

Vieses algorítmicos e o imperativo de enfrentar o racismo estrutural no Brasil

Autoras: Isabela da Silva e Cíntia Rosa Pereira de Lima

Vieses algorítmicos e o imperativo de enfrentar o racismo estrutural no Brasil

Quando no verão de 1955, John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon compartilharam com o mundo a ideia de que qualquer aspecto da aprendizagem ou atividade humana poderiam ser descritos de forma precisa, permitindo que uma máquina fosse construída para simulá-laⁱ, cunhando-lhe o termo “inteligência artificial” (“IA”), despertou-se na sociedade uma legítima expectativa - tão promissora quanto incerta.

E, embora o termo continue a chamar atenção mais de meio século depois de seu primeiro uso, há controvérsias quanto a sua proposição. Isso porque além de não ser consistente fazer uma analogia entre a autonomia de robôs e a racionalidade humana - dotado de ética, moral e valores que refletem não apenas a personalidade individual, mas a compreensão coletiva da sociedade -, a própria intenção de associar o termo autonomia (*lato sensu*) à máquina pode não ser ideal, tendo em vista que, muitas vezes, o que se chama autonomia na verdade representa uma “variedade de combinações algorítmicas viabilizadas por um software”ⁱⁱ. Portanto, trata-se muito mais de uma reprodução de certos tipos de comando do que, de fato, uma escolha, realizada de forma livre e consciente.

As divergências terminológicas e conceituais quanto à IA refletem o fato de se tratar de um campo em aberto, com significativos avanços nas últimas décadas - especialmente em função do advento do *deep learning*ⁱⁱⁱ -, mas que ainda está longe de suas potencialidades. Nesse sentido, Kai-Fu Lee^{iv} defende que é possível explicar a evolução da IA a partir de “quatro ondas”:

(i) da internet, por exemplo, por meio do uso de algoritmos de IA que aprendem nossas preferências e passam a nos recomendar conteúdos;

(ii) de negócios, como utilizado por bancos e seguradoras que realizam mineração de dados^v;

(iii) de percepção, por exemplo, o aumento do uso de sensores inteligentes; e

(iv) de autonomia, referindo-se à união das três ondas anteriores, permitindo a fusão das capacidades otimização quanto aos grandes conjuntos de dados complexos e percepção sensorial do ambiente externo, por exemplo, os carros autônomos.

Entretanto, o regulador precisa estar atento aos riscos que a IA apresenta. Um caso emblemático, julgado em julho de 2022, após o Departamento de Justiça dos Estados Unidos ter reconhecido o uso de critérios raciais, étnicos, religiosos, sexuais e de renda para definir anúncios de moradia para usuários do *Facebook*, atentando contra o *Fair Housing Act* (Lei de Habitação Justa).^{vi}

Outro caso que vale a pena ser mencionado é o escândalo envolvendo a *Amazon* após o preterimento de candidaturas de mulheres em detrimento a de homens por um sistema de IA para recrutamento, implicando em discriminação de gênero.^{vii}

Nota-se, ainda, o fato de influenciadores negros terem menor “entrega” e, portanto, engajamento de seus conteúdos em redes sociais^{viii}, diminuindo a possibilidade deles se tornarem influenciadores competitivos no mercado digital.

Isso ocorre porque existem “vieses algorítmicos”,^{ix} termo cunhado para descrever as distorções e/ou injustiças resultantes das decisões automatizadas feitas por algoritmos de inteligência artificial^x. A evolução do conceito de viés algorítmico para o direito internacional tem sido gradual, mas significativa. Organizações internacionais, como a União Europeia e as Nações Unidas, têm reconhecido a importância de abordar vieses algorítmicos em suas políticas e diretrizes. Por exemplo, o Regulamento Geral de Proteção de Dados (“GDPR”) da UE inclui disposições que abordam a transparência e responsabilidade no uso de algoritmos de IA que impactam os direitos e liberdades das pessoas^{xi}.

Além disso, iniciativas globais, como a “Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial” (2018)^{xii}, destacam a necessidade de garantir a equidade e a não discriminação em sistemas automatizados. No entanto, apesar do reconhecimento crescente, desafios persistem na aplicação efetiva do direito internacional para mitigar vieses algorítmicos, pois sua complexidade técnica e a falta de transparência em sua implementação continuam a ser obstáculos importantes.

O grande desafio a ser contornado em busca da transparência da IA é a *black box*, expressão usada para descrever a dificuldade de explicabilidade (“XIA”)^{xiii} e auditabilidade^{xiv} de sistemas de IA^{xv}.

Ainda que se possa estar longe de uma saída conclusiva quanto aos vieses algorítmicos, alguns Estados tentam criar seus próprios instrumentos jurídicos para prevenir e mitigar a ocorrência de discriminações e injustiças eventualmente

ocasionadas pelos sistemas de IA. É o caso dos Estados Unidos e da União Europeia, por exemplo, respectivamente: o *Algorithmic Accountability Act* (“AAA”)^{xvi}; o *Artificial Intelligence Act* (“AI Act”), primeiro marco da IA do mundo.^{xvii}

No Brasil, o Projeto de Lei nº 2.338/2023 pretende regular o desenvolvimento e o uso da IA no país, sendo que o rol dos princípios está no art. 3º, dentre eles, destaca-se o inciso IV, o princípio da não discriminação; e, o inciso VI, o princípio da transparência, explicabilidade, inteligibilidade e auditabilidade.

No caso estadunidense, por exemplo, outros documentos tiveram - e ainda têm - papel importante para orientação de sua população, como o *National Defense Authorization Act for Fiscal Year 2021* (“NDAA 2021”) e o *Federal Trade Commission Act* (2023).

O NDAA 2021 foi responsável pelo estabelecimento de uma Comissão Nacional de Inteligência Artificial (*National Security Commission on Artificial Intelligence* - “NSCAI”). A seção 5103 traz diversos recortes sobre o tema, como a defesa da diversidade e da inclusão, o impacto à saúde mental em programas de natalidade e combate às “barreiras raciais” (que deveria ser melhor definido como racismo algorítmico^{xviii}) e socioeconômicas, a luta contra a discriminação em processos seletivos, a necessidade de produção de relatórios como medida para avaliação de impacto, entre outros.

Após a divulgação do NDAA 2021, também foi publicado a *Federal Trade Commission Act* (“FTC Act”), em 2023, que atribuiu um papel crucial da FTC na aplicação das leis antitruste, na proteção de dados e da privacidade dos consumidores e na promoção da concorrência justa nos mercados. Em 2023, o FTC aprovou uma resolução que autoriza o uso do processo compulsório em investigações não públicas envolvendo produtos e serviços que usam ou afirmam ser produzidos usando inteligência artificial ou que alegam detectar seu uso^{xix}.

Além destes, o *Algorithmic Accountability Act*,^{xx} atualizado em 2023^{xxi}, trata-se de um marco na discussão sobre vieses algorítmicos nos EUA e entre suas disposições estão a previsão de:

(i) avaliação de impacto de equidade, que obriga as empresas que desenvolvem algoritmos de IA a conduzirem avaliações de impacto de equidade para identificar e mitigar vieses algorítmicos;

(ii) transparência e divulgação: exige transparência na implementação e funcionamento dos algoritmos, incluindo a divulgação de dados de treinamento,

métodos de seleção e peso de variáveis, e resultados das avaliações de impacto de equidade; e

(iii) responsabilidade e prestação de contas, que estabelece responsabilidade legal para as empresas em caso de discriminação ou danos causados por vieses algorítmicos, com possibilidade de penalidades significativas em caso de não conformidade.

Por outro lado, a União Europeia tem tratado o tema de maneira diferente. O "Regulamento Geral de Proteção de Dados" (GDPR)^{xxii}, que entrou em vigor em maio de 2016, menciona o combate à discriminação contra vieses algorítmicos. Embora o GDPR não se concentre exclusivamente em vieses algorítmicos, ele estabelece disposições importantes relacionadas à transparência, responsabilidade e proteção dos direitos individuais no processamento de dados, incluindo aqueles realizados por algoritmos de IA. Alguns artigos podem ser destacados, por exemplo, o art. 10º, que trata da governança de dados; o artigo 14º, a respeito da supervisão humana; e o art. 70, que contempla questões de *enforcement* ao atribuir competências às autoridades nacionais.

Após 2018, uma série de diretrizes e orientações jurídicas foram publicadas, revelando seu caráter não vinculante, típica de *soft law*, tais como os documentos *Ethics Guidelines for Trustworthy AI* (2019)^{xxiii}; *Policy and investment recommendations for a Trustworthy AI* (2019)^{xxiv} e *White Paper on Artificial Intelligence: A European approach to excellence and trust* (2020)^{xxv}. Com uma abordagem principiológica, a Comissão Europeia buscou sedimentar as bases para o fortalecimento de sua "visão *Good AI Society*"^{xxvi}, mas, mais do que isso, buscou fazer com que a seriedade com que discute o tema da regulação de inteligência artificial ultrapassasse suas fronteiras. A partir disso, o efeito bruxelas^{xxvii} observado nos últimos anos quanto às legislações da UE passa a ganhar novos desdobramentos, dado que não se busca apenas o prestígio internacional por ser pioneira quanto à regulação da tecnologia mais disruptiva da história, mas influenciar o jogo político-econômico de forma a garantir sua posição da divisão internacional informacional^{xxviii} e, com isso, manter o *status* de potência sem que necessariamente domine o mercado a ponto de disputá-lo com grandes *players*, como EUA e China.

A respeito do *AI Act*, cabe menção a algumas de suas previsões mais importantes:

(i) avaliação e gestão de risco, especialmente para modelos de alto risco, os quais devem passar por avaliações de versões, análise e mitigação de riscos sistêmicos, testes de adversidades, reportar incidentes graves à Comissão Europeia, garantir segurança cibernética e eficiência energética;

(ii) uso de vigilância biométrica por governos, que a lei prevê que tenha seu uso restrito a situações específicas, como a busca de vítimas de determinados crimes, prevenção de ameaças terroristas e busca de suspeitos de crimes graves;

(iii) sistemas de categorização biométrica, havendo proibição de sistemas que utilizam características sensíveis (como crenças políticas, religiosas, orientação sexual, raça, entre outras) e

(iv) *scraping* indiscriminada de imagens faciais, havendo proibição à coleta não direcionada de imagens faciais da internet ou de câmeras de vigilância para criar bancos de dados de reconhecimento facial.

É notório que a forma apressada com que algumas discussões têm ocorrido para formulação da EBIA e também na CJSUBIA, alinhada a pouca capilarização desses debates na ponta, com a sociedade, podem significar a menor popularidade do tema de IA e sensibilização com questões importantes desse universo. Como exemplo pode-se dar destaque à urgente necessidade de letramento digital, em todos os níveis de ensino, ou mesmo a defesa de uma infraestrutura digital pública no país, o que não ocorrerá sem a devida mobilização da sociedade, mas igualmente do setor privado, público e acadêmico - num movimento semelhante ao realizado pela UE. Afinal, como o país espera ser capaz de enfrentar questões importantes como os vieses algorítmicos sem que antes consiga compreender as raízes do racismo estrutural que ainda corrói as instituições e o seio da sociedade brasileira?

Muito anterior aos vieses da máquina, os vieses e preconceitos já estão presentes no ser humano e suas criações reproduzem as próprias mazelas observadas em seus criadores. Nesse sentido, o sistema imperialista que “herdou” o racismo para manutenção do subdesenvolvimento, como defendido por Clóvis Moura^{xxix}, também tem deixado como herança para o atual ciclo econômico, da economia digital, este mesmo mecanismo de dinamização dos interesses de classes dominantes para obtenção de mais-valia. Com incontáveis dados pelo mundo, será possível ampliar de maneira exponencial comportamentos de violências e agressões no mundo real e virtual, retroalimentando-se e tendo como resultado o aprofundamento das desigualdades sociais pré-existentes.

Não à toa, a discriminação contra *influencers* negros é uma realidade que vem sendo denunciada, além da problematização do uso de sistemas de IA - como sistemas de reconhecimento facial - na segurança pública, em que se teme o maior encarceramento de pessoas negras no Brasil, quando essa já é uma realidade gritante no país^{xxx}, bem como a criação de imagens de IA associando pessoas negras a armas e fuzis^{xxx}.

Assim, o mundo virtual e real alimentam-se reciprocamente e impactam diretamente a vida de bilhões de pessoas todos os dias. E esse impacto começa, inclusive, com a produção de supercomputadores e a maior emissão de gases de efeito estufa^{xxxii}, bem como a degradação ambiental que atinge ainda mais pessoas negras e indígenas^{xxxiii}, historicamente marginalizadas na sociedade, e que sofrerão os efeitos das mudanças ambientais e climáticas e do esgotamento de recursos naturais de uma forma mais intensa do que grupos abastados na divisão internacional informacional.

ⁱ McCarthy, John *et al*, 1955, p. 99.

ⁱⁱ Porto, 2018, p. 97.

ⁱⁱⁱ O “aprendizado profundo” é o termo utilizado para descrever um tipo de modelo estatístico de probabilidade, no qual há possibilidade de previsão de cenários futuros por meio da máquina.

^{iv} Kai-Fu Lee, 2019, p. 131.

^v Refere-se ao processo de análise de grandes conjuntos de dados para encontrar padrões, anomalias e correlações, possibilitando, assim, a previsão de resultados.

^{vi} Disponível em: <<https://exame.com/tecnologia/meta-corrigira-ferramenta-discriminatoria-para-anuncios-de-moradia/>>. Acesso em: 23 jun. 2024.

^{vii} Disponível em: <<https://forbes.com.br/negocios/2018/10/amazon-desiste-de-ferramenta-secreta-de-recrutamento/>>. Acesso em: 24 jun. 2024.

^{viii} Disponível em: <<https://diplomatique.org.br/racismo-algoritmico/>>. Acesso em: 24 jun. 2024.

^{ix} Veja também: SILVA, Maria Fernanda; OLIVEIRA, Cristina Godoy Bernardo de. O impacto social causado pelo uso de algoritmos discriminatórios e a superveniência da LGPD. Disponível em: < [O impacto social causado pelo uso de algoritmos discriminatórios - Migalhas](#)>, acessado em 10 de julho de 2024.

^x Citron; Pasquale, 2014.

^{xi} Disponível em: <<https://gdpr-info.eu/>>. Acesso em: 24 jun. 2024.

^{xii} Disponível em: <https://www.sbmac.org.br/wp-content/uploads/2021/02/Portugue%CC%82s-UdeM_Decl-IA-Resp_LA-Declaration_vf.pdf>. Acesso em: 24 jun. 2024.

^{xiii} Trata-se de um conjunto de processos e métodos que garantem aos usuários humanos compreenderem e confiarem nos resultados obtidos com os algoritmos de aprendizado de máquina.

^{xiv} A auditabilidade é característica inextricável à transparência. Por meio dela, é possível que os comportamentos dos sistemas de IA sejam validados, auxiliando em processos de identificação e correção de desvios e erros desses sistemas. Além disso, a auditoria geralmente ocorre por meio de autoridades independentes, proporcionando ainda mais segurança e confiabilidade nos resultados.

^{xv} Knight, 2017.

^{xvi} Projeto ainda não aprovado. Disponível em: <<https://www.congress.gov/bill/116th-congress/house-bill/2231>>. Acesso em: 05 jul. 2024.

-
- xvii Aprovado em março de 2024. Disponível em: <<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>>. Acesso em: 05 jul. 2024.
- xviii Termo utilizado para explicar o reforço às dinâmicas de cunho racista na sociedade realizado por interfaces e sistemas automatizados (Silva, 2020, p. 31).
- xix Mais informações em: <<https://www.ftc.gov/news-events/news/press-releases/2023/11/ftc-authorizes-compulsory-process-ai-related-products-services>>. Disponível em: 05 jul. 2024.
- xx Sobre uma análise complementar: LIMA, Cíntia Rosa Pereira de; FALEIROS JÚNIOR, José Luiz de Moura. Depois de quase 10 anos de importantes debates: o Regulamento Geral Europeu sobre Inteligência Artificial é aprovado. Disponível em: <[Depois de quase 10 anos de importantes debates - Migalhas](#)>, acesso em 11 de julho de 2024.
- xxi Disponível em: <<https://www.wyden.senate.gov/imo/media/doc/algorithmic-accountability-act-of-2023-summary.pdf>>. Acesso em: 05 jul. 2024.
- xxii Disponível em: <https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_pt>. Acesso em: 05 jul. 2024.
- xxiii Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>>. Acesso em: 05 jul. 2024.
- xxiv Disponível em: <<https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence>>. Acesso em: 05 jul. 2024.
- xxv Disponível em: <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065>>. Acesso em: 05 jul. 2024.
- xxvi A visão *Good AI Society* da União Europeia é baseada no conjunto de valores que orientam a formulação de políticas públicas pela Comissão Europeia quanto à IA. Esses valores são expressos em todos os documentos anteriormente mencionados.
- xxvii Termo cunhado por Anu Bradford para explicar a influência regulatória da Europa sobre o mundo, exercendo um poder unilateral como agente regulador internacional (Rodrigues, 2021, p. 205).
- xxviii Segundo Fabrício Polido, o termo refere-se à transformação observada na divisão internacional do trabalho no contexto da era da informação (Polido, 2020, p. 233).
- xxix Moura, 1983.
- xxx Segundo relatório da Rede de Observatórios da Segurança de 2019, o estado da Bahia, com uma das maiores populações negras no país, realizou 51% das prisões registradas no Brasil com o uso de tecnologia de reconhecimento facial, considerando que entre março e outubro de 2019, 151 pessoas foram presas na Bahia, Rio de Janeiro, Santa Catarina e Paraíba. Disponível em: <<https://observatorioseguranca.com.br/rede-de-observatorios-da-seguranca-lanca-primeiro-relatorio-no-ceara/>>. Acesso em: 05 jul. 2024.
- xxxi Mais informações em: <<https://www1.folha.uol.com.br/tec/2024/06/ia-do-whatsapp-gera-imagens-de-criancas-armadas-e-associa-fuzis-a-pessoas-negras.shtml>>. Acesso em: 05 jul. 2024.
- xxxii O Google aumentou suas emissões de carbono em quase 50% ao longo de cinco anos devido à incorporação da IA generativa em diversas áreas de sua atuação; distanciando-se da meta de atingir zero emissões líquidas de gases de efeito estufa até 2030. Disponível em: <<https://valor.globo.com/empresas/noticia/2024/07/03/poluicao-do-google-cresce-48-em-5-anos.ghtml>>. Acesso em: 05 jul. 2024.
- xxxiii Disponível em: <<https://www12.senado.leg.br/noticias/materias/2022/06/27/negros-e-indigenas-sao-os-mais-afetados-por-catastrofes-ambientais-aponta-debate>>. Acesso em: 05 jul. 2024.