supermonitoramento e superpoliciamento dessas populações.

A crítica à neutralidade algorítmica também encontra respaldo nos estudos de Noble (2018), que demonstra como sistemas computacionais frequentemente reproduzem estruturas históricas de exclusão racial e social sob a aparência de objetividade matemática. Segundo a autora, os vieses algorítmicos não constituem desvios excepcionais do sistema, mas manifestações das próprias hierarquias existentes na sociedade e incorporadas aos processos tecnológicos. Isso evidencia que a governança da inteligência artificial não pode restringir-se à correção técnica de *datasets* ou ao aumento da precisão estatística dos modelos. O debate deve alcançar as próprias finalidades econômicas e institucionais que orientam o desenvolvimento tecnológico.

Assim, ao se reconhecer que a programação de algoritmos não é neutra, o risco se torna ainda mais grave quando esses sistemas lidam com dados sensíveis, justamente por estarem ligados a dimensões marcadas por camadas adicionais de vulnerabilidade, como raça, etnia, gênero e religião. Quando manipulados de maneira enviesada, esses dados tendem a intensificar a discriminação, o que reforça desigualdades preexistentes e perpetua preconceitos. Desse modo, a falta de neutralidade na programação de algoritmos coloca em risco direitos fundamentais de indivíduos e grupos historicamente marginalizados.

É nesse ponto que emerge, de maneira mais nítida, a discriminação algorítmica: decisões automatizadas passam a tratar indivíduos e grupos de maneira desigual ao absorver e reproduzir vieses históricos presentes nos dados, replicando e ampliando distorções preexistentes e contribuindo para a manutenção de preconceitos arraigados na sociedade (Barocas; Hardt; Narayanan, 2019).

Essa discussão possui implicações diretas para a proteção de direitos fundamentais e para a interpretação da LGPD. A identificação de que sistemas algorítmicos incorporam escolhas políticas e econômicas reforça a necessidade de mecanismos robustos de transparência, prestação de contas e avaliação de impacto algorítmico. Não se trata apenas de exigir explicabilidade técnica das decisões automatizadas, mas de permitir o escrutínio das racionalidades econômicas subjacentes aos modelos utilizados. A governança democrática da inteligência artificial pressupõe, portanto, reconhecer que eficiência estatística e maximização econômica não podem constituir os únicos parâmetros legítimos para sistemas capazes de afetar direitos, oportunidades e liberdades individuais.

Em face desses riscos, ganha relevo a noção de *accountability* algorítmica, que articula transparência, explicabilidade, auditabilidade e responsabilização dos agentes envolvidos no desenvolvimento e no uso de sistemas automatizados. A mera referência genérica à transparência revela-se insuficiente quando o algoritmo produz efeitos relevantes sobre direitos fundamentais, sendo necessário exigir documentação adequada dos modelos, registro dos dados empregados, indicação dos critérios de decisão e possibilidade efetiva de revisão humana. Isso pressupõe a existência de trilhas de auditoria, mecanismos internos e externos de controle e procedimentos institucionais capazes de identificar e corrigir vieses, inclusive quando decorrentes de dados sensíveis, dados inferidos ou variáveis *proxy*, de modo a evitar que a delegação de escolhas a sistemas algorítmicos resulte em esvaziamento da responsabilidade jurídica.

Nesse contexto, a relação entre a pretensa neutralidade da programação e o uso de dados sensíveis torna-se particularmente crítica. A retórica da neutralidade tende a criar a ilusão de que os algoritmos seriam instrumentos imparciais e objetivos, quando, na realidade, são profundamente influenciados por vieses que refletem contextos culturais, sociais e históricos, bem como interesses econômicos e institucionais.

Quando dados sensíveis são processados por sistemas apresentados como "neutros", o risco de discriminação se potencializa, pois decisões automatizadas tendem a reproduzir desigualdades preexistentes sob a aparência de mera racionalidade técnica.

Isso posto, a ideia de neutralidade na programação precisa ser constantemente questionada, especialmente em relação ao tratamento de dados sensíveis, para garantir que esses algoritmos sejam projetados e implementados de maneira justa e equitativa.

Por essa razão, em vez de se apoiar na noção vaga de neutralidade algorítmica, o debate jurídico e regulatório deve partir de um padrão ético-jurídico afirmativo, orientado pelos princípios da igualdade material, da não discriminação e da centralidade da dignidade da pessoa humana. Isso significa exigir que sistemas de decisão automatizada sejam concebidos com requisitos mínimos de explicabilidade, contestabilidade e prestação de contas, de modo que as pessoas afetadas possam compreender, questionar e, se necessário, obter a revisão das decisões que as atingem. Sob essa perspectiva, o desenvolvimento e o uso de algoritmos que operam com dados sensíveis ou com variáveis aptas a revelá-los não podem ser vistos como exercícios tecnicamente neutros, mas como práticas reguladas de distribuição de oportunidades e riscos, sujeitas ao controle jurídico e institucional.

3 Dados sensíveis e sua relevância na discriminação algorítmica

A tutela jurídica de dados pessoais como um corolário do direito à privacidade nos leva a considerar que a autodeterminação informativa, ou o poder de controle sobre os próprios dados, deve ser a tônica quando buscamos a proteção específica dos dados sensíveis, especialmente se tais dados podem gerar tratamentos desiguais. O reconhecimento do direito fundamental à igualdade no art. 5º, *caput*, da Constituição Federal tutela também o direito ao tratamento sem distinções de qualquer natureza. Ao mesmo tempo, entre os objetivos fundamentais da República Federativa do Brasil, constantes do art. 3º da CF, está o de "promover o bem de todos, sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação". Soma-se ao reconhecimento constitucional da proteção

da igualdade e da não discriminação a previsão na LGPD da impossibilidade do tratamento para fins discriminatórios ilícitos ou abusivos, isto é, o princípio da não discriminação (art. 6º, IX).

Essa preocupação é ainda mais relevante quando consideramos a definição legislativa de dados sensíveis. Do ponto de vista normativo, dois principais instrumentos legais oferecem diretrizes sobre o que constitui um dado sensível. O primeiro é a Lei Geral de Proteção de Dados – LGPD, criada em 2018 e em vigor desde 2020 (Brasil, 2018). A LGPD define dado sensível como qualquer dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organizações de caráter religioso, filosófico ou político, dados referentes à saúde ou à vida sexual, dados genéticos ou biométricos, quando vinculados a uma pessoa natural (art. 5º, II).

O segundo é o Regulamento Geral de Proteção de Dados da União Europeia – GDPR (União Europeia, 2016). Lançado em 2016 e em vigor desde 2018, o GDPR define dado sensível de forma semelhante à LGPD. No entanto, há uma diferença semântica importante: o normativo europeu emprega o verbo "revelar" em sua redação, de modo que são considerados sensíveis aqueles cujo tratamento possa revelar a origem racial ou étnica, entre outras características protegidas (artigo 9.1).

Ademais, o Considerando 51 do GDPR ressalta que os dados pessoais especialmente sensíveis merecem proteção específica, justamente porque, pela sua natureza e pelo contexto em que são tratados, o processamento pode implicar riscos significativos para os direitos e as liberdades fundamentais das pessoas envolvidas.

O Considerando 71, por sua vez, estabelece que o responsável pelo tratamento deve proteger os dados pessoais levando em conta os potenciais riscos aos direitos e aos interesses dos titulares, de modo a prevenir, entre outros efeitos, resultados discriminatórios contra pessoas em razão de sua origem racial ou étnica, opinião política, religião ou convicções, filiação sindical, estado genético ou de saúde, ou orientação sexual.

Em outra via, para além da divergência terminológica, há uma distinção funcional relevante entre as estruturas das normas. Enquanto o GDPR estabelece, no seu artigo 9.1, uma proibição geral de tratamento de categorias especiais de dados, admitindo exceções estritas, a LGPD

adota uma postura de condicionamento do tratamento. Um ponto de tensão reside na base legal de "prevenção à fraude e à segurança do titular", prevista no art. 11, II, *g*, da lei brasileira. Por ser uma cláusula aberta e frequentemente invocada em sistemas de decisão automatizada (como análise de crédito e biometria facial), essa base legal pode tornar-se um "porto seguro" para o uso extensivo de dados sensíveis. Em contraste, o regime europeu exige que tais exceções sejam interpretadas restritivamente e acompanhadas de salvaguardas específicas para os direitos fundamentais, impondo um ônus de transparência e proporcionalidade mais gravoso ao controlador.

Nesse contexto, é importante recordar que o dado sensível remete diretamente à esfera da intimidade do indivíduo, o que abrange aspectos como fisionomia, crenças, ideologias e outras características que compõem sua personalidade. Quando tratados de maneira inadequada, especialmente sob justificativas amplas e pouco controladas, esses dados podem expor informações pessoais sem o conhecimento ou consentimento do titular, de maneira a gerar violações de privacidade e potenciais danos à sua dignidade e ao próprio exercício de direitos fundamentais.

Por outro lado, a concepção de dados sensíveis não pode limitar-se à enumeração normativa prevista na LGPD e/ou no GDPR, sob pena de proteção insuficiente. Importa destacar que tais efeitos discriminatórios podem ocorrer mesmo quando atributos sensíveis, como sexo, raça ou etnia, não são processados diretamente, bastando que o sistema se valha de outras variáveis correlacionadas a essas características.

Nesse aspecto, a doutrina contemporânea vem sustentando uma definição material ou funcional de dado sensível, segundo a qual o caráter sensível de uma informação deve ser aferido também pelo contexto de tratamento e pelo potencial discriminatório que ela pode gerar, independentemente de estar expressamente listada no rol legal.

Solove (2023) argumenta que a regulação baseada exclusivamente em categorias formais de dados sensíveis é inadequada, pois tais categorias são arbitrárias, possuem fronteiras vagas e frequentemente não capturam as situações de maior risco. O autor propõe que "dados são o que dados fazem" (*data is what data does*), isto é, que a proteção jurídica deve centrar-se no dano e no risco concretos que o tratamento pode produzir, e não apenas na natureza abstrata da informação.

Sob essa perspectiva, mesmo dados aparentemente não sensíveis podem assumir natureza sensível quando seu tratamento, isoladamente ou em combinação com outras variáveis, revelar atributos protegidos ou produzir efeitos discriminatórios. Nesse caso, a problemática ganha contornos ainda mais críticos quando se observa que dados sensíveis podem ser inferidos a partir de dados triviais ou aparentemente neutros. Como o dado sensível inferido não foi fornecido diretamente pelo titular, ocorre potencial violação da autodeterminação informativa, uma vez que o tratamento de categorias especiais de dados ocorre sem que o indivíduo tenha plena ciência ou controle sobre essa nova informação gerada pelo sistema. Nessa hipótese, a proteção jurídica deve ser funcional: a inferência de um atributo protegido – como orientação sexual, condição de saúde ou origem racial – deve atrair imediatamente o regime protetivo rigoroso do art. 11 da LGPD, independentemente de a fonte original ser um dado comum, sob pena de esvaziamento das salvaguardas legais.

Desse modo, a tutela jurídica deve ser interpretada de maneira ampla, abrangendo toda e qualquer informação cujo tratamento possa acarretar discriminação, independentemente de sua classificação formal, reconhecendo-se que o risco discriminatório é o critério material central para justificar regimes especiais de proteção.

Pode-se afirmar, em consequência, que a coleta e o tratamento de dados – sejam formalmente classificados como sensíveis ou não – desempenham papel central no processo de discriminação algorítmica quando fornecem informações aptas a segmentar, a perfilar e a excluir indivíduos com base em características pessoais ou escolhas protegidas.

Como decorrência da concepção funcional e orientada ao risco proposta por Solove (2023), o foco da tutela jurídica deve recair sobre o uso concreto dos dados e seus efeitos discriminatórios, reconhecendo-se que qualquer dado pode tornar-se sensível quando empregado para gerar distinções injustas ou arbitrárias. Isso ocorre porque a programação de muitos algoritmos envolve escolhas deliberadas quanto à coleta e ao uso de informações sensíveis ou correlacionadas a atributos protegidos, com o propósito explícito de influenciar comportamentos, segmentar públicos ou restringir, de forma direta ou indireta, o exercício de direitos fundamentais.

Quando algoritmos processam dados como raça, orientação sexual ou convicções religiosas – ou variáveis que funcionem como *proxy* para essas categorias –, não apenas capturam aspectos íntimos da identidade, mas podem empregá-los de maneira a perpetuar preconceitos, reforçar desigualdades estruturais e converter características pessoais protegidas em critérios de exclusão, de desvantagem ou de subordinação (O'Neil, 2016).

Por essa razão, os dados sensíveis são muito mais suscetíveis a serem utilizados para gerar discriminação do que outros tipos de dados. Ao se observar o conteúdo definido pelo legislador, verifica-se que esses elementos, historicamente, têm sido fatores que contribuíram para a segregação social. Konder (2020, p. 455) corrobora essa visão ao afirmar:

> Como destacado, algo une todos esses exemplos: os dados sensíveis são dados pessoais especialmente suscetíveis de utilização para fins discriminatórios, como estigmatização, exclusão ou segregação, de modo que seu tratamento atinja a dignidade de seu titular, lesionando sua identidade pessoal ou privacidade. O próprio anteprojeto da legislação identifica que o fim precípuo do tratamento diferenciado dos dados sensíveis é impedir a discriminação da pessoa humana com base nas suas informações. Por essa razão somente podem ser sensíveis os dados referentes à pessoa humana, em virtude do valor intrínseco da sua dignidade.

Em linha adicional, Bioni (2019, p. 112) caracteriza dados sensíveis como propensos a abusos, pois refletem identidades e escolhas pessoais que podem ser alvo de preconceito e discriminação. Afirma esse autor: "[o]s dados sensíveis, por sua própria natureza, têm um potencial maior de causar danos ao titular, pois tocam aspectos fundamentais de sua identidade e podem ser utilizados para fins discriminatórios".

Prossegue o autor, demonstrando que os dados sensíveis fazem parte de uma tipologia diferente, justamente porque o conteúdo desses dados oferece uma possibilidade de vulnerabilidade baseada em um viés discriminatório. Dados que expressam informações de natureza sexual, religiosa, racial e política necessitam de proteção específica, pois refletem elementos fundamentais da personalidade de uma pessoa e são suscetíveis a gerar distinções sociais negativas.

Convém registrar, também, que algoritmos treinados com dados históricos marcados por vieses tendem a replicar esses padrões, produzindo

discriminações diretas ou indiretas. Mesmo quando dados sensíveis não são utilizados de forma explícita, eles podem estar fortemente correlacionados a outras variáveis consideradas pelo modelo, o que acaba por gerar resultados discriminatórios sob a aparência de neutralidade (O'Neil, 2016).

Com base nisso, pode-se afirmar que os vieses presentes nos dados constituem um dos principais problemas no uso de algoritmos, pois esses sistemas de inteligência artificial aprendem a partir de bases históricas que frequentemente carregam preconceitos e desigualdades estruturais. Conforme apontam Barocas, Hardt e Narayanan (2019), quando alimentados com informações enviesadas, os algoritmos tendem a reproduzir e até amplificar tais distorções, perpetuando discriminações existentes na sociedade.

Além disso, falhas no design do algoritmo podem agravar esse cenário. Zarsky (2016) destaca que a escolha das variáveis utilizadas no desenvolvimento dos modelos algorítmicos, ainda que de forma não intencional, pode introduzir vieses que comprometem a imparcialidade dos resultados. Muitas vezes, variáveis correlacionadas a fatores sensíveis – como raça, gênero ou condição socioeconômica – são incorporadas indiretamente, gerando decisões discriminatórias de forma velada.

Nesse sentido, a discriminação por meio de variáveis *proxy* representa um dos desafios mais complexos para a dogmática jurídica, porquanto permite que o algoritmo segregue grupos protegidos sem nunca processar diretamente um dado sensível (Wachter; Mittelstadt, 2019). Na prática, o sistema utiliza variáveis aparentemente neutras que guardam uma correlação estatística tão alta com atributos sensíveis que acabam por substituí-los. Um exemplo pedagógico recorrente é a utilização do CEP ou código postal em modelos de *credit scoring* ou precificação de seguros.

Nos Estados Unidos, em cidades marcadas por histórico de *redlining* (negativa sistemática de crédito e serviços financeiros em razão do local de residência) e de segregação espacial, a localização geográfica passa a funcionar, na prática, como substituto imediato da origem racial ou do nível socioeconômico (EUA, 2023). Assim, ao atribuir pontuações mais baixas ou prêmios mais elevados a moradores de determinados bairros, o algoritmo, por via transversa, reproduz racismo institucional ou discriminação de classe sob a aparência tecnicamente

neutra de "risco geográfico", típicos de formas de discriminação indireta baseadas em variáveis *proxy*.

Para tentar coibir essas práticas, o modelo norte-americano possui legislação antidiscriminatória específica, como o *Equal Credit Opportunity Act* – ECOA (EUA, 1974) e o *Fair Housing Act* – FHA (EUA, 1968), que proíbem discriminação com base em raça, cor, religião, sexo, estado civil, nacionalidade ou idade, inclusive quando ocorre de forma indireta por meio de variáveis proxy. Recentemente, agências reguladoras americanas (*Consumer Financial Protection Bureau* – CFPB [CFPB, 2022]) têm interpretado essas leis de forma a alcançar algoritmos que produzam efeitos discriminatórios, mesmo sem intenção expressa e sem processar diretamente dados protegidos. Essa abordagem, conhecida como disparate *impact doctrine*, responsabiliza o controlador por resultados discriminatórios estatisticamente comprováveis, exigindo que demonstre que o critério utilizado é justificado por necessidade legítima de negócio e que não há alternativa menos discriminatória disponível (Barocas; Selbst, 2016).

Outro aspecto relevante é a possibilidade de interpretações errôneas decorrentes de dados mal processados ou ambíguos. Citron e Pasquale (2014) observam que erros na coleta, no tratamento ou na análise das informações podem levar a conclusões equivocadas, o que afeta negativamente indivíduos e grupos específicos, sobretudo quando os algoritmos são aplicados em contextos sensíveis, como crédito, saúde ou justiça criminal.

Por fim, Pasquale (2015) chama a atenção para os chamados feedback loops, ou ciclos de retroalimentação, nos quais decisões iniciais enviesadas acabam sendo continuamente reforçadas ao longo do tempo. Isso ocorre porque os resultados dos algoritmos passam a alimentar novos conjuntos de dados, consolidando padrões discriminatórios e tornando cada vez mais difícil a correção das distorções originais.

Um exemplo é o chamado policiamento preditivo, em que dados históricos de abordagens e registros criminais em determinados bairros alimentam sistemas que, por sua vez, direcionam novamente a atuação policial para esses mesmos territórios. Nesse caso, reforçam-se ciclos de vigilância concentrada sobre grupos já vulnerabilizados. Cite-se, nesse sentido, o software de policiamento preditivo PredPol, utilizado em diversas jurisdições nos Estados Unidos (Lum; Isaac, 2016).

O sistema operava a partir de dados históricos de criminalidade para projetar "pontos quentes" de policiamento; contudo, ao processar registros que refletiam práticas de vigilância historicamente concentradas em bairros de minorias raciais, o algoritmo tendia a reforçar, repetidamente, a presença policial nessas mesmas áreas. Esse movimento configura um ciclo vicioso: a maior presença policial em determinados territórios produz, por sua vez, mais registros e detenções, que retornam ao sistema como novos dados de treino, confirmando as previsões algorítmicas e consolidando o estigma sobre populações vulnerabilizadas sob a aparência de mera neutralidade matemática.

Desse modo, compreender esses fatores é essencial para reconhecer as limitações das soluções algorítmicas e a necessidade de mecanismos mais robustos de governança, supervisão e transparência. Somente com estruturas consistentes de controle, auditoria e prestação de contas é possível mitigar riscos e evitar que tais tecnologias aprofundem desigualdades sociais já existentes, em vez de corrigi-las.

Diante disso, tornam-se essenciais a intervenção do Estado e a mobilização da sociedade para mitigar práticas ilegais e abusivas de tratamento de dados, de modo a proteger os direitos dos indivíduos e impedir que dados sensíveis sejam utilizados de forma prejudicial. Nesse sentido, essa atuação em conjunto, de forma contínua, é obrigatória para coibir práticas discriminatórias, de modo a promover maior equidade e assegurar a efetiva proteção dos direitos fundamentais.

Nesse sentido, a União Europeia deu um passo além do GDPR ao aprovar, em 2024, o Regulamento de Inteligência Artificial – AI Act (União Europeia, 2024), que estabelece regime específico para sistemas de IA de alto risco, incluindo aqueles utilizados em recrutamento, concessão de crédito, acesso a serviços essenciais e aplicação da lei. O AI Act exige, para esses sistemas, avaliação prévia de conformidade, documentação técnica detalhada, supervisão humana efetiva, transparência sobre funcionamento e uso de dados, e mecanismos de monitoramento contínuo para detecção de vieses discriminatórios.

No Brasil, esse debate regulatório avança com o Projeto de Lei (PL) n. 2.338/2023 (Brasil, 2023), que estabelece diretrizes para o desenvolvimento e o uso de sistemas de inteligência artificial, com foco em sistemas de alto risco. O PL prevê requisitos de transparência, de gestão de riscos,

de supervisão humana e de auditoria externa para sistemas que possam afetar direitos fundamentais, incluindo decisões em áreas como crédito, recrutamento, saúde e segurança pública. Embora ainda em tramitação, o projeto reforça a tendência de que a regulação de algoritmos no Brasil deva convergir para padrões de explicabilidade, *accountability* e avaliação de impacto discriminatório, alinhando-se ao regime de proteção de dados da LGPD e às diretrizes internacionais de governança algorítmica.

Essa preocupação regulatória justifica-se, em grande medida, pela constatação empírica de que sistemas automatizados, quando mal concebidos ou alimentados por dados enviesados, podem reproduzir e até aprofundar desigualdades sociais preexistentes, conforme demonstram casos emblemáticos ocorridos em diferentes setores e jurisdições. Na Amazon, o seu sistema de recrutamento penalizava currículos que continham o termo *women's*, o que caracterizava viés de gênero diretamente relacionado à composição dos dados de treinamento, historicamente dominados por perfis masculinos (Dastin, 2018).

Outra ocorrência amplamente divulgada foi a do Apple Card, em que relatos apontaram que mulheres receberam limites de crédito significativamente mais baixos do que homens com perfis financeiros semelhantes (Vigdor, 2019).

No caso Amazon, a presença do termo *women's* em currículos serviu como marcador indireto de gênero; no caso *Apple Card*, embora não se tenha comprovado o uso direto da variável sexo, o resultado discriminatório sugere que outras variáveis correlacionadas – como histórico de renda, padrões de consumo ou comportamento financeiro historicamente diferenciados entre homens e mulheres – produziram efeito equivalente. Isso reforça a necessidade de adoção de uma concepção funcional ou material de dado sensível.

Em outra situação, na esfera penal, o sistema Compas tornou-se emblemático ao classificar de forma desproporcional réus negros como de alto risco de reincidência em comparação com réus brancos. Essa situação revela a maneira pela qual algoritmos utilizados em decisões judiciais podem refletir e reforçar preconceitos raciais estruturais, colocando em risco princípios fundamentais de justiça e equidade (ProPublica; Angwin *et al*., 2016).

Por meio de um algoritmo de saúde utilizado nos Estados Unidos, pacientes negros foram sistematicamente subtratados, porque o modelo empregava custos de saúde como *proxy* para necessidade de cuidados. Como pacientes negros historicamente receberam menos investimentos em saúde – não por menor necessidade, mas por barreiras estruturais de acesso –, o sistema interpretou erroneamente gastos mais baixos como menor demanda por tratamento (Paul, 2019).

No Brasil, especificamente no Rio de Janeiro, uma mulher foi erroneamente identificada como foragida por um sistema de reconhecimento facial da Polícia Militar (Dutra, 2024). Daiane, psicóloga e coordenadora de Promoção de Igualdade Racial de Nova Iguaçu, tem dez anos de experiência na luta por igualdade racial e no atendimento a vítimas de preconceito.

No dia 30 de abril de 2024, enquanto participava de uma conferência, ela foi abordada por três policiais militares do projeto Segurança Presente. A vítima relatou que os policiais a cercaram e começaram a compará-la com imagens registradas em um aparelho celular.

Durante a abordagem, os militares informaram que ela havia sido identificada pelo sistema de reconhecimento facial e, embora já tivessem percebido que não se tratava da pessoa procurada, exigiram que ela apresentasse um documento de identificação.

Ela então argumentou com os agentes que ela estava sendo vítima de racismo em uma conferência sobre igualdade racial. A psicóloga declarou seus direitos e destacou a frequência com que a polícia confunde pessoas negras com criminosos e suspeitos.

Esse incidente ressalta como sistemas de reconhecimento facial, quando mal utilizados, podem perpetuar preconceitos raciais e contribuir para a discriminação sistêmica. A reação de Daiane, ao argumentar sobre o racismo sofrido e a frequência com que pessoas negras são equivocadamente associadas a criminosos, demonstra a necessidade urgente de revisar e de regulamentar o uso dessas tecnologias, garantindo que não reforcem desigualdades e injustiças sociais.

Vê-se, assim, que o uso de sistemas de reconhecimento facial pela segurança pública suscita questionamentos adicionais sobre a base legal e a proporcionalidade do tratamento de dados biométricos. Não

obstante o art. 11, II, *g*, da LGPD autorize o tratamento de dados sensíveis para prevenção à fraude e à segurança do titular, essa base legal foi concebida para proteção do próprio indivíduo (como autenticação biométrica em bancos), não para vigilância massiva ou identificação em via pública sem consentimento. A aplicação extensiva dessa exceção para fundamentar reconhecimento facial em segurança pública, sem lei específica que estabeleça salvaguardas, prazos de armazenamento, direitos de contestação e mecanismos de auditoria independente, pode configurar tratamento ilícito de dados sensíveis, especialmente quando o sistema apresenta vieses raciais comprovados. Esse debate reforça a necessidade de regulamentação específica que discipline o uso de tecnologias biométricas pelo poder público, sob pena de violação dos princípios da finalidade, da necessidade e da não discriminação.

Além disso, a situação descrita ilustra como a programação enviesada de algoritmos de reconhecimento facial pode intensificar a discriminação, especialmente quando dados sensíveis, como a cor da pele, são utilizados de forma inadequada. Quando sistemas são desenvolvidos sem a devida atenção à diversidade e às peculiaridades dos dados sensíveis, o risco de cometer erros e perpetuar preconceitos históricos aumenta significativamente.

Um dos principais fatores que explica esse viés é o déficit de representatividade nas bases de dados utilizadas para treinar os modelos de inteligência artificial. Estudos demonstram que bases amplamente empregadas no treinamento de algoritmos de reconhecimento facial, como VGGFace2, MS-Celeb-1M e Casia-WebFace, contêm proporções desbalanceadas de imagens, com até 84,5% de rostos brancos ou caucasianos (Silva, 2024).

Essa sub-representação de grupos minoritários – especialmente pessoas negras, asiáticas e indígenas - faz com que os algoritmos apresentem taxas de erro dramaticamente superiores ao processar esses rostos. Pesquisa do MIT Media Lab revelou que sistemas comerciais de reconhecimento facial apresentavam erro de até 34,7% ao identificar mulheres negras, enquanto a margem de erro para homens brancos ficava abaixo de 1% (Buolamwini; Gebru, 2018). Quando aplicados em contextos de segurança pública, esses sistemas perpetuam padrões de discriminação racial, de forma a submeter pessoas negras a riscos desproporcionais de identificação errada, abordagens policiais injustificadas e violações de direitos fundamentais.

Para mitigar esses riscos, a literatura técnica tem proposto ferramentas de transparência e documentação, como *Datasheets for Datasets* (Gebru *et al.*, 2021) e *Model Cards for Model Reporting* (Mitchell *et al.*, 2019). Essas metodologias exigem que desenvolvedores documentem, de forma estruturada e pública, informações sobre composição demográfica dos dados de treinamento, metodologia de coleta, limitações conhecidas, taxas de erro desagregadas por grupo (raça, gênero, idade), contextos de uso recomendados e contraindicados, e medidas de mitigação de vieses implementadas. Embora essas práticas ainda sejam voluntárias e pouco disseminadas no Brasil, elas representam mecanismos concretos de operacionalização dos princípios de transparência e de *accountability* previstos na LGPD, e deveriam ser exigidas por reguladores, contratantes públicos e pela sociedade civil como condição mínima para uso de sistemas automatizados em contextos sensíveis.

Verifica-se, portanto, que a discriminação algorítmica pode perpetuar desigualdades sociais, violar direitos civis e humanos e minar a confiança pública em sistemas automatizados. Nessa linha, a exposição inadequada desses dados pode resultar em diversas formas de estigmatização e de abusos. Assim, a proteção dos dados sensíveis é essencial para evitar discriminação e garantir a privacidade dos indivíduos, além de assegurar que suas escolhas pessoais não prejudiquem sua participação na vida coletiva, incluindo o exercício de direitos fundamentais.

Ademais, proteger essa categoria de dados é fundamental para preservar a confiança entre indivíduos e organizações responsáveis pelo tratamento, assegurando que informações sensíveis não sejam convertidas em instrumentos de exclusão ou de subordinação. Diante desse cenário, torna-se obrigatório examinar quais medidas jurídicas e técnicas podem ser adotadas para limitar a discriminação promovida por algoritmos. Nesse sentido, cabe esclarecer algumas previsões normativas que oferecem parâmetros relevantes para o enfrentamento do problema.

Entre essas previsões, destaca-se, em primeiro lugar, o princípio da não discriminação, introduzido pela LGPD em seu art. 6º, IX. Esse princípio estabelece a impossibilidade de realização do tratamento de dados pessoais para fins discriminatórios ilícitos ou abusivos, descrevendo de forma expressa a vedação à discriminação no âmbito do processamento de informações pessoais.

Com efeito, o princípio da não discriminação foi concebido, em grande medida, para abranger situações ocorridas em ambiente digital, funcionando como parâmetro central na avaliação da licitude de tratamentos automatizados de dados pessoais e como instrumento de controle material sobre escolhas tecnológicas que possam gerar ou perpetuar exclusões injustas.

Nesse sentido, Doneda assevera que a LGPD, ao incorporar o princípio da não discriminação, visa garantir que os dados pessoais não sejam utilizados de maneira a resultar em discriminação injusta contra indivíduos. O autor salienta que a discriminação pode ocorrer não apenas em razão de preconceitos explícitos, mas também devido a vieses implícitos presentes nos algoritmos que processam esses dados, evidenciando a importância do princípio como instrumento de controle tanto de práticas abertas quanto de formas dissimuladas de tratamento desigual (Doneda, 2020).

Acrescenta, ainda, que os dados sensíveis, por sua natureza, requerem uma proteção especial para evitar a discriminação. Ele observa que a LGPD classifica dados sobre origem racial ou étnica, convicção religiosa, opinião política, dados genéticos, dados biométricos e outros como sensíveis, reconhecendo o potencial desses dados para serem usados de forma discriminatória.

Outra previsão normativa central no combate à discriminação algorítmica é o princípio da transparência no tratamento de dados pessoais. A LGPD o consagra em seu art. 6º, VI, garantindo aos titulares informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e sobre os respectivos agentes envolvidos, com ressalva aos segredos comercial e industrial. De igual modo, encontra previsão no artigo 5.1, alínea *a*, do GDPR.

Sobre esse critério, Frazão (2021b) argumenta que a opacidade constitui um dos maiores desafios no campo da inteligência artificial e do aprendizado de máquina, sobretudo quando esses sistemas são empregados para decisões que afetam de modo significativo a vida das pessoas. Segundo a autora, a falta de visibilidade sobre o funcionamento interno dos algoritmos pode resultar em decisões injustas e discriminatórias, uma vez que os processos decisórios permanecem inacessíveis e incompreensíveis aos indivíduos atingidos. Tal opacidade cria um ambiente institucional

no qual se torna extremamente difícil questionar, auditar ou corrigir decisões algorítmicas, o que favorece a perpetuação de injustiças e a naturalização de vieses discriminatórios embutidos nos sistemas.

Nessa perspectiva, a transparência garantida pelos normativos acima deve ser interpretada de forma expansiva no contexto da inteligência artificial, superando a mera publicidade formal para alcançar o dever de explicabilidade. A opacidade dos algoritmos de aprendizado de máquina, frequentemente protegida pelo argumento de segredo comercial ou industrial, não pode servir de escudo para decisões arbitrárias ou discriminatórias. É imperativo que as organizações não apenas divulguem os dados utilizados, mas assegurem que a lógica decisória seja compreensível e auditável pelos afetados, permitindo que o titular identifique se a decisão automatizada foi pautada em variáveis *proxy*, em dados inferidos ou em correlações estatísticas injustas que reproduzem preconceitos históricos sob a aparência de neutralidade técnica.

Nessa linha, as organizações devem implementar auditorias regulares de seus algoritmos para identificar a presença de vieses e de padrões discriminatórios. Tais auditorias devem ser conduzidas por terceiros independentes, de modo a assegurar imparcialidade técnica e rigor metodológico. Essa exigência encontra respaldo normativo no art. 20 da LGPD, assim como no artigo 22.1 do GDPR, que estabelecem o direito do titular de solicitar a revisão de decisões tomadas exclusivamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluindo decisões destinadas a definir seu perfil pessoal, profissional, de consumo, de crédito ou aspectos de sua personalidade.

Tal direito de revisão constitui garantia essencial à proteção dos indivíduos em contextos nos quais decisões automatizadas podem produzir impactos significativos em suas vidas. Ele assegura, ademais, que os titulares dos dados possam contestar e buscar correção em situações nas quais o processamento automatizado resulte em discriminação ou em prejuízo indevido, reforçando o princípio da não discriminação e viabilizando o controle material sobre escolhas algorítmicas.

A eficácia dessa garantia, contudo, depende de uma supervisão humana que seja qualitativa e substancial, não meramente burocrática ou protocolar. Deve-se atentar para o risco do chamado viés de automação

(*automation bias*), fenômeno em que o revisor humano tende a validar acriticamente o resultado gerado pela máquina por excesso de confiança na precisão estatística do sistema, convertendo a revisão em mera chancela formal de decisões potencialmente discriminatórias.

Para que a revisão atenda efetivamente ao princípio da não discriminação, o agente humano deve possuir não apenas autonomia decisória, mas também capacidade técnica para compreender e auditar os critérios empregados pelo algoritmo, bem como identificar e corrigir ciclos de retroalimentação (feedback loops) que possam estar produzindo ou perpetuando a estigmatização de grupos vulnerabilizados.

Como adverte Ajunwa (2020), a automação cria um paradoxo: a ferramenta desenhada para eliminar o preconceito humano acaba por institucionalizá-lo sob um manto de neutralidade técnica. Esse cenário é agravado pelo viés de automação, pois os operadores tendem a delegar a autoridade moral e decisória ao sistema, tornando a revisão prevista no art. 20 da LGPD uma etapa meramente formal, desprovida de análise crítica sobre os potenciais danos discriminatórios. Nesse sentido, a efetividade do direito de revisão não depende apenas de sua previsão normativa, mas da criação de condições institucionais e técnicas que permitam ao revisor humano exercer efetivo controle sobre a lógica e os critérios do sistema automatizado.

Além disso, o direito de revisão opera como mecanismo de controle e de *accountability*, induzindo as organizações a adotarem padrões mais elevados de diligência e de responsabilidade no desenvolvimento e na implementação de algoritmos, sobretudo quando envolvem o tratamento de dados sensíveis. Desse modo, a LGPD não apenas resguarda direitos individuais, mas também promove uma transformação estrutural no uso de tecnologias de processamento de dados e orienta para padrões de equidade, de justiça e de respeito à dignidade da pessoa humana.

Essa transformação é especialmente relevante quando se considera que o uso de dados sensíveis em processos automatizados de tomada de decisão constitui aspecto crítico que demanda atenção redobrada. Por estarem intrinsecamente vinculados à identidade e à dignidade das pessoas, tais dados, quando manipulados de forma inadequada, podem exacerbar preconceitos e perpetuar desigualdades historicamente arraigadas.

Essas falhas não configuram meros problemas técnicos ou operacionais; elas revelam programação enviesada e ausência de representatividade nos conjuntos de dados utilizados para treinamento dos algoritmos, expondo a dimensão estrutural da discriminação algorítmica.

Nessa perspectiva, a proteção de dados sensíveis transcende a esfera da privacidade individual; constitui, em maior dimensão, a questão central de equidade e de justiça social. Assegurar que essas categorias de dados sejam tratadas com o mais elevado padrão de responsabilidade é condição necessária tanto para prevenir discriminações quanto para viabilizar um uso ético, inclusivo e socialmente justo da tecnologia.

A efetividade dessas proteções, contudo, exige compromisso permanente com a revisão e o aperfeiçoamento contínuos dos sistemas algorítmicos, de modo a assegurar que sejam projetados e operados em conformidade com os princípios da dignidade humana, da transparência e da igualdade material, sem reproduzir ou amplificar vieses e injustiças preexistentes. A LGPD, ao reconhecer expressamente a vulnerabilidade inerente aos dados sensíveis e ao instituir garantias como o princípio da não discriminação (art. 6º, IX) e o direito de revisão de decisões automatizadas (art. 20), configura-se como instrumento jurídico indispensável na defesa dos direitos fundamentais e na construção de um ecossistema tecnológico orientado pela equidade, pela transparência e pela dignidade da pessoa humana.

4 Considerações finais

A presente investigação permitiu concluir que a discriminação algorítmica não constitui mero erro acidental ou falha técnica isolada, mas frequentemente representa subproduto de escolhas políticas e econômicas dissimuladas sob a aparência de neutralidade técnica. A resposta ao problema inicialmente formulado demonstra que a discriminação algorítmica mantém relação estreita com o tratamento de dados sensíveis – não apenas quando tais dados são coletados diretamente, mas sobretudo quando são inferidos, correlacionados ou reproduzidos por meio de variáveis *proxy* e por ciclos de retroalimentação. Disso decorre a conclusão de que a pretensa neutralidade algorítmica não se sustenta como categoria suficiente para legitimar decisões automatizadas com impacto sobre direitos

fundamentais, uma vez que os sistemas incorporam escolhas humanas, contextos institucionais e assimetrias sociais preexistentes.

Os principais achados do estudo podem ser sintetizados em quatro pontos. Primeiro, algoritmos não são instrumentos neutros, mas estruturas sociotécnicas moldadas por interesses econômicos e institucionais, capazes de reproduzir preconceitos históricos não apenas por vieses acidentais nos dados, mas também porque são orientados por objetivos de maximização de lucro, segmentação de mercado e eficiência operacional que frequentemente entram em conflito com imperativos de equidade e não discriminação. A seleção de dados, métricas, objetivos e critérios de classificação reflete escolhas deliberadas que traduzem assimetrias de poder em requisitos técnicos, convertendo o que aparenta ser racionalidade matemática em instrumento de reprodução de desigualdades estruturais.

Segundo, a centralidade dos dados sensíveis deve ser compreendida em sentido material e funcional, abrangendo não apenas categorias normativas, mas também dados inferidos a partir de informações aparentemente neutras, correlações estatísticas, variáveis *proxy* (como CEP, padrões de consumo ou histórico profissional que funcionam como substitutos de raça, gênero ou classe social) e ciclos de retroalimentação (feedback loops) que consolidam e perpetuam padrões discriminatórios. Essa concepção exige rigor interpretativo que a mera classificação estática não alcança, conforme sustentado por Solove (2023) ao propor que "dados são o que dados fazem" (*data is what data does*).

Terceiro, os riscos discriminatórios se agravam em ambientes marcados por opacidade, por ausência de explicabilidade e de transparência substancial, e por insuficiência de mecanismos de auditoria independente e de contestação efetiva, agravados pelo fenômeno do viés de automação (*automation bias*), em que revisores humanos tendem a validar acriticamente decisões algorítmicas, esvaziando a proteção formal prevista no art. 20 da LGPD.

Quarto, a LGPD oferece instrumentos relevantes de tutela, sobretudo por meio do regime especial de dados sensíveis (art. 11), dos princípios da não discriminação (art. 6º, IX) e da transparência (art. 6º, VI), bem como pelo direito de revisão de decisões automatizadas (art. 20), embora este último apresente limitações ao permitir revisão por outro sistema (§ 2º) e não exigir intervenção humana obrigatória,

diferentemente do regime mais rigoroso do GDPR (artigo 22). A efetividade desses instrumentos depende de interpretação expansiva orientada pela explicabilidade, pela *accountability* algorítmica e pela exigência de transparência substancial – não meramente formal –, somada à necessidade de auditoria externa independente e de mecanismos robustos de avaliação de impacto discriminatório, conforme sinalizam diretrizes internacionais (*AI Act* da União Europeia) e projetos legislativos em tramitação no Brasil (PL n. 2.338/2023).

Do ponto de vista prático, as conclusões indicam a necessidade de fortalecimento da governança algorítmica em ambientes públicos e privados, com implementação do que se pode denominar "*accountability* por design". Para legisladores e operadores do direito, isso implica exigir padrões mais claros de documentação, de explicabilidade e de auditabilidade em sistemas automatizados que afetem direitos fundamentais, bem como desenvolver critérios de controle capazes de identificar discriminações indiretas estruturalmente reproduzidas por variáveis *proxy*, dados inferidos e feedback loops.

Para instituições como o Ministério Público, as conclusões reforçam a importância de uma atuação que transite da mera formalidade para a intervenção técnica, preventiva e repressiva em casos de perfilização opaca, reconhecimento facial, pontuação automatizada e outras formas de tratamento automatizado que possam converter desigualdades históricas em decisões aparentemente técnicas.

Nesse cenário, cabe a tais órgãos de controle e fiscalização exigir a comprovação de justeza e equidade, como com a apresentação de relatórios de impacto à proteção de dados pessoais (RIPD) que contenham testes específicos de viés (*bias testing*) e métricas de equidade desagregadas por grupos protegidos, especialmente em sistemas de alto risco aplicados à segurança pública, justiça criminal e reconhecimento facial.

A transparência exigida deve ser substancial – não apenas nominal –, assegurando que o direito de revisão previsto no art. 20 da LGPD seja exercido por agentes dotados de autonomia decisória e capacidade técnica para identificar, interromper e corrigir o viés de automação, impedindo que a delegação de escolhas a sistemas algorítmicos resulte em esvaziamento da responsabilidade institucional e em legitimação acrítica de práticas discriminatórias.

Em conclusão, este estudo apresenta limitações. Por se tratar de pesquisa predominantemente teórica, bibliográfica e documental, não se desenvolveu investigação empírica própria sobre bases de dados, modelos específicos ou práticas institucionais concretas de auditoria algorítmica. Essa limitação não compromete a validade dos resultados, mas indica a conveniência de pesquisas futuras voltadas à análise empírica de algoritmos utilizados na administração pública brasileira, à avaliação de impacto regulatório da LGPD sobre sistemas de decisão automatizada e ao desenvolvimento de parâmetros técnicos e operacionais para auditoria, supervisão e governança de sistemas automatizados no setor público e no setor privado, com ênfase em mecanismos concretos de *accountability*, de explicabilidade e de equidade.

Referências

AJUNWA, Ifeoma. The paradox of automation as anti-bias intervention. **Cardozo Law Review**, Nova York, v. 41, n. 5, p. 1671-1702, 2020. Disponível em: https://tinyurl.com/2z6ch8sz . Acesso em: 7 maio 2026.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Fairness and machine learning**: limitations and opportunities. Cambridge: MIT Press, 2019.

BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. **California Law Review**, Berkeley-California, v. 104, n. 3, p. 671-732, 2016.

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

BRASIL. Congresso Nacional. Senado Federal. **Projeto de Lei n. 2.338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília: Senado Federal, 2023. Disponível em: https://tinyurl.com/76fz639t. Acesso em: 9 maio 2026.

BRASIL. Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei n. 12.965, de 23 de abril de 2014 (Marco Civil da Internet). **Diário Oficial da União**, Brasília, 15 ago. 2018. Disponível em: https://tinyurl.com/2x797zr9. Acesso em: 28 jul. 2025.

BROWN, Simone. **Dark matters**: on the surveillance of blackness. Durham: Duke University Press, 2015.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender shades: intersectional accuracy disparities in commercial gender classification. **Proceedings of Machine Learning Research**, [*s. l.*], v. 81, p. 1-15, 2018. Disponível em: https://tinyurl.com/yjbb98mk. Acesso em: 7 maio 2026.

CFPB – CONSUMER FINANCIAL PROTECTION BUREAU. **Consumer financial protection circular 2022-03**: adverse action notification requirements in connection with credit decisions based on complex algorithms. Washington: CFPB, 2022.

CITRON, Danielle K.; PASQUALE, Frank A. The scored society: due process for automated predictions. **Washington Law Review**, Seattle-WA, v. 89, n. 1, p. 1-33, 2014.

DASTIN, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women. **Reuters**, San Francisco, 10 out. 2018. Disponível em: https://tinyurl.com/ykxz2rc4. Acesso em: 28 jul. 2025.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**: fundamentos da Lei Geral de Proteção de Dados. 2. ed. São Paulo: Revista dos Tribunais, 2020.

DUTRA, Daniele. Mulher é confundida com foragida por sistema facial da PM: "Racismo". **UOL Notícias**, Rio de Janeiro, 11 jul. 2024. Disponível em: https://tinyurl.com/3wruztcw. Acesso em: 7 ago. 2024.

ESTADOS UNIDOS – EUA. **Equal Credit Opportunity Act** (ECOA). Pub. L. No. 93-495, 15 U.S.C. § 1691 *et seq*. Washington: Government Printing Office, 1974.

ESTADOS UNIDOS – EUA. **Fair Housing Act**. Pub. L. No. 90-284, Title VIII, 82 Stat. 81, 42 U.S.C. § 3601 et seq. Washington: Government Printing Office, 1968.

ESTADOS UNIDOS – EUA. Federal Reserve History. Redlining. **Federal Reserve History**, Washington, 2 jun. 2023. Disponível em: https://tinyurl.com/2cj65mtc. Acesso em: 6 maio 2026.

FEENBERG, Andrew. **Transforming technology**: a critical theory revisited. Oxford: Oxford University Press, 2002.

FRAZÃO, Ana. Discriminação algorítmica: ciência dos dados como ação política. **Jota Info**, São Paulo, 2021a. Disponível em: https://tinyurl.com/mrer8zz4. Acesso em: 10 ago. 2024.

FRAZÃO, Ana. Discriminação algorítmica: por que algoritmos preocupam quando acertam e erram? **Jota Info**, São Paulo, 2021b. Disponível em: https://tinyurl.com/3xvdueuu. Acesso em: 28 jul. 2025.

FRAZÃO, Ana. Discriminação algorítmica: a responsabilidade dos programadores e das empresas. **Jota Info**, São Paulo, 2021c. Disponível em: https://tinyurl.com/yx88ck2j . Acesso em: 10 ago. 2024.

GEBRU, Timnit *et al*. Datasheets for datasets. **Communications of the ACM**, Nova York, v. 64, n. 12, p. 86-92, 2021. DOI: https://doi.org/10.1145/3458723.

KITCHIN, Rob. **The Data revolution**: Big Data, open data, data infrastructures and their consequences. London: SAGE Publications, 2017.

KONDER, Carlos Nelson. O tratamento de dados sensíveis à luz da Lei 13.709/2018. *In*: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no Direito brasileiro**. 2. ed. São Paulo: Thomson Reuters Brasil, 2020. p. 441-459.

LUM, Kristian; ISAAC, William. To predict and serve? Predictive policing and law enforcement. **Significance**, London, v. 13, n. 5, p. 14-19, out. 2016. Disponível em: https://tinyurl.com/4mmhu6tu . Acesso em: 6 maio 2026.

MANN, Monique; MATZNER, Tobias. Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination. **Big Data & Society**, [*s. l.*] v. 6, n. 2, p. 1-11, jul./dez. 2019. DOI: https://doi.org/10.1177/2053951719895805.

MITCHELL, Margaret *et al*. Model cards for model reporting. *In*: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAT*), Atlanta, 2019. **Proceedings** [...]. Nova York: ACM, 2019. p. 220-229. DOI: https://doi.org/10.1145/3287560.3287596.

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

O'NEIL, Cathy. **Weapons of math destruction**: how Big Data increases inequality and threatens democracy. New York: Crown Publishing Group, 2016.

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

PAUL, Kari. Health care algorithm used in US hospitals favors white patients over sicker black patients. **The Guardian**, Londres, 25 out. 2019. Disponível em: https://tinyurl.com/98hv8sua. Acesso em: set. 2026.

PROPUBLICA; ANGWIN, Julia *et al*. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. **ProPublica**, Nova Iorque, 23 maio 2016. Disponível em: https://tinyurl.com/2s4758we. Acesso em: 28 jul. 2025.

SILVA, Matheus Thiago Domingos da. **Vieses em algoritmos de reconhecimento facial**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Centro de Engenharia Elétrica e Informática, Universidade de Campina Grande, Campina Grande, 2024. Disponível em: https://tinyurl.com/49tbth3x. Acesso em: 7 maio 2026.

SOLOVE, Daniel J. Data is what data does: regulating based on harm and risk instead of sensitive data. **GWU Legal Studies Research Paper,** Washington, D.C., n. 2023-22, jan. 2023. Disponível em: https://tinyurl.com/tkvaf37w. Acesso em: 7 maio 2026.

SRNICEK, Nick. **Platform Capitalism**. Cambridge: Polity Press, 2017.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 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**, Luxemburgo, 4 maio 2016. Disponível em: https://tinyurl.com/4j4h5hfe. Acesso em: 28 jul. 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024. Estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento da Inteligência Artificial) e que altera os Regulamentos (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828. **Jornal Oficial da União Europeia**, Luxemburgo, 12 jul. 2024. Disponível em: https://tinyurl.com/42kr84bp. Acesso em: 9 maio 2026.

VIGDOR, Neil. Apple Card investigated after gender discrimination complaints. **The New York Times**, Nova Iorque, 10 nov. 2019. Disponível em: https://tinyurl.com/jtvspr2s. Acesso em: 28 jul. 2025.

WACHTER, Sandra; MITTELSTADT, Brent. A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI. **Columbia Business Law Review**, New York, v. 2019, n. 2, p. 494-620, 2019.

WINNER, Langdon. **The whale and the reactor**: a search for limits in an age of high technology. Chicago: University of Chicago Press, 1986.

ZARSKY, Tal Z. The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. **Science, Technology, & Human Values**, [*s. l.*], v. 41, n. 1, p. 118-132, 2016.