DISCRIMINAÇÃO ALGORÍTMICA E A DIGNIDADE HUMANA: O IMPACTO AO RESPEITO, À AUTONOMIA E À AUTODETERMINAÇÃO

Priscila Sauthier[1]

Editor Responsável: Leonardo Rodrigues
Revisão textual: Willian Albino

RESUMO

A Inteligência Artificial (IA) e os algoritmos compõem uma tecnologia de aprendizado avançado que participa cada vez mais do cotidiano das pessoas, seja nos espaços digitais, no uso de aplicativos, no gerenciamento de e-mails ou em serviços online. Sua popularidade baseia-se na promessa de tornar a execução de tarefas e a resolução de problemas mais rápidas, baratas e eficientes. Contudo, sua atuação na vida das pessoas motivou uma preocupação não somente sobre como os dados são coletados, mas também quais são as consequências ocasionadas aos indivíduos que são submetidos à sua análise. Em destaque, a discriminação no algoritmo. Destarte, o presente artigo possui o escopo de analisar as ocorrências desse fenômeno e como ele afeta o princípio da dignidade da pessoa humana, em especial aos seus elementos constitutivos: respeito, autonomia e autodeterminação. Para isso, adota-se uma pesquisa qualitativa, de caráter exploratório-dedutivo, que utiliza reportagens que abordam casos exemplificativos relevantes ao tema, legislação, livros e artigos científicos interdisciplinares como fontes bibliográficas. Identificou-se que a tecnologia não possui neutralidade e que a prática da

[1] Especialista em Direito Público (Constitucional, Administrativo e Tributário) e em Direito Penal e Processual Penal pela Universidade Estácio de Sá. Especializanda em Ética em Inteligência Artificial pela Universidade Federal da Paraíba. Bacharela em Direito pela Universidade Presidente Antônio Carlos (UNIPAC). ORCID: https://orcid.org/0009-0007-4593-6979. E-mail: sauthierpriscila@gmail.com

Recebido: 15/08/2025. Aprovado: 21/01/2026

generalização no tratamento das informações pode implicar resultados indignos, sujeitando humanos à condição de objetos. Sua capa lógica alimenta uma imagem equivocada de inquestionabilidade que incentiva a abdicação da autonomia individual. E, em meio às representações danosas, negativas de oportunidades, exclusões de serviços e obstáculos ao acesso à justiça concreta, a autodeterminação é ferida. Assim, torna-se indispensável a consideração da dignidade humana na idealização e desenvolvimento de soluções de IA, a ser aplicada por instituições privadas, públicas e profissionais de tecnologia.

Palavras-chave: Inteligência Artificial; Autonomia; Instrumentalização do indivíduo; Ética tecnológica.

ALGORITHMIC DISCRIMINATION AND HUMAN DIGNITY: THE IMPACT ON RESPECT, AUTONOMY AND SELF-DETERMINATION

ABSTRACT

Artificial Intelligence (AI) and algorithms constitute an advanced learning technology that is increasingly integrated into daily life, whether in digital Spaces, through the use of applications, e-mail management, or online services. Its popularity is rooted in the promise of making task execution and problem-solving faster, more cost-effective and efficient. However, its influence on people's lives has raised concerns regarding not only how data is collected but also the consequences for individuals subjected to its analysis. Of particular note is the occurrence of algorithmic discrimination. Therefore, this article aims to analyze the occurrences of this phenomenon and how it affects the principle of human dignity, specifically its constitutive elements: respect, autonomy and self-determination. To this end, a qualitative research approach is adopted, characterized as exploratory-deductive, utilizing reports of exemplary cases relevant to the topic, legislation, books, and interdisciplinary scientific articles as bibliographic sources. It was identified that technology is not neutral,

and the practice of generalization in information processing can result in indignities, reducing humans to the status of objects. Its logical facade fuels a misguided image of unquestionability that encourages the abdication of individual autonomy. Also, amidst harmful representations, the denial of opportunities, exclusion from services and barriers to substantive justice, human self-determination is undermined. Consequently, the consideration of human dignity becomes indispensable in the design and development of AI-driven solutions, to be applied by private and public institutions, as well as technology professionals.

Keywords: Artificial Intelligence; Autonomy; Instrumentalization of the individual; Technology ethics.

INTRODUÇÃO

As pessoas estão constantemente interagindo com modelos de Inteligência Artificial e, consequentemente, algoritmos. Seja em computadores ou nos *smartphones*, em espaços online, como plataformas de mídias sociais, ou no uso de assistentes virtuais, modelos conversacionais, como o ChatGPT, aplicativos para gerenciamento de agenda, e-mails e serviços de *streaming* de músicas, filmes e séries. São tecnologias de aprendizado avançado e de contínuo desenvolvimento que alcançaram popularidade sob a promessa de tornar o exercício de tarefas e a resolução de problemas mais rápidos, baratos e eficientes, especialmente por meio da automatização.

Logo, o mercado produziu opções de sistemas com algoritmos que podem tanto ser anexados a serviços e espaços digitais quanto se tornar ferramentas de apoio de previsão e decisão. Os primeiros são munidos da capacidade de identificar interesses, desejos, opiniões e informações pessoais de seus usuários para o oferecimento da personalização de experiências. Os segundos possuem as funções de análise de conjuntos de dados e auxílio na solução de problemas.

Surge, em seguida, uma preocupação não somente pelo modo como as informações das pessoas são recolhidas, mas também pelas consequências que o seu tratamento ocasiona. Modelos, especialmente quando se revelam verdadeiras

representações de caixas-pretas, vêm provocando, durante os anos, a possibilidade de processamentos de informação manifestarem em discriminação. Um problema que não pode ser ignorado. E, como contraposição a esse tipo de resultado, o qual reduz o valor das pessoas, ressalta-se o princípio da dignidade da pessoa humana.

Dessa forma, busca-se analisar, no presente artigo, as ocorrências de discriminação promovidas pelo algoritmo e como esse fenômeno afeta os elementos constitutivos da ideia de dignidade, devida a todos os seres humanos que utilizam ou são submetidos ao julgamento de modelos de Inteligência Artificial. A reflexão colaborará para confirmar a importância de cultivar a presença do princípio da dignidade da pessoa humana, reconhecido internacionalmente e por Constituições internas, nas relações dos humanos com a tecnologia mais recente.

Para tanto, adota-se uma pesquisa qualitativa, de caráter exploratório-dedutivo, que utiliza reportagens, reunindo exemplos ilustrativos sobre o tema, legislação, livros e artigos científicos interdisciplinares como fontes bibliográficas. Salienta-se que, em procura de pesquisas científicas que identificassem e avaliassem casos concretos de discriminação por meio do algoritmo, foram consultadas as bases de dados SciELO, usando a combinação dos termos “Inteligência Artificial” AND “viés racial” e “Inteligência Artificial” AND “Direito”; e o Portal de Periódicos CAPES, com o termo “*Riscanvi*”, bem como utilizou-se o método de candeia de referências retrospectiva para enriquecer a apreciação sobre a problemática.

Primeiramente, apresenta-se um entendimento basilar sobre a concepção de dignidade da pessoa humana, com a especial contribuição de doutrinadores do Direito, para, em seguida, averiguar se a discriminação, como manifestação da ação do algoritmo, constitui um meio de desumanização diante da prática da generalização danosa. Após, procede-se à análise de como ela afeta os elementos da obrigação do respeito, da autonomia e da autodeterminação dos indivíduos.

1. A DIGNIDADE DA PESSOA HUMANA

A ideia de dignidade da pessoa humana alcançou destacada representação em ordenamentos jurídicos do mundo no período subsequente ao fim da Segunda Guerra Mundial, este marcado pela crise do positivismo jurídico, essencialmente formalista e indiferente a valores éticos, que oportunizou o uso da lei como justificativa de defesa para o cometimento de atos nazifascistas violentos (Piovesan, 2021). A resposta foi tornar a dignidade da pessoa humana um pressuposto para a garantia da justiça, a qual se manifesta no Direito (Rocha, 2009).

De um lado, ocorreu uma movimentação da comunidade internacional em reunir um conjunto de valores comuns, reconhecendo a existência de direitos inerentes a toda pessoa humana (Reis, 2023), que nasce livre e igual em dignidade. Emergiu o Direito Internacional dos Direitos Humanos, fundamentado nesse conceito. Paralelamente, no âmbito interno dos textos constitucionais ocidentais, procurou-se a inclusão de parâmetros éticos por meio do reconhecimento da força axiológica dos princípios, reservando especial atenção à dignidade (Piovesan, 2021).

No Brasil, a dignidade da pessoa humana foi expressamente elencada como fundamento do Estado Democrático de Direito na Constituição da República Federativa de 1988, concedendo-lhe o título de princípio fundamental, em seu art. 1º, inc. III (Brasil, 1988). Uma posição inovadora no ordenamento jurídico brasileiro que reconheceu, conforme pontua Sarlet (2011), uma dimensão de liberdade ao ser humano, o qual não pode ser considerado um mero objeto para a atividade estatal, mas sim a sua finalidade. Demonstra-se, como princípio jurídico, uma natureza dual que abrange tanto uma face negativa, impondo limites ao poder estatal, quanto uma face positiva-prestacional, ao estipular que cabe ao Estado o dever de ativamente proteger, promover e concretizar uma vida digna aos seus cidadãos.

Em movimento de contraposição à ideologia nazista de categorizar, inferiorizar e subjugar humanos a fins degradantes (Rocha, 2009), instituiu-se a dignidade da pessoa humana como um valor intrínseco a toda e qualquer pessoa, o qual compreende o elemento da autodeterminação consciente e responsável sobre a própria vida e implica a

pretensão do respeito mútuo entre pessoas. Seu objetivo é garantir o mínimo de estima que toda pessoa merece diante de sua condição de ser humano, detentor do direito de ser feliz (Moraes, 2023). Uma qualidade, destaca-se, de caráter incondicional, diante da inexistência de condição capaz de provocar a perde de seu valor, e incomparável, pois o seu valor se sobrepõe a todos os outros valores e a eles não se sujeita (Formosa, 2017).

O valor intrínseco interpreta-se em duas dimensões: como elemento ontológico da dignidade e como fundamento de direitos fundamentais. O primeiro, de natureza filosófica, refere-se ao reconhecimento da natureza especial da pessoa (Barroso, 2022). Diferentemente das coisas, ela é possuidora de dignidade e, por isso, há a indispensabilidade de considerá-la como um fim em si mesma e nunca como meio para fins de outrem (Carvalho, 2017). O segundo aplica-se no incentivo, por exemplo, ao direito a igualdade. Compreende-se que, por serem dotadas de mesmo valor, as pessoas não podem sofrer discriminações arbitrárias e merecem ser reconhecidas tanto em suas particularidades quanto como representantes de grupos sociais minoritários (Barroso, 2022).

Outrossim, a autonomia, como elemento ético da dignidade, engloba a autodeterminação, ou seja, a capacidade de cada pessoa desenvolver a sua personalidade (Barroso, 2022) e conceder sentido à sua vida, independente de forças externas, sejam elas de caráter benevolente ou paternalista. Destarte, nenhuma figura de autoridade possui o direito de constranger uma pessoa a se tornar um instrumento de alcance de algum resultado (Carvalho, 2017).

A Ministra do Supremo Tribunal Federal, Cármen Lúcia (2009), acrescenta que a dignidade da pessoa humana se tornou matriz do direito contemporâneo e possui uma acepção que não se limita ao físico, fundamentando-se na integridade e inviolabilidade do ser humano em dimensões existenciais. Tal conceito é essencial para sistemas jurídicos que procuram acalentar direitos fundamentais e combater a desumanização movida por tecnologias que fazem parte de uma política mercadológica canibal, a qual provoca a exclusão justamente porque a inclusão social não garante a produção de lucro.

2. DISCRIMINAÇÃO NO ALGORITMO: DA GENERALIZAÇÃO À DESUMANIZAÇÃO

Para atender a demanda de recomendação de conteúdo multimídia em serviços de streaming, geração de imagem, atendimento ao cliente, escolha de candidatos para uma vaga de emprego ou monitoramento para avaliação de performance em esportes, dentre outras funções relacionadas à sua aptidão em calcular previsões e aconselhar decisões, modelos de IA e seus algoritmos são habitualmente projetados para analisar uma grande quantidade de informações e simplificá-las para gerar a resposta desejada da forma mais eficiente possível.

Para isso, o algoritmo atua como um conjunto de instruções que municia um passo a passo para a resolução de problemas, de cunho lógico e matemático (Eysenck, M; Eysenck, C., 2023), o qual precisará de uma seleção de variáveis. Por exemplo, se o objetivo é que o modelo identifique um cavalo, será necessária a rotulação das características básicas do animal. Contudo, se, ao invés de definir o que é um cavalo, precisássemos que a máquina identificasse e distinguisse rostos masculinos de rostos femininos? A tarefa torna-se mais complicada e pode ser afetada por perspectivas pessoais, conscientes ou não (Taulli, 2020).

Com o aperfeiçoamento da técnica de aprendizagem, o *deep learning*, também conhecido como aprendizado profundo, se tornou a vanguarda da Inteligência Artificial. A partir dele, os modelos começaram a processar uma massiva quantidade de dados e adquiriram a capacidade de identificar os recursos nos dados, pixel por pixel, fazendo as relações necessárias e interpretando padrões que os seres humanos muitas vezes não possuem capacidade de detectar (Taulli, 2020). No entanto, essa abordagem requer um treinamento mais extenso para apresentar níveis de desempenho elevados (Eysenck, M; Eysenck, C., 2023).

Godoy, Reyes e Urrutia (2024), membros diretivos da Associação de Ética em Dados e Inteligência Artificial da Universidade do Chile, opinam que, ao realizar o treinamento de uma IA, espera-se que ela demonstre o melhor desempenho na resolução de um problema de forma geral e não somente nos dados escolhidos para o treinamento. O objetivo, assim, não é conseguir que a IA detecte a diferença de um cavalo nas imagens

utilizadas para treiná-la, mas que ela seja capaz de detectar a diferença em imagens externas, não participantes do treinamento, porque, desse modo, ela demonstra a capacidade de generalizar.

Nota-se que, mesmo que o avanço dos métodos de aprendizagem tenha permitido uma crescente independência dos modelos, o cerne da garantia de sua eficiência mantém-se na busca pela relação entre os dados. Uma competência de organizar as informações, relacionando-as e definindo conceitos gerais para facilitar a sua análise e posterior conclusão. No entanto, se a potencialização da IA com a *Big Data* e com a maior capacidade geral dos componentes computacionais é capaz de gerar benefícios, igualmente se apresenta a possibilidade de causar prejuízos.

Os modelos de IA de recrutamento, por exemplo, receberam e continuam sendo alvo de grande entusiasmo pelas empresas, especialmente as *Big Techs* (empresas de grande impacto tecnológico e comercial). No entanto, em alguns casos, como a IA da Amazon, criada em 2014 para a automatização do processo de seleção de pessoas, foi marcada pela desconsideração de candidatas apenas por serem mulheres. O sistema, especificamente, considerou a palavra “mulher” como termo negativo após analisar os perfis contratados pela empresa nos últimos dez anos, sendo estes predominantemente masculinos (Dastin, 2018). Para Harari (2024), isso significa que o algoritmo é capaz de assimilar a misoginia arraigada nas empresas, pois encontra padrões preconceituosos em seus bancos de dados.

Em 2018, foi a vez da plataforma LinkedIn. A rede social corporativa precisou corrigir o algoritmo utilizado no seu sistema de vagas de emprego, o qual demonstrou um comportamento tendencioso. Embora o nome, idade, sexo e raça não estivessem incluídos no algoritmo original, o aprendizado de máquina da IA conseguiu detectar padrões de comportamento por gênero e, no fim, indicou mais homens para as vagas abertas do que mulheres (D'Almeida, 2021), inibindo a oferta igualitária de oportunidades.

Frermann *et al.* (2023), contribuíram para essa reflexão ao confirmarem em sua pesquisa que algoritmos podem identificar o gênero inclusive em uma análise cega de currículo e realizar previsões acerca da qualidade do candidato. Ainda, descobriram que, mesmo tentando impedir o resultado discriminatório pelo ChatGPT, modelo de IA

generativa, ele ocorreu na identificação da condição de mãe entre as ~~das~~ candidatas. Os currículos femininos, afinal, possuem maior probabilidade de apresentar lacunas temporais em razão da licença maternidade.

Conforme pontua Cathy O'Neil (2020), conceder à máquina o papel de analisar milhares de currículos e selecionar os melhores candidatos pode transparecer, primeiramente, uma boa solução diante da economia de tempo e de uma suposta imparcialidade e objetividade do modelo, pois o processo não envolveria um humano capaz de ser preconceituoso. Não obstante, deve-se atentar ao fato que as aplicações matemáticas que sustentam o funcionamento de qualquer modelo são criadas e influenciadas previamente pelas escolhas de humanos. E humanos podem falhar.

As escolhas sobre o cargo que se objetiva preencher, quais são os dados que serão utilizados, quais informações terão um peso maior na avaliação do currículo, dentre outras, são realizadas previamente por um humano para a posterior aplicação pelo modelo. Escolhas que podem conter percepções particulares, equívocos ou vieses humanos que serão embutidos nos algoritmos (O'Neil, 2020). São elementos que participarão da programação desse sistema e impactarão as saídas da avaliação do modelo.

Como o caso explicitado por Cathy O'Neil (2020) sobre o sistema de previsão de crimes *PredPol*, o qual baseia-se na localização geográfica dos crimes. A ideia de usar o modelo para auxiliar na segurança pública não é, em si, um demonstrativo de má intenção, mas a escolha de expandir a abrangência dos crimes, de violentos aos de menor grau, direcionou a presença policial para bairros empobrecidos, com a maioria dos moradores sendo negros e hispânicos. O que, então, culminou em um maior número de abordagens dessas populações e uma maior pontuação a essas regiões. Ocorrendo o que a pesquisadora definiu como ciclo nocivo de *feedback*, no qual a própria polícia auxilia na produção de dados que justificam o retorno contínuo a essas áreas e aumentam as chances de prisões dos moradores. No fim, ao designar o algoritmo sem considerar a segregação urbana já historicamente existente, permitiu-se a sua reafirmação danosa.

Outrossim, destaca-se a escolha sobre as bases de dados usadas para o treinamento dos modelos. No caso de modelos de análise facial, por exemplo, se escolhidas bases que possuem em sua maioria imagens representando pessoas brancas

caucasianas, enquanto são desfalcadas de boas representações que englobem diferentes nacionalidades, grupos étnico-raciais ou gênero, o aprendizado estará prejudicado e significará, posteriormente, um acúmulo de erros no reconhecimento de pessoas não caucasianas. O teste feito pelo Instituto Nacional de Padrões e Tecnologia (NIST, na sigla em inglês), comprovou essa ocorrência, detectando em programas feitos nos Estados Unidos uma taxa de erro entre 10 e 100 vezes na identificação de um asiático ou um afro-americano em comparação a um caucasiano. Já quando os projetos foram feitos na Ásia, a frequência dos falsos positivos se destacou mais na avaliação de pessoas caucasianas do que asiáticas (Rivera, 2025). É incontestável, a diversidade nos bancos de dados também é importante para se evitar situações discriminatórias.

Ante o exposto, observa-se que não é possível afirmar que a tecnologia é neutra (Noble, 2018; Odle *et al.*, 2025), pois são criadas, por meio dos algoritmos, versões do mundo real (Callejón, 2023) simplificadas e cabíveis de leitura por operações matemáticas e estatísticas. Nesse sentido, a generalização ou a criação de perfis podem compor o meio possível para a análise dos massivos conjuntos de informações e para a transferência da linguagem do mundo real para a das máquinas, no entanto, suas versões são influenciadas por ideologias, desejos e interesses humanos, de maneira consciente ou inconsciente. Inclusive, pela escolha das bases de dados para treinamento dos modelos.

Na opinião de Tarcízio Silva (2022), quando os algoritmos intercedem de forma a incluir ou excluir pessoas e afetam o seu valor frente aos objetos e ao capital, mantendo verdadeiras relações de poder em prol de poucos e criando impactos sociais negativos, eles extrapolam a sua função básica de linhas de código. Possibilitando, dessa forma, que a estratificação social historicamente construída, impregnada com injustiças (O'Neil, 2020), se torne componente das arquiteturas e linguagens tecnológicas (Noble, 2018) e, por elas, redistribuída. Cria-se, a partir disso, a oportunidade para que a discriminação negativa, caracterizada pelos tratamentos distintos que desconsideram grupos de pessoas em razão de premissas irrazoáveis ou ilegítimas e pela assimilação de conceitos culturais dominantes (Barroso, 2022), seja reafirmada e fortalecida pelas dimensões automatizadoras dos algoritmos. Odle *et al.* (2025) denominam esse fenômeno de viés algorítmico.

Outrossim, a problemática de manutenção de um *status quo* desigual, o qual privilegia alguns enquanto desumaniza outros, adquire maior gravidade ao se considerar que os modelos de IA são, em sua maioria, uma caixa-preta. Ou seja, os pressupostos de funcionamento ou as regras contidas nos modelos são mantidos propositalmente em segredo, sob a justificativa de que o algoritmo é um "molho secreto" de propriedade privada (O'Neil, 2020) ou possui uma complexidade que impossibilita a prestação de contas e o entendimento pelas pessoas (Harari, 2024). E, uma vez que as pessoas não possuem acesso às razões ou não conseguem compreender todos os elementos que direcionaram ao tratamento que receberam, mesmo ao sentirem que foi injusto, terão dificuldades para contestá-lo.

2.1 Do comprometimento do respeito

Recentemente, em uma análise de seis diferentes modelos de IA Generativa, Assis e Moura (2025), relataram que, ao digitar *prompts* ou comandos para a busca de imagens de homens negros trabalhando, 75% geraram representações relacionadas a alguma atividade laboral braçal ou manual, sendo que 37,5% traziam os homens trajando um avental. Ao especificarem ainda mais o pedido, solicitando a imagem de um homem negro brasileiro trabalhando, obtiveram um resultado de 100% das imagens referenciando atividades braçais. Na opinião dos autores, o resultado reforçou a reprodução de estereótipos históricos de servidão e subalternidade. Por sua vez, Alenichev *et al.* (2023) demonstraram que o modelo *Midjourney Bot* Versão 5.1 foi incapaz de criar imagens do conjunto descritivo "médicos negros africanos e pacientes crianças brancas".

Nesse diapasão, em estudo sobre esses tipos de ferramentas, Moura e Braga (2023) constataram uma sexualização do corpo feminino ao pesquisarem a geração de imagens de mulheres brancas e negras com deficiência. Em 52 imagens, 36% evidenciaram partes do corpo como seios, busto ou quadril. Em relação às mulheres negras, o resultado apresentou figuras mais sensualizadas em mais que o dobro em comparação às mulheres brancas. Ainda, os modelos demonstraram dificuldades em criar uma imagem que corresponda a uma pessoa com deficiência em 90% do total gerado.

Salienta-se, nos casos em amostra, que as conclusões da análise algorítmica demonstraram na reprodução de certos grupos sociais, em diferentes graus, a percepção de terceiros sobre eles. Da invisibilidade à hipervisibilidade, a representação desses grupos por meio das ferramentas tecnológicas baseou-se tão somente nos interesses de pessoas que não compõem esses determinados grupos. Conforme pontua Carrera (2020), em seu estudo sobre o sexismo e racismo em bancos de imagens digitais, nega-se para algumas pessoas as representações da vivência da felicidade, riqueza, sucesso e beleza enquanto, para outras, não se reconhece sequer a sua existência fora de certos parâmetros.

Um problema, inclusive, denunciado por Safiya Noble (2018) à época de sua análise crítica sobre os sistemas de classificação, o mecanismo comercial de busca do Google e a correlação entre *black girls* (ou garotas negras) e pornografia. Em sua pesquisa, ela argumenta que, movido pelo olhar ideológico patriarcal branco e pelo interesse do capital, o sistema de propagandas tornou comum o referenciamento em destaque de mulheres e meninas negras para atender o consumo de pornô fetichista. Tal dinâmica incentiva um direcionamento de pensamento que reforça sub-representações danosas à própria concepção de identidade dessas mulheres.

Ora, ao realizar uma busca, a mensagem entregue a essas meninas é que o seu corpo e o seu ser são uma mercadoria, passíveis de descarte. O maior risco dessa instrumentalização imposta sobre elas consiste no ataque à sua relação com o seu próprio eu (Rivera, 2025) ao ponto de fazê-las acreditar nessa mensagem. Uma deturpação da imagem do ser que, inclusive, limita a possibilidade de darem sentido às suas vidas, sob o pretexto da utilidade que possuem para outras pessoas.

Do mesmo modo, quando os grandes modelos de linguagem, que sustentam as plataformas de IA generativa, causam uma hiperexposição de respostas que descrevem mulheres como trabalhadoras domésticas e ligadas ao lar familiar enquanto homens são associados a “negócios” e “carreira”. Um estereótipo de gênero limitador que persiste em determinar papéis tradicionais que privilegiam homens (UNESCO, 2024). Adicionalmente, cita-se o *FaceApp*, um aplicativo famoso em criar figuras envelhecidas de rostos, utilizado para a edição de fotos, que alterou a cor da pele dos usuários para torná-los mais semelhantes ao imaginário social de beleza europeia (Gnipper, 2017), reforçando uma

percepção originada no processo histórico de colonização e subjugação de culturas (Silva, 2022).

Essas são amostras exemplificativas reais que elucidam a associação de certas caraterísticas ou funções a específicos grupos, formando um conceito geral deles, mas de uma maneira que os distingue desfavoravelmente de outras pessoas. Estas últimas, em comparação, recebem um tratamento que as eleva a um arquétipo do que seria o normal ou o padrão aceitável. Trata-se de um constructo de visão binária de superior e inferior que desmerece aspectos identitários de indivíduos, promovendo estereótipos, com generalizações impróprias, que marginalizam culturalmente pessoas para legitimar arranjos sociais desequilibrados (Silva, 2022).

Assim, ao desconsiderar as individualidades das pessoas, agrupando-as em noções exclusivamente benéficas para um terceiro e não para elas, os modelos as rebaixam a meros objetos de satisfação de outrem. Enquanto a produção de modelos e o *design* de algoritmos focarem em públicos-alvo específicos e sujeitarem os demais à consideração ou desconsideração baseada em parâmetros de utilidade, rejeita-se o mínimo de respeito que cabe a todo ser humano, que não deve ser precificado.

3. Do dano à autodeterminação e à autonomia

Sistemas algorítmicos cada vez mais participam de processos de decisões, especialmente sistemas automatizados. Para Díaz e Liberal (2024), se esse algoritmo for capaz de se antecipar e direcionar a decisão humana, ele adquire um caráter paternalista. E sua presença pode refletir inconscientemente na crença de que a IA possui uma melhor capacidade de emitir decisões racionais e desprovidas de vieses do que humanos. Impulsionando, por consequência, a delegação da capacidade de decisão às máquinas, mesmo sob o risco de reduzir a autonomia humana. Um fenômeno conhecido como palavra de máquina ou *word-of-machine effect*.

Observa-se o caso de sistemas de suporte que avaliam riscos de reincidência criminal como o *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) ou Perfil de Gestão de Delinquentes Correcionais para Sanções Alternativas. O

sistema de IA em questão apontava 45% mais chances de réus negros obterem pontuações mais altas de risco quando comparados com réus brancos que possuíam características semelhantes, conforme estudo independente da corporação ProPublica (Larson; Angwin, 2016). Para Loomis, isso significou uma pontuação de alto risco e influenciou uma condenação de oito anos e seis meses de prisão (Angwin *et al*., 2016).

Os algoritmos do COMPAS fundamentaram o resultado probabilístico em bases de dados sobre pessoas que antecederam Loomis e das quais ele não tinha meios de ter ciência sobre quem eram (Rivera, 2025), mas foi o suficiente para um juiz sentenciar uma maior limitação à sua liberdade. Mesmo o sistema sendo opaco, confiou-se na palavra de máquina e deixou-se de realizar uma análise qualitativa da situação individual do réu.

Em experiência semelhante, mas incorporando a participação de profissionais supervisores no processo, cita-se o *RisCanvi*, utilizado na administração prisional da Catalunha. Descobriu-se que a sua aparência de cientificidade, além da impossibilidade de conhecer o funcionamento do algoritmo, criou uma camada de incontestabilidade sobre o programa. De um lado, prejudicou-se a correção de erros e o direito de defesa frente a injustiças, principalmente ao considerar que o seu uso é omitido às pessoas privadas de liberdade (Aróstegui, 2023). De outro, os supervisores convencidos da suposta objetividade e neutralidade do sistema, em sua maioria de 96,8%, limitavam-se a validar a avaliação do algoritmo, deixando de emitir de forma explícita os seus próprios juízos de valor. O grau de autonomia humana, dentro desse ciclo, torna-se claramente questionável (Díaz; Liberal, 2024).

Acrescenta Aróstegui (2023) que essa centralização de poder na combinação algorítmica implica a concepção da pessoa privada de liberdade como uma fonte de perigo que deve ser estudada, classificada e controlada. Um tratamento que mais a aproxima da condição de coisa do que uma pessoa de direitos. A IA, dessa forma, fomenta uma forma de justiça preditiva que cria perfis pessoais discriminadores, os quais podem ser alimentados pelo preconceito do programador do algoritmo, consciente disso ou não, assim como por fatores estáticos como o histórico criminal dos réus (Pedroso; Santos, 2024).

Há de se destacar que, ao se preterir o critério humano em favor de algoritmos de recomendações ou predições (Díaz; Liberal, 2024), sob uma equivocada percepção de que a tecnologia necessariamente acarreta a neutralidade, minou-se a autonomia humana em papéis essenciais de decisão, sem, contudo, evitar a produção de resultados discriminatórios. Frente a isso, é importante compreender que os modelos podem ser "opiniões embutidas em matemática" (O'Neil, 2020, p. 22) e, portanto, não é recomendável ceder ao pensamento de que algoritmos sempre acertarão em seus resultados, tampouco de que é preferível substituir completamente a análise humana.

Sob essa perspectiva, ressalta-se que as decisões automatizadas, não devidamente justificadas, podem afetar ainda a independência das pessoas sujeitas à avaliação algorítmica, as quais não receberão oportunidades de forma igualitária. Pelo contrário, terão acesso a um número reduzido de opções, prejudicando a sua liberdade de decisão sobre as suas próprias vidas.

Um processo de contratação, exemplifica-se, pode utilizar uma tecnologia de análise emocional e atribuir mais emoções de teor negativo aos sorrisos de homens negros em comparação a homens brancos. A consequência, além de causar um dano representacional, é o aumento das chances de profissionais de cor serem rejeitados nas entrevistas (Zwart, 2024). Enquanto a prestação de um serviço público, como o de emissão de passaporte, pode ser prejudicada frente a um sistema que possui dificuldades em reconhecer fotografias dependendo da cor da pele e do gênero (Ahmed, 2020).

Nesse sentido, sistemas de reconhecimento facial que cometem falsos positivos, sobremaneira de pessoas não caucasianas, incentivam o risco de acusações e prisões injustas de pessoas inocentes, o que já ocorreu: Nijeer Parks foi identificado equivocadamente em 2021 como sendo o autor de um furto de loja (Sarlin, 2021); Thaís Santos, auxiliar administrativa, foi considerada fugitiva duas vezes em um mesmo evento (Santiago, 2023); Shaun Thompson foi confundido por um homem procurado e detido até a apresentação do seu passaporte (Clayton, 2024); a prisão de Robert Williams em 2020 pela Polícia de Detroit acarretou um processo judicial que se encerrou em 2024 com um acordo de restrição do uso da tecnologia de reconhecimento facial pela polícia (American Civil Liberties Union, 2024).

Outro exemplo interessante a apresentar é o caso do modelo de apoio à decisão sobre saúde denominado *Optum*, o qual empregou os custos de tratamentos médicos como parâmetro de avaliação e acarretou a exclusão de pacientes negros. A ferramenta concluiu que como era gasto menos dinheiro com pacientes negros, os quais possuíam os mesmos níveis de necessidade de tratamento de pacientes brancos, eles ficavam menos doentes e, ao final, não recebiam a mesma qualificação para tratamento que o outro grupo. Embora explicitamente não existisse uma regra para a identificação racial, o resultado foi uma discriminação racial motivada por condições financeiras (Paul, 2019).

Sobre o tópico, Tina Smith Nelson, advogada supervisora da *American Association of Retired Persons Legal Counsel for the Elderly* (uma organização sem fins lucrativos que presta serviços gratuitos nas áreas jurídica e social para idosos nos Estados Unidos) ao opinar sobre um sistema algorítmico que desconsiderou a dificuldade de mobilidade de idosos e pessoas com deficiência, apontou como inquietante a ação de delegar para algoritmos a decisão sobre o fornecimento de atendimentos médicos (McCormick, 2021).

Há de se concordar com ela, afinal, ao se normalizar a concepção de que não somente é possível, mas teoricamente preferível ceder a competência de decisão aos algoritmos, abdica-se da autonomia individual humana sob uma equivocada crença de que o algoritmo, por ter uma natureza matemática, não seria capaz de perpetuar o tratamento discriminatório. No entanto, os casos supramencionados demonstram o contrário desse pensamento.

Na verdade, eles esclarecem que os algoritmos podem ser amplificadores da discriminação, sujeitando as pessoas a situações de vulnerabilidade, como a negativa de oportunidades, a exclusão de serviços comuns e até obstáculos no acesso à justiça concreta. Um problema que mina, primordialmente, a liberdade do ser humano em escolher os seus próprios caminhos, objetivos e as perspectivas do que deseja ser em sua vida. São interferências externas que desvalorizam o ser humano e suas capacidades, não restando dúvida quanto ao desrespeito ao valor axiológico da dignidade da pessoa humana.

CONSIDERAÇÕES FINAIS

Os modelos de Inteligência Artificial e os algoritmos são tecnologias que atualmente participam do cotidiano das pessoas e podem provocar impactos sociais negativos em suas vidas, pois nem toda a saída ou resultado de sua análise representa neutralidade ou justiça.

Observa-se que a discriminação movida pelo algoritmo não nasceu com ele. Sua origem é humana, mas ela alcança novas oportunidades de atuação ante a competência analítica e generalizadora da tecnologia, capaz de manter a produção de resultados desiguais e reforçar a exclusão de grupos sociais específicos. Embora o algoritmo seja, em sua essência, um conjunto estruturado de instruções, constata-se que ele é fortemente influenciado por escolhas humanas, o seu uso para a identificação de padrões possibilita que seres humanos sejam tratados como objeto, bem como a sua capa lógica pode conceber uma imagem equivocada de inquestionabilidade. Elementos que, conforme os exemplos recolhidos, apontados e analisados no presente estudo, facilitaram o enfraquecimento dos alicerces do valor-fonte dignidade, quais sejam, a autonomia, a autodeterminação e o respeito.

Nesse contexto, uma vez que a dignidade humana é um princípio base para o reconhecimento de direitos, englobando a igualdade e a não discriminação, e para a manutenção da vida feliz do ser humano, torna-se, além de interessante, imperativo que ela seja aplicada para a compreensão da necessidade de rever o desígnio concedido aos algoritmos e para servir de parâmetro de construção de qualquer modelo de Inteligência Artificial.

Portanto, entusiastas e solicitantes de modelos de IA para a execução de serviços, incluindo tanto empresas privadas quanto instituições públicas, bem como os profissionais que participam da idealização e desenvolvimento dessa tecnologia não podem se abster de aplicar o princípio. Sua atuação deve conduzir um olhar crítico que não se limite a interesses econômicos ou particulares, mas sim que envolva políticas de comprometimento ético e responsável com as soluções tecnológicas. Nesse ponto, torna-se recomendável, diante dos casos de discriminação algorítmica que persistem ao longo dos anos, estabelecer diretrizes normativas de governança de algoritmos e de sistema automatizados, balizados na dignidade humana, que garantam essa consciência ética.

REFERÊNCIAS

AHMED, Maryam. UK passport photo checker shows bias against dark-skinned women. **BBC News**. [*S. l.*], 7 out. 2020. News. Disponível em: https://www.bbc.com/news/technology-54349538. Acesso em: 13 jul. 2025.

ALENICHEV, Arsenii; KINGORI, Patricia; GRIETENS, Koen Peeters. Reflections before the storm: the AI reproduction of biased imagery in global health visuals. **The Lancet Global Health**, [*S.l.*], v. 11, n. 10, p. 1496-1498, 2023. Disponível em: https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(23)00329-7/fulltext?ref=axbom.com. Acesso em: 09 jun. 2025.

AMERICAN CIVIL LIBERTIES UNION. Williams v. City of Detroit: Face Recognition False Arrest. **Court Cases**, [*S.l.*], 29 jan. 2024. Privacy & Technology. Disponível em: https://www.aclu.org/cases/williams-v-city-of-detroit-face-recognition-false-arrest. Acesso em: 8 jun. 2025.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. **ProPublica**, *[S. l.]*, 23 maio 2016. Disponível em: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Acesso em: 7 jun. 2025.

ARÓSTEGUI, Lorena Alemán. El uso de RISCANVI en la toma de decisiones penitenciarias. **Estudios Penales y Criminológicos**, [*S. l.*], v. 44, n. Ext., p. 1-43, 10 nov. 2023. Disponível em: https://doi.org/10.15304/epc.44.8884. Acesso em: 12 ago. 2025.

ASSIS, Juliana de; MOURA, Maria Aparecida. Semioses algorítmicas e viés racial: um estudo de imagens criadas pela IA generativa. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, Brasil, v. 30, p. 1–24, 2025. Disponível em: https://periodicos.ufsc.br/index.php/eb/article/view/103495. Acesso em: 11 jul. 2025.

BARROSO, Luís Roberto. **Curso de Direito Constitucional Contemporâneo**: os conceitos fundamentais e a construção do novo modelo. 10. ed. São Paulo: SaraivaJur, 2022.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 06 de ago. 2025.

CALLEJÓN, Francisco Balaguer. **La Constitución del Algoritmo**. 2.ed. Zaragoza: Fundación Manuel Giménez Abad, 2023. *E-book*. Disponível em: https://www.fundacionmgimenezabad.es/sites/default/files/Publicar/publicaciones/documentos/e9_2ed_constitucion_algoritmo_seg.pdf. Acesso em: 05 jul. 2025.

CARRERA, Fernanda. Racismo e sexismo em bancos de imagens digitais: Análise de resultados de busca e atribuição de relevância na dimensão financeira/profissional. *In*: SILVA, Tarcízio (org.). **Comunidades, algoritmos e ativismos digitais**: olhares afrodiaspóricos. São Paulo: LiteraRUA, 2020. p. 139-155.

CARVALHO, Kildare Gonçalves. **Direito Constitucional**: Direito Constitucional Positivo. 22. ed. Belo Horizonte: Del Rey, 2017.

CLAYTON, James. 'I was misidentified as shoplifter by facial recognition tech'. **BBC Newsnight,** [*S.l.*], 25 mai. 2024. Disponível em: https://www.bbc.com/news/technology-69055945. Acesso em: 10 jun. 2025.

DASTIN, Jeffrey. Insight - Amazon scraps secret AI recruiting tool that showed bias against women. **Reuters,** San Francisco, 10 out. 2018. Disponível em: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/. Acesso em: 7 jun. 2025.

DÍAZ, Ujué Agudo; LIBERAL, Karlos G. **El algoritmo paternalista**: Cuando mande la inteligencia artificial. Iruñea-Pamplona: Katakrak Liburuak, 2024. *E-book*.

D'ALMEIDA, Nicole. Com IA viciada, LinkedIn criou algoritmo para corrigir algoritmo: entenda. **Uol Tilt**, São Paulo, 1 nov. 2021. Inteligência Artificial. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2021/11/01/linkedin-inteligencia-artificial-tendenciosa.htm. Acesso em: 7 jun. 2025.

EYSENCK, Michael; EYSENCK, Christine. **Inteligência Artificial X Humanos**: O que a ciência cognitiva nos ensina ao colocar frente a frente a mente humana e a IA. Tradução de Gisele Klein. 1 ed. Porto Alegre: Artmed, 2023.

FORMOSA, Paul. **Kantian ethics, dignity, and perfection**. New York: Cambridge University Press, 2017.

FRERMANN, Lea; NJOTO, Sheilla; CHEONG, Marc; RUPPANNER, Leah. When it comes to jobs, AI does not like parents: New research finds that AI doesn't just discriminate against women in the workforce – but also has a problem with parents. **Pursuit – University of Melbourne**, [*S.l.*], 23 jul. 2023. Sciences & Technology. Disponível em: https://pursuit.unimelb.edu.au/articles/when-it-comes-to-jobs-ai-does-not-like-parents. Acesso em: 7 jun. 2025.

GNIPPER, Patricia. Polêmica: aplicativo que altera as feições do usuário é acusado de racismo. O FaceApp conta com uma inteligência artificial para melhorar um sorriso, envelhecer rostos e deixá-los mais bonitos; o problema é que esse filtro também clareia o tom de pele das pessoas. **Techmundo.** [*S. l.*], 25 abr. 2017. Software. Disponível em: https://www.tecmundo.com.br/apps/116069-polemica-aplicativo-altera-feicoes-usuario-acusado-racismo.htm. Acesso em: 10 jun. 2025.

GODOY, Johnny; REYES, Camilo Carvajal; URRUTIA, Felipe. Es la ética un límite para la IA?. **Cuadernos de Beauchef**, Chile, v. 8, n. 1, p. 9-19, 2024. Disponível em: https://revistasdex.uchile.cl/index.php/cdb/article/view/15213/15245. Acesso em: 02 ago. 2025.

HARARI, Yuval Noah. **Nexus**: Uma breve história das redes de informação, da Idade da Pedra à inteligência artificial. Tradução de Berilo Vargas e Denise Bottmann. São Paulo: Schwarcz S.A., 2024.

LARSON, Jeff; ANGWIN, Júlia. Technical Response to Northpointe: Northpointe asserts that a software program it sells that predicts the likelihood a person will commit future crimes is equally fair to black and white defendants. We re-examined the data, considered the company's criticisms, and stand by our conclusions.

ProPublica, *[S. l.]*, 2016. Machine Bias. Disponível em: https://www.propublica.org/article/technical-response-to-northpointe. Acesso em: 7 jun. 2025.

MCCORMICK, Erin. What happened when a 'wildly irrational' algorithm made crucial healthcare decisions: Advocates say having computer programs decide how much help vulnerable people can get is often arbitrary – and in some cases downright cruel. **The Guardian**. [*S. l.*], 2 jul. 2021. US News. Disponível em: https://www.theguardian.com/us-news/2021/jul/02/algorithm-crucial-healthcare-decisions. Acesso em: 30 jul. 2025.

MORAES, Alexandre de. **Direito Constitucional**. 39. ed. Barueri: Atlas, 2023.
MOURA, Carlos Manoel Silva; BRAGA, Tiago Emmanuel Nunes. A Inteligência Artificial e a criação de conteúdo: os vieses que habitam entre nós. **Anais do Workshop de Informação, Dados e Tecnologia - WIDaT**, *[S. l.]*, v. 6, 2023. Disponível em: https://labcotec.ibict.br/widat/index.php/widat2023/article/view/11. Acesso em: 9 jul. 2025.

NOBLE, Safiya Umoja. **Algorithms of Oppression**: How Search Engines Reiforce Racism. New York: New York University Press, 2018.

ODLE, Brooke; FINLEY, Katherine; LONGFIELD, Victoria; SERRÃO, Rodrigo. Going Beyonde The Technical: Exploring Ethical and Societal Implications of Machine. *In*: BERRY, Carlotta A.; MARSHALL, Brandeis Hill (ed.). **Mitigating Bias in Machine Learning.** New York: McGraw Hill LLC, 2025. p. 17-31.

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

PAUL, Karin. Healthcare algorithm used across America has dramatic racial biases: System sold by Optum estimates health needs based on medical costs, which are much less than for white patients, report finds. **The Guardian**. [*S. l.*], 25 out. 2019. Health. Disponível em: https://www.theguardian.com/society/2019/oct/25/healthcare-algorithm-racial-biases-optum#:~:text=Less%20money%20is%20spent%20on,the%20same%20level%20of%20illness. Acesso em: 11 jul. 2025.

PEDROSO, João; SANTOS, Andreia. Inteligência Artificial e Justiça Criminal: Riscos e Desafios. **Revista Sociologia Online**, Lisboa, n. 35, p. 134-155, ago. 2024. Disponível em: https://doi.org/10.30553/sociologiaonline.2024.35.6. Acesso em: 12 ago. 2025.

PIOVESAN, Flávia. **Direitos humanos e o direito constitucional internacional**.19. ed. São Paulo: Saraiva Educação, 2021.

REIS, Ana Rosa. Campanha celebra os 75 anos da Declaração Universal dos Direitos Humanos. **Centro de Informação das Nações Unidas para o Brasil (UNIC Rio):** Rio de Janeiro, 30 ago. 2023. Notícias. Disponível: https://brasil.un.org/pt-br/244006-campanha-celebra-os-75-anos-da-declara%C3%A7%C3%A3o-universal-dos-direitos-humanos. Acesso em: 23 mai. 2025.

RIVERA, Laura G. de. **Esclavos del algoritmo**: Manual de resistencia em la era de la inteligência artificial. Barcelona: Penguin Random House Grupo Editorial, 2025.

ROCHA, Cármen Lúcia Antunes. O princípio da dignidade da pessoa humana e a exclusão social. **Jurisprudência Catarinense**, Florianópolis, Brasil, v. 35, n. 117, p. 71-107, abr./jun. 2009. Disponível em:

https://bdjur.stj.jus.br/server/api/core/bitstreams/6bc3ff89-a160-459f-8475-22ed060d4373/content. Acesso em: 04 ago. 2025.

SANTIAGO, Abinoan. 'Me urinei de medo ao ser levada no camburão da PM por erro de IA'. **Uol Tilt**, Florianópolis, 15 nov. 2023. Inteligência Artificial. Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2023/11/15/erro-camera-reconhecimento-facial.htm. Acesso em: 10 jun. 2025.

SARLET, Ingo Wolfgang. **Dignidade da pessoa humana e direitos fundamentais na Constituição Federal de 1988**. 9. ed. rev. atual. Porto Alegre: Livraria do Advogado Editora, 2011.

SARLIN, Jon. EUA: Polícia prende inocente a partir de sistema de reconhecimento facial: Embora a tecnologia de reconhecimento facial tenha se tornado mais precisa, pesquisa mostrou que ela é suscetível a erros com pessoas de pele mais escura. **CNN Brasil**, [*S.l.*], 03 mai. 2021. Internacional. Disponível em: https://www.cnnbrasil.com.br/internacional/sistema-de-reconhecimento-facial-enviou-este-homem-inocente-para-a-prisao/. Acesso em: 10 jun. 2025.

SILVA, Tarcízio. **Racismo algorítmico**: inteligência artificial e discriminação nas redes sociais. São Paulo: Edições Sesc São Paulo, 2022.

TAULLI, Tom. **Introdução à Inteligência Artificial**: Uma abordagem não técnica. São Paulo: Apress Novatec, 2020.

UNESCO (2024). **Challenging systematic prejudices: an Investigation into Gender Bias in Large Language Models**. Paris: UNESCO, 2024. *E-book*. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000388971. Acesso em: 10 jul. 2025.

ZWART, Hans de. Racist technology in Action: AI detection of emotion rates Black basketball players as 'angrier' than their White counterparts. **Racism and Technology Center**, [*S.l.*], 21 jul. 2024. Newsletter 46. Disponível em:
https://racismandtechnology.center/2024/07/21/racist-technology-in-action-ai-detection-of-emotion-rates-black-basketball-players-as-angrier-than-their-white-counterparts/. Acesso em: 30 jul. 2025.