

# INTELIGÊNCIA ARTIFICIAL, RACISMO E EDUCAÇÃO: AOS PASSOS DOS LIAMES E PORMENORES

Jonathan Braz de Souza  
Universidade Estadual Paulista, *Campus Marília*

Camila Rodrigues Batista Neta  
Universidade Estadual Paulista, *Campus Marília*

## RESUMO

O seguinte trabalho tem por objetivo realizar um panorama acerca da confluência entre Inteligência Artificial, educação e racismo. Através de obras com teor relevante para a temática, lança-se luz para uma problemática pouco explorada, em comparação às questões envoltas nos benefícios da IA. Nesse sentido, em primeiro momento, buscamos traçar um resgate histórico de momentos importantes para a Inteligência Artificial. Posteriormente, afinamos nosso trabalho acerca de conceitos-chave sobre a IA Generativa e, por fim, problematizamos o caráter “imparcial” e “neutro” dos algoritmos e o seu viés racista ao que tange questões educacionais. Por fim, problematizar tais temáticas é se preocupar com uma prática tecnológica inclusiva, seja na adoção de políticas públicas ou até mesmo desses recursos dentro de instituições de ensino. Dessa maneira, questionar o viés do algoritmo é um modo de denunciar a maneira como essa tecnologia se desenvolveu e seus limites para uma sociedade plural e democrática.

**Palavras-chave:** Inteligência Artificial; Racismo; Educação.

## ABSTRACT

The following paper aims to provide an overview of the confluence between Artificial Intelligence and racism. Through works relevant to the topic, we shed light on a problem that has been little explored, compared to the questions surrounding the benefits of AI. To this end, we first seek to trace a historical overview of important moments in Artificial Intelligence. Subsequently, we narrow our work to key concepts of Generative AI and, finally, we problematize the "impartial" and "neutral" nature of algorithms and their racist bias. Finally, problematizing these issues involves focusing on inclusive technological practices, whether in the adoption of public policies or even in the use of these resources within educational institutions. Thus, questioning algorithmic bias is a way to expose the way this technology has developed and its limitations for a pluralistic and democratic society.

**Keywords:** Artificial Intelligence; Racism; Education.

## Introdução

A inteligência artificial (IA) não é apenas uma tecnologia contemporânea nem um conjunto de técnicas isoladas: é um campo científico e político. Suas raízes vão das reflexões filosóficas e matemáticas do século XIX até as transformações sociais e econômicas do século XXI e que, de certa maneira, presenciamos cotidianamente nas páginas dos diversos noticiários presentes em nosso país. Ao permear questões históricas da IA, não buscamos apenas registrar uma cronologia de invenções técnicas, mas compreender concepções, como: “inteligência” e “autonomia” foram construídas em determinados sistemas. Este texto analisa, primeiramente, os principais momentos, debates e problemas que marcaram

o desenvolvimento da IA, posteriormente, o enfoque na IA Generativa e, por fim, busca-se relacionar questões raciais e educacionais, como o racismo algorítmico, a partir da emergência da IA e seus impactos no mundo contemporâneo. A partir de autores como Turing (1950); Russell e Norvig (2013); Noble (2018); Zuboff (2021); Boratto (2023); buscamos esmiuçar questões relacionadas à Inteligência Artificial, o racismo algorítmico e questões educacionais. Dessa maneira, a metodologia a ser utilizada será uma abordagem qualitativa, tendo como cerne o levantamento bibliográfico de livros, artigos, revistas e periódicos que retratam de maneira interdisciplinar o tema da inteligência artificial, o racismo algorítmico e suas implicações no campo educacional.

## **Justificativa**

A história da IA começa muito antes do termo "inteligência" ter sido cunhado. A imaginação técnica sobre máquinas pensantes repousa em dois vetores principais: a formalização lógica/matemática e o desenvolvimento de máquinas que realizam cálculos. Em termos históricos, no século XIX, expoentes como Ada Lovelace já suscitavam discussões sobre as capacidades das máquinas de computação. Lovelace, em suas famosas notas sobre a Máquina Analítica de Babbage, sugeriu que máquinas poderiam manipular símbolos segundo regras – posicionando a questão de que a “inteligência” poderia surgir da aplicação sistemática de procedimentos formais. Embora as máquinas da era vitoriana fossem mecânicas e limitadas, o enquadramento lógico-computacional que Ada e outros ajudaram a consolidar abriu espaço para pensar a mente humana em termos de manipulação simbólica.

No século XX, a lógica matemática e a ciência da computação ofereceram o arcabouço formal para a ideia de “mecanizar” o raciocínio: a noção de algoritmo. Essa noção foi propiciada pelo trabalho de logicistas como Hilbert e Gödel, e de Turing, comumente conhecido pela máquina de Turing. Esse último, em 1950, publicou o texto clássico "Computing Machinery and Intelligence" (1950), no qual formulou o famoso teste e discutiu objeções filosóficas sobre se as máquinas poderiam "pensar". O artigo de Turing instigou debates sobre comportamento observável versus estados internos, sugerindo uma mudança na pergunta “as máquinas pensam?” para uma pergunta operacional sobre desempenho comunicativo e funcional. O texto de Turing ainda é um marco para a história conceitual da IA e é citado como ponto de partida formal do campo da Inteligência Artificial.

Em aspectos terminológicos, a nomenclatura "Inteligência Artificial" foi formalmente introduzida na em um seminário em Dartmouth, no verão de 1956, redigida por John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon. Esses são somente alguns nomes, mas, de acordo com Russell e Norvig (2013) “[o] seminário de Dartmouth não trouxe nenhuma novidade, mas apresentou uns aos outros todos os personagens importantes da história” (p. 17). A proposta imaginava um programa de pesquisa que reunisse ideias de lógica, neurofisiologia e teoria da informação para explorar a

possibilidade de fazer uma máquina aprender com a experiência. O encontro de verão de Dartmouth, por mais que não seja propagado nas questões relacionadas à Inteligência Artificial, pode ser lembrado como o momentos de institucionalização do campo, pois, a partir dele, formaram-se laboratórios, programas de pesquisa e uma rede internacional de pesquisadores que passaram a trabalhar sob a bandeira da IA. Nos anos imediatamente posteriores, as linhas dominantes de pesquisa privilegiaram abordagens simbólicas, chamada de IA "simbólica". A ideia central era que o raciocínio humano podia ser modelado como a manipulação de símbolos e regras explícitas.

Ao mesmo tempo que a abordagem simbólica prosseguia, pesquisadores inspirados nas ciências biológicas propuseram modelos inspirados em neurônios: as redes neurais artificiais. Em 1958, Frank Rosenblatt descreveu o perceptron, mas o que é um perceptron? Em termos gerais, o perceptron é um modelo simples de neurônio artificial capaz de aprender a partir de sinais de entrada. Embora a trajetória da pesquisa em redes neurais sofreu um revés importante com a publicação de *Perceptrons* por Marvin Minsky e Seymour Papert. Minsky e Papert, de acordo com Russell e Norvig (2013), demonstraram limitações como excesso de promessas e a falta de escalabilidade prática de certas abordagens neurais. O resultado foi uma redução substancial de financiamento e interesse em redes neurais nas décadas seguintes; somado a exageros de promessas e a resultados práticos decepcionantes em algumas áreas de IA simbólica, isso contribuiu para o chamado “primeiro inverno da IA”. O inverno da IA pode ser caracterizado como a redução de financiamento, desilusão institucional e crítica intelectual são momentos importantes da história porque mostram que a trajetória tecnológica não é linear. Eles sublinham que progressos científicos dependem de recursos, de expectativas públicas e de condições institucionais.

Nas décadas seguintes, a combinação de técnicas de aprendizado profundo, disponibilidade massiva de dados (big data) e computadores mais potentes (GPUs) conduziram uma nova “primavera” da IA. Diferentemente do inverno da IA, a primavera pode ser caracterizada pelo ressurgimento de temáticas relacionadas à IA, financiamento por parte do governo e políticas de incentivo ao campo, como atualmente, principalmente após o lançamento do ChatGPT pela empresa OpenAI. Do ponto de vista sócio-histórico, a IA sempre gerou debates sobre seu impacto no trabalho, nas estruturas de poder e na distribuição de riqueza. Desde os primeiros prognósticos de automação até os debates contemporâneos sobre deslocamento de empregos, as transformações tecnológicas suscitam promessas de produtividade, nem sempre alcançáveis, e receios quanto à concentração de benefícios. Autores críticos como Cathy O’Neil chamam a atenção para efeitos perversos de modelos algorítmicos aplicados em grande escala: quando sistemas de predição influenciam decisões de crédito, educação, contratação e justiça, eles podem reproduzir ou intensificar desigualdades sociais. Seu livro, *Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia* (2020), popularizou a noção de que modelos opacos e aplicados em escala podem causar danos sistêmicos a populações vulneráveis, chamando atenção para a necessidade de auditoria, transparência e regulação de algoritmos.

A história desses impactos pode permear sistemas de pontuação, modelos de previsão e políticas automatizadas que produziram decisões discriminatórias. Além do emprego, há questões relativas à vigilância e ao capital de dados. Plataformas digitais, governos e corporações construíram infraestruturas de coleta e processamento de dados que alimentam modelos de IA; isso levanta questões sobre consentimento, privacidade, soberania de dados e formas de poder digital. Historicamente, a capacidade técnica de prever e influenciar comportamentos – por publicidade direcionada, ranking e recomendações – tem efeitos políticos e culturais profundos (por exemplo, influências em eleições, polarização, modelos de moderação de conteúdo). Essas dinâmicas mostram como a IA se torna um vetor de reconfiguração institucional e geopolítica.

Nos anos 2010-2024, com o crescimento exponencial da adoção de sistemas de IA, emergiu uma literatura robusta sobre ética, responsabilidade e governança. Instituições acadêmicas, empresas e governos passaram a refletir sobre princípios, diretrizes e propostas regulatórias para uma "IA ética". Entre as discussões centrais estão: justiça e não discriminação; transparência e explicabilidade; responsabilidade e mecanismos de recurso; segurança e robustez; privacidade e proteção de dados; e implicações laborais. A discussão sobre regulação tornou-se importante em virtude de episódios públicos de falhas algorítmicas ou vazamentos de dados e de crescente pressão social. Movimentos por auditoria independente de algoritmos, por frameworks de “direitos digitais” e por mecanismos legais de responsabilização têm ganhado força. Essas iniciativas se inserem numa tradição histórica de tentativa de domesticar tecnologias com potencial disruptivo.

Em contraponto, a história cultural da IA é plural em seu conteúdo de entretenimento: romances, filmes e debates públicos moldaram expectativas e medos sobre máquinas inteligentes. Diversas são as obras que podem ser mencionadas, como: *Ex\_machina*: instinto artificial; *Chappie*; *Eu, Robô*; *A.I. - Inteligência Artificial*; *Inteligência Artificial: ascensão das máquinas*; *Her*; *Blade Runner*, o caçador de andróides; *Matrix* e entre outros. Desde obras clássicas de ficção científica, como citado anteriormente, até representações midiáticas contemporâneas, as imagens de IA, como: robôs benevolentes, inteligências hostis, entidades superinteligentes, influenciam agendas de pesquisa e receptividade pública. Além disso, acordos políticos e narrativas geopolíticas (por exemplo, a corrida por liderança em IA entre Estados-nação) transformaram a IA em arena de políticas industriais, investimentos estratégicos e disputas por padrões. A história recente mostra como as tecnologias digitais passam a ser medidas de poder internacional e de capacidade econômica, como constatamos no noticiário sobre os conflitos de grandes potências devido a matérias-primas para o desenvolvimento dessas tecnologias.

Um ponto que atravessa a história da IA é a ambiguidade do termo “inteligência”. Em contextos técnicos, a inteligência denota desempenho em tarefas específicas como classificar imagens e/ou traduzir frases. Em contextos filosóficos e populares, tende a implicar compreensão, consciência ou agência. Essa ambivalência gera mal-entendidos — por exemplo, a atribuição de capacidades “humanas” a sistemas que, na prática, apenas otimizaram uma função de perda. Historicamente, esse equívoco tem alimentado

tanto expectativas infundadas quanto alarmistas. A reflexão histórica aponta para a necessidade de desagregar “inteligência” em sub-habilidades (percepção, raciocínio, planejamento, interação social) e de avaliar sistemas em suas especificidades.

## **Inteligência Artificial Generativa: Um passo adiante na esfera da IA**

No cerne da Inteligência Artificial o aspecto generativo se desenvolveu como uma área que reluz um fascínio, apesar dos debates envolvidos na utilização dessa tecnologia. A IA Generativa se refere a sistemas capazes de criar novos conteúdos, como textos, imagens, músicas, códigos ou até vídeos, a partir de padrões aprendidos em grandes volumes de dados ou big data. Em vez de apenas seguir instruções rígidas ou realizar cálculos pré-definidos, a IA generativa consegue produzir algo novo, original e muitas vezes surpreendente, imitando de forma impressionante certos aspectos da criatividade humana. É como se, a partir de tudo que já viu e aprendeu, ela fosse capaz de “imaginar” novas possibilidades. Para entender melhor como isso funciona, precisamos primeiro compreender alguns conceitos fundamentais sobre como as IAs aprendem. As técnicas mais conhecidas envolvem aprendizado supervisionado, aprendizado não supervisionado e, de forma mais ampla, machine learning. Faremos um panorama acerca desses três conceitos, por serem relevantes para a compreensão de questões pormenores na área.

De maneira mais descritiva, esse método consiste em a máquina aplicar uma fórmula com múltiplos parâmetros, sendo que o aprendizado ocorre ao substituir valores iniciais aleatórios nessa fórmula e, posteriormente, tentar realizar primeiras previsões assertivas a partir de exemplos fornecidos. Nesse processo, a máquina consegue identificar as previsões equivocadas, corrigindo as predições com base nos exemplos previamente armazenados. A repetição desse método milhares de vezes possibilita que a máquina ajuste essas previsões a um referencial, denominado modelo de treinamento. (Boratto, 2023, p. 25)

O machine learning ou aprendizado de máquina é um campo da inteligência artificial que se concentra em ensinar computadores a aprender com dados. Em vez de programar manualmente todas as regras que um sistema deve seguir, os desenvolvedores fornecem grandes quantidades de exemplos, e o sistema “descobre” sozinho os padrões que ligam esses dados a certos resultados. Isso torna possível criar algoritmos capazes de se adaptar a novas situações sem precisar ser reprogramados do zero. Podemos pensar no machine learning como a “escola” onde a IA vai para aprender: lá, ela pode receber aulas muito guiadas ou aulas mais livres, dependendo do tipo de aprendizado. Já no aprendizado supervisionado, o sistema é treinado com exemplos que já têm a resposta correta. Imagine que você quer ensinar uma IA a reconhecer fotos de gatos. Você mostraria milhares de imagens, cada uma com a etiqueta “gato” ou “não gato”. Assim, o algoritmo vai associando características das imagens à presença ou ausência de um gato. É “supervisionado” porque, a cada passo, ele recebe um feedback claro: “Acertou” ou “Errou”.

O aprendizado supervisionado consiste em todo tipo de aprendizado que se fundamenta na existência de um modelo que avalia as respostas aos estímulos de

entrada. Já o aprendizado não supervisionado caracteriza-se pela inexistência de um modelo que avalia as respostas aos estímulos de entrada (Russel; Norvig, 2013).

É como um professor corrigindo todas as tarefas dos alunos, dizendo exatamente onde estão os erros. Esse tipo de aprendizado é muito eficiente para tarefas como reconhecimento de voz, classificação de e-mails como spam ou não spam, e até diagnósticos médicos baseados em exames de imagem. Por fim, o aprendizado não supervisionado é mais livre: o sistema recebe apenas os dados, sem respostas prontas. Ele precisa encontrar sozinho padrões, agrupamentos ou estruturas ocultas nessas informações. Voltando ao exemplo dos gatos, utilizado anteriormente, nesse caso você entregaria um conjunto enorme de fotos, mas sem dizer quais têm gatos e quais não têm. O algoritmo, então, poderia agrupar imagens parecidas e, quem sabe, descobrir por conta própria que algumas dessas imagens têm elementos comuns, como orelhas pontudas, bigode ou olhos brilhantes. Esse tipo de aprendizado é usado para análise de comportamento de clientes, segmentação de mercado e até compressão de dados.

A IA generativa utiliza elementos tanto do aprendizado supervisionado quanto do não supervisionado, mas vai além: ela não apenas reconhece padrões, ela cria algo novo a partir deles. Por exemplo, modelos de linguagem como o GPT (Generative Pre-trained Transformer) são treinados com imensos volumes de texto da internet, livros, artigos e outros conteúdos. Esse treinamento pode incluir fases supervisionadas (onde o modelo aprende com correções humanas) e fases não supervisionadas (onde o modelo analisa e prevê palavras com base no contexto). Depois, quando pedimos para escrever algo ou criar uma frase, ele combina tudo que “aprendeu” para gerar um resultado original. Além disso, há uma dificuldade em definir o que acontece nesse processo de “aprendizado” dentro dos diferentes tipos de modelos. O teor generativo possui a capacidade de gerar conteúdo abrindo um leque de aplicações e possibilidades: escritores podem usar IA para criar rascunhos de histórias, designers podem gerar ideias visuais, músicos podem compor novas melodias com auxílio de algoritmos, e até engenheiros podem projetar peças com formatos otimizados por IA. No entanto, junto com essas oportunidades, surgem também questões éticas, como o uso indevido de conteúdo protegido por direitos autorais ou a possibilidade de criar desinformação convincente. A própria IA do Google, Veo3, foi uma febre, devido ao caráter realista dos vídeos produzidos, acendendo o alerta para a disseminação e divulgação de notícias falsas.

Um aspecto interessante da IA generativa é que ela aprende com probabilidades. Quando vai gerar uma frase, por exemplo, o modelo não simplesmente “escolhe” a palavra mais comum – ele considera várias opções prováveis e pode escolher caminhos menos óbvios para criar variedade e naturalidade no texto. É por isso que, ao pedir para escrever uma história de ficção, a IA pode inventar personagens, diálogos e cenários que nunca apareceram antes, mesmo que se inspire fortemente em padrões já existentes. Essa combinação de técnicas e a capacidade de produzir conteúdos inéditos estão transformando diversas áreas. Apesar das possibilidades, essa “criatividade” não é a mesma que a humana. A IA não cria porque tem emoções ou intenções; ela cria porque combina informações de forma

estatística e rápida. (Goodfellow; Bengio; Courville, 2016). Todavia, para quem observa o resultado final, a diferença pode ser sutil – e é aí que surge tanto o encanto quanto a preocupação. Afinal, se uma máquina consegue escrever como um ser humano, quem garante que saberemos diferenciar um conteúdo verdadeiro de um inventado?

### **Inteligência Artificial e Racismo: Quando o prompt “julga” a cor da pele e suas implicações educacionais**

Apesar da questão levantada na seção anterior, algo pouco mencionado quando falamos sobre Inteligência Artificial é o seu cunho discriminatório. Embora a IA seja frequentemente apresentada como um conjunto de ferramentas objetivas e imparciais, isentas de preconceitos humanos, na prática essas tecnologias refletem e, por vezes, amplificam desigualdades já presentes na sociedade, como afirma Tarcízio Silva (2022), em sua obra *Racismo algorítmico: Inteligência artificial e discriminação nas redes sociais*. Esse fenômeno decorre do fato de que sistemas de IA são treinados a partir de dados históricos, linguísticos e visuais que carregam as marcas de contextos sociais desiguais. Assim, ao invés de corrigir discriminações, esses sistemas podem reproduzi-las de forma automatizada e invisível, dando-lhes um teor de neutralidade técnica. Um dos casos mais emblemáticos que expõe essa relação ocorreu com o software de reconhecimento facial utilizado por órgãos policiais nos Estados Unidos. Em 2018, um estudo do *MIT Media Lab*, conduzido por Joy Buolamwini, revelou que três dos sistemas comerciais mais populares – desenvolvidos por Microsoft, IBM e Face+ – apresentavam taxas de erro muito mais altas para identificar o gênero de mulheres negras em comparação com homens brancos. Enquanto a taxa de acerto para homens brancos superava 99%, para mulheres negras o erro chegava a 35%. Esse resultado não apenas mostrou um viés técnico, mas também apontou para uma cadeia de decisões que privilegiava a coleta de imagens de pessoas brancas, especialmente homens, durante o treinamento dos algoritmos. O problema do viés racial em IA também se manifesta em sistemas de previsão criminal, como o COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), utilizado em vários estados norte-americanos para avaliar o risco de reincidência de indivíduos. Uma investigação do *ProPublica* demonstrou que o sistema atribuía notas de risco mais altas a réus negros em comparação a réus brancos com histórico criminal semelhante, mesmo quando os primeiros não reincidiam. O caso é particularmente grave porque, em decisões judiciais, tais pontuações podem influenciar penas, liberdade condicional e até a prisão preventiva, ampliando a seletividade penal que historicamente atinge comunidades negras.

É importante notar que esse viés não é um defeito meramente técnico, mas o reflexo de um processo histórico. Nos Estados Unidos, por exemplo, as bases de dados criminais utilizadas para treinar sistemas como o COMPAS foram construídas a partir de décadas de policiamento mais intenso em bairros majoritariamente afroamericanos e latinos. Isso gerou um acúmulo de registros policiais que não necessariamente refletem maior criminalidade nessas comunidades. A IA, nesse sentido, funciona como um espelho de uma estrutura racista pré-existente, apenas transformando-a em números e probabilidades aparentemente “objetivas” e “imparciais”. Outro campo em que o racismo se manifesta de outra forma é o da moderação automatizada de conteúdo em redes sociais. Sistemas treinados para detectar discurso de ódio frequentemente apresentam resultados enviesados contra usuários negros. Safiya Umoja Noble, em seu livro *Algoritmos da Opressão: como o Google fomenta e lucra com o racismo* (2018), documenta como buscas no Google no início dos anos 2010 retornavam resultados sexualizados ou depreciativos quando se pesquisavam termos relacionados a mulheres negras. Embora o Google tenha posteriormente ajustado seus algoritmos, o episódio revela como classificadores automáticos, baseados em dados da própria internet, podem reforçar estereótipos racistas presentes na produção e no consumo de conteúdo online.

No Brasil, a discussão sobre IA, educação e racismo começa a ganhar mais espaço. Casos como o uso de sistemas de reconhecimento facial pela polícia em grandes eventos e no carnaval levantam preocupações. Em 2019, no Rio de Janeiro, uma mulher negra foi presa erroneamente após ser identificada por um desses sistemas, quando, na realidade, não se tratava da pessoa procurada. O caso ganhou repercussão e foi denunciado por organizações que apontaram a falta de transparência e os riscos de criminalização desproporcional à população negra. A questão central é que: os algoritmos não operam no vácuo, eles são construídos por pessoas e alimentados por dados que carregam vieses sociais. Mesmo quando programadores não têm a intenção de discriminar, a ausência de diversidade nas equipes de desenvolvimento contribui para a reprodução de visões limitadas do mundo. Um levantamento da *AI Now Institute* mostrou que, nas principais empresas de tecnologia, menos de 5% dos funcionários da área de IA se identificavam como negros. Essa sub-representação dificulta a identificação e correção de vieses que atingem diretamente minorias raciais. Para combater o racismo algorítmico, alguns pesquisadores e ativistas têm defendido abordagens como a “auditoria algorítmica”, que consiste em submeter sistemas de IA a testes independentes que revelem seus padrões discriminatórios antes da implementação em larga escala. Além disso, propõe-se o desenvolvimento de conjuntos de dados mais representativos e o estabelecimento de regulamentações que obriguem empresas e governos a serem transparentes sobre como e com quais dados seus modelos são treinados. A análise sobre IA e racismo também exige uma compreensão de como o poder se distribui no mundo digital. A coleta massiva de dados, característica da economia da vigilância descrita por Shoshana Zuboff (2021), não é apenas uma questão de privacidade, mas também de controle social. Quando esses dados são usados para categorizar, prever e direcionar comportamentos de forma diferenciada segundo marcadores raciais, reforça-se um ciclo de desigualdade que, no caso de populações negras e indígenas, já está profundamente enraizado na história colonial.

O desafio, portanto, não é apenas técnico, mas ético e político e, também, educacional. A utilização de ferramentas de Inteligência Artificial no campo educativo tem sido uma constante. Instituições privadas argumentam a partir desse recurso como uma proposta de diferenciação para o processo de aprendizagem. Diversos cursos, atualmente, possuem chatbots de inteligência artificial como atrativo e como um “facilitador de processos educacionais”. A aplicação de uma IA nesse campo implica repensar a própria lógica educacional, assim a maneira como a produção tecnológica é realizada, tendo em mente seu modo de desenvolvimento, incluindo mais vozes e experiências diversas nas etapas de concepção, implementação e avaliação. Implica, ainda, em questionar usos que, mesmo tecnicamente possíveis, geram impactos sociais desiguais. Ao reconhecermos que algoritmos são produtos culturais, e não meramente matemáticos, podemos começar a pensar em caminhos para construir tecnologias que não apenas refletem, mas também ajudem a transformar, uma realidade ainda marcada por estruturas e indivíduos racistas.

Nesse sentido, os aspectos do racismo estrutural, nos é colocado como um desafio triplo: sua superação em sociedade – já admitindo o árduo exercício que o envolve – e trazer luz para esse aspecto quando se refere ao mundo das IA.

Com isso, utilizaremos de duas problemáticas apontadas por Silva (2020) para especificar alguns dos vários desafios quando se fala sobre a invisibilidade negra em redes sociais, incluindo o modo como esse sujeito é percebido. Se antes mencionamos os aspectos de valores relacionados à criminalidade, agora, inclui-se também aspectos mais relacionados à individualidade possível e a questão de gênero; no primeiro, existe a questão de que o “eu” já vêm percebido por vários estigmas e que resta a IA, encaixá-lo nisso, já no segundo, envolve a errônea assimilação de que a mulher negra está sempre envolta por um fetiche exagerado por mais ingênua que seja o seu fazer. Nas palavras do autor:

A negação de responsabilidade ocorre na tentativa de enquadramento da questão como externa a sua ferramenta, como apenas uma reprodução de estereótipos existentes. E, usando a estratégia da suposta neutralidade racial que nega tanto a realidade factual brasileira quanto a de seu país-sede, os Estados Unidos, alega, como justificativa, que o fato poderia acontecer com qualquer grupo. [...] Em outro caso brasileiro, mulheres descobriram que buscas como “mulher negra dando aula” resultam também em conteúdo pornográfico. Ao responder ao portal educacional que cobriu o problema, a empresa emitiu uma nota dizendo: “o conjunto de resultados para o termo mencionado não está à altura desse princípio [intenção dos usuários] e pedimos desculpas àqueles que se sentiram impactados ou ofendidos”. Ou seja, a empresa se esquivava e delega o ônus da ofensa às vítimas. (Silva, 2020. p. 83-84)

A passagem acima revela uma problemática que atravessa a discussão acerca das tecnologias digitais. A negação de responsabilidade, por parte das grandes empresas de tecnologia ou big techs, diante dos efeitos discriminatórios e ofensivos que suas ferramentas podem provocar. Tal negação, expressa um enquadramento das questões relacionadas ao preconceito como mera reprodução de estereótipos realizados pelas máquinas. Ao alegar que o racismo presente nos resultados ou decisões algorítmicas é apenas uma simples reprodução automática das desigualdades sociais, essas empresas tentam se

posicionar como atores distantes das consequências de suas criações ou ao menos dos modelos que desenvolvem, recusando assim, uma dimensão política e social dos seus próprios produtos tecnológicos. Tal estratégia de neutralidade é uma ferramenta com um alto poder e influência de invisibilizar a realidade que estrutura as diversas camadas da sociedade. A justificativa de que falhas algorítmicas podem ocorrer com “qualquer grupo” é uma tentativa de uniformizar experiências distintas, apagando as especificidades dos impactos raciais e das desigualdades que atravessam as sociedades. Isso equivaleria a dizer que as máquinas, pelo seu teor tecnológico, estariam acima dos conflitos sociais e dos processos históricos, quando na verdade, elas são construídas justamente a partir desses contextos, por pessoas que vivem e (re)produzem suas condições culturais, políticas e econômicas.

## Considerações finais

Frente ao exposto, embora essas situações aqui descritas pareçam nos encurralar, e, de fato, trata-se de uma realidade perturbadora, uma das possíveis soluções seriam uma maior pluralidade de bancos de dados, sem que dependessem somente dos usuários apontarem tendências questionáveis sobre as informações que procuram ou que obtêm – de uma forma que parece quase natural ao algoritmo, seria o reconhecimento por parte de grandes empresas. É a partir do reconhecimento dos muitos estudos que têm sido feitos, somados às queixas dos internautas, que as informações obtidas pelas IAs, deveriam partir de fontes melhores. Não porque aquilo está dado ou somente porque há muitas pessoas que irão concordar com aquele dado. Afinal, como mencionamos ainda há pouco, trata-se de uma questão estrutural e que vem se mostrando cada vez mais evidente também no escopo das *Machine Learning*.

À luz desses apontamentos, o que fizemos ao longo deste texto foi questionarmos sobre os aspectos do que se é entendido por inteligência artificial, de maneira a promover um panorama histórico sobre palavras-chave dessa área do conhecimento, como machine learning, aprendizado supervisionado e aprendizado não supervisionado, assim como Inteligência Artificial Generativa. Além disso, tal resgate propicia compreender os pormenores desse campo do conhecimento e entender os limites éticos dessa ferramenta. A partir de um outro olhar, tensionamos o modo como esses algoritmos foram desenvolvidos e questionamos suas limitações e exclusões de determinadas minorias, bem como o reconhecimento disso perpassa por questões que são maiores do que as que pretendemos relatar.

## REFERÊNCIAS

- BORATTO, Murilo do Carmo. Inteligência Artificial: Breve histórico, conceitos e reflexões. *In*: ALVES, Lynn. **Inteligência Artificial e educação**: refletindo sobre os desafios contemporâneos. Salvador: EDUFBA, 2023, p. 21-32.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA: MIT Press, 2016.

NOBLE, Safiya Umoja. **Algoritmos da Opressão**: como o Google fomenta e lucra com o racismo. Trad: Felipe Damorim. Santo André: Editoria Rua do Sabão, 2018.

O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. Tradução: Rafael Abraham. 1º edição. São Paulo: Editora Rua do Sabão, 2020.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. Tradução: Regina Célia Simille. Rio de Janeiro: Elsevier, 2013.

TURING, Alan. **Computing Machinery and Intelligence**. Mind 49: 433-460, 1950.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância**: a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2021.