INTELIGÊNCIA ARTIFICIAL E VIESES ALGORÍTMICOS: DESAFIOS ÉTICOS E JURÍDICOS EM TEMPOS DE INOVAÇÃO

ARTIFICIAL INTELLIGENCE AND ALGORITHMIC BIASES: ETHICAL AND LEGAL CHALLENGES IN TIMES OF INNOVATION

Carlos Eduardo Santana[1]
Thomas Willian Galagarotto Fumagalli[2]
Orion Ferreira Lima [3]

Resumo: Este trabalho examina as implicações éticas no uso da inteligência artificial (IA) e seu impacto na legislação brasileira. A pesquisa demonstra que a IA, quando utilizada de forma enviesada, pode reproduzir discriminações algorítmicas como o racismo algorítmico. Investigamos a responsabilidade pelo uso indevido de algoritmos, com ênfase em tecnologias de reconhecimento facial que promovem discriminações. Além disso, discutimos a ética relacionada à IA, abordando os desafios da responsabilização moral de desenvolvedores e provedores. Nesse contexto, analisamos como a IA carece de autonomia, liberdade e voluntariedade, características essenciais para a moralidade humana. O estudo se apoia nas contribuições do Projeto de Lei 2338/2023, que visa a regulamentação ética da IA no Brasil.

Palavras-chave: Inteligência Artificial. Viés Algorítmico. Autonomia. Legislação Brasileira.

Abstract: This paper examines the ethical implications of artificial intelligence (AI) and its impact on Brazilian legislation. The research shows that AI, when used in a biased way, can reproduce algorithmic discrimination, such as algorithmic racism. We investigate the responsibility for the misuse of algorithms, with a focus on facial recognition technologies that promote discrimination. Furthermore, we discuss the ethics related to AI, addressing the challenges of moral accountability for developers and providers. In this context, we analyze how AI lacks autonomy, freedom, and voluntariness, essential characteristics for human morality. The study draws on contributions from Bill 2338/2023, which aims to establish ethical regulation for AI in Brazil.

Keywords: Artificial Intelligence. Algorithmic Bias. Autonomy. Brazilian Legislation.

Introdução

O presente trabalho pretende refletir as implicações ética do uso da Inteligência Artificial (IA) e como a legislação brasileira tem se posicionado perante os desafios tecnológicos que atingem à dignidade humana.

Inicialmente pretendemos revisitar o conceito de inteligência artificial em seus aspectos históricos. Nesse sentido, é relevante considerar os impactos da Segunda Guerra

[1] Discente do Curso de Filosofia. Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Filosofia da Faculdade João Paulo II – FAJOPA como parte das exigências para obtenção do título de Licenciado em Filosofia, sob orientação do Prof. Ms. Thomas Willian Galgarotto Fumagalli. E-mail: ceduardo268@gmail.com

[2] Docente do Departamento de Filosofia – orientador do trabalho de conclusão de curso. E-mail: thomaswf@hotmail.com

[3] Docente do Departamento de Filosofia – membro da banca de trabalho de conclusão de curso. E-mail: orionferreira@yahoo.com.br

Mundial no avanço tecnológico e o uso instrumental desses recursos no período pós-guerra.

Desde as descobertas de Turing até o presente momento, a ciência computacional tem transitado por momentos de esperança e decepção, uma vez que servem tanto aos interesses do capitalismo neoliberal quanto ao progresso que permite a conservação e perpetuação da vida. Por essa razão, se faz necessário estabelecer uma agenda ética de debate acerca do usos dessas tecnologias na vida.

Devido sua capacidade única de raciocinar sempre fizeram parte do pensamento humano as reflexões acerca do mundo em que se vive e sobre si próprio. Decodificar sua origem e a do que se vê ao redor tem sido objeto de estudo desde a antiguidade. Neste sentido, o conhecido *cogito cartesiano* em sua máxima "*penso, logo existo*", confere uma estrutura lógica para relacionar a questão do uso do cérebro humano no ato mais simples de pensar como atestado de sua própria existência (DESCARTES, 1983).

À medida que a reflexão sobre a natureza do pensamento humano evoluiu, a IA emergiu como um campo que busca replicar essa capacidade de raciocínio. Atualmente, a IA tem se tornado uma força transformadora em diversos setores da sociedade, desde a saúde e educação até a segurança pública e o entretenimento. Com a capacidade de processar grandes volumes de dados e aprender com eles, a IA promete soluções inovadoras para problemas complexos. No entanto, esse potencial é acompanhado de desafios éticos significativos, especialmente no que diz respeito aos vieses algorítmicos. Num segundo momento dedicaremos a analisar em que medida a legislação brasileira se encontra preparada para lidar com os dilemas éticos e jurídicos que emanam no uso da inteligência artificial no cotidiano. Destaca-se, nesse sentido, o Projeto de Lei

2338/2023 (PL) do chefe do Senado Brasileiro, Rodrigo Pacheco, em que se procura delimitar o uso da inteligência artificial no que tange aos riscos que ela pode causar à dignidade humana. Veremos que são tentativas de cunho jurídico que procuram compreender os limites da tecnologia e como ela pode afetar a vida das pessoas, sobretudo, das mais vulneráveis.

A crescente implementação de tecnologias de IA em nossas vidas levanta preocupações sobre a maneira como essas ferramentas operam e as consequências que podem gerar. Os vieses algorítmicos surgem quando sistemas de IA, alimentados por dados históricos, reproduzem ou amplificam discriminações existentes, impactando desproporcionalmente grupos marginalizados. A presença de racismo algorítmico é um exemplo alarmante de como a tecnologia pode perpetuar injustiças sociais, levantando questões sobre a responsabilidade dos desenvolvedores e das organizações que implementam essas tecnologias.

Neste contexto, a responsabilidade ética dos desenvolvedores de IA se torna uma questão central. É imperativo que aqueles que projetam e implementam esses sistemas considerem as implicações morais de suas escolhas, especialmente ao lidar com dados que podem refletir preconceitos sociais. A falta de diversidade nas equipes de desenvolvimento de IA e a natureza dos dados utilizados frequentemente contribuem para a perpetuação desses vieses, evidenciando a necessidade de um exame crítico das práticas atuais no setor.

A origem dessa responsabilidade remete a investigações fundamentais na área da computação, especialmente as de Alan Turing, que, ao afirmar que “pensar é calcular”, estabeleceu as bases para a disciplina da Inteligência Artificial (IA). Essa nova área do conhecimento não apenas confere ao ser humano um status de existência, mas também o posiciona como um "criador". Como ressaltam Russell e Norvig (2013), “o campo da inteligência artificial vai ainda mais além: ele tenta não apenas compreender, mas também construir entidades inteligentes” (p. 22). Essa dualidade de papel — como pensadores e criadores — traz à tona a necessidade de uma reflexão profunda sobre as implicações éticas e sociais da IA.

Em nossa última sessão, pretendemos ressaltar a importância de estabelecer uma agenda de investigação e discussão acerca das dimensões éticas e legais da IA, a fim de garantir que seu desenvolvimento e aplicação não apenas respeitem os direitos humanos, mas também promovam a justiça social.

1 HISTÓRICO GERAL E DEFINIÇÃO DE IA

O marco histórico ideal para entender esse novo campo denominado IA está situado no contexto da Segunda Guerra Mundial (1939-1945), mais especialmente em seu pós guerra e subsequente período de avanço tecnológico (Arbix, 2021). De maneira totalmente triste e infeliz do qual não deve ser motivo de orgulho algum para a humanidade, o período que se sucedeu aos fatos foi dotado de grande avanço nas ciências e nas tecnologias devido ao empenho das nações envolvidas em mostrar poder bélico e financeiro. Mesmo em meio aos efeitos deletérios produzidos pela guerra, conforme Arbix (2021), houve tamanha evolução especialmente na área da tecnologia nunca antes vista.

Como prova disso, a antecipação do fim da guerra é marcada, devido aos estudos no período de batalhas do matemático britânico Allan Turing (1912-1954), que criou uma máquina que interceptava mensagens secretas de estratégias de combate dos alemães afim de conseguir antecipadamente saber o que os mesmos tramavam. Sua história ficou conhecida mundialmente no filme hollywoodiano O Jogo da Imitação (2014), do diretor norueguês Morten Tyldum.

Turing, com seu artigo “Computing Machinery and Intelligency” de 1950 reflete sobre o fato de uma máquina ser ou não capaz de pensar, pois, segundo ele, “pensar é calcular”. Sendo assim, surgem as primeiras tratativas daquilo que mais tarde reproduziria nas máquinas as redes neurais de processamento (RNP) semelhantes às do homem. Suas contribuições, especialmente a elaboração do seu Teste de Turing, que envolvia fazer uma máquina imitar a atividade humana de modo que, ao ser submetida a um interrogatório por escrito, o interlocutor humano não pudesse distinguir se as respostas vinham de um humano ou de uma máquina, impulsionaram o avanço inicial da IA como a conhecemos hoje. (RUSSEL & NORVING, 2013)

Após esse impulso, o ponto de partida oficial acadêmico foi a Conferência de Dartmouth, em 1956, onde o termo "Inteligência Artificial" foi cunhado, e a IA começou a se firmar como um campo acadêmico e científico independente (RUSSELL & NORVIG, 2013).

Desde então, o campo da IA evoluiu em ciclos de otimismo e decepção. Na década de 1960, houve um grande entusiasmo em relação às possibilidades da IA, mas, nas décadas seguintes, a área sofreu com a falta de avanços significativos e problemas técnicos que não puderam ser superados na época, período conhecido como o "Inverno

da IA" (RUSSELL & NORVIG, 2013, p. 17). O que veio a se repetir na década de 1990, segundo Cozman e Neri (2021) com um inverno "menos rigoroso, mas não menos acabrunhante para a comunidade acadêmica da área" (p. 23).

Consta ainda que

> Por um lado, durante essa década muitas diferentes técnicas computacionais foram desenvolvidas; em particular, houve significativo uso de probabilidades e lógica para representar conhecimento, bem como estatística para aprendizado e teoria de utilidade e de controle para tomada de decisão. (Cozman e Neri, 2021, p. 23-24)

Assim, progresso em áreas que consolidam a IA hoje como aprendizado de máquina (*machine learning*), aprendizado profundo (*deep* learning), processamento de linguagem natural e maior poder de processamento dos computadores, bem como a inovação das *big datas*, a forma de armazenamento e conexão de dados em grande volume e alta velocidade devido a redes de *internet* como o 5G, trouxeram a IA de volta ao centro das atenções no final do século XX e início do século XXI. (Cozman e Neri, 2021; Lage, 2022)

Marcando esse momento, em 1997 um grande holofote retornou a ela quando em uma partida de xadrez ocorre a vitória do computador Deep Blue, da IBM (International Business Machines Corporation), sobre Garry Kasparov, campeão mundial naquele momento. Outro grande momento foi a vitória em um jogo conquistada pelo computador Watson, também da IBM, em 2011. No jogo *Jeopardy*, humanos foram vencidos pelo computador que analisou uma massa de dados em linguagem natural e respondeu todas as perguntas mais rapidamente. (COZMAN, PLONSKI & NERI, 2021)

Assim, muitos já tentaram definir o conceito de IA e a cada trabalho que aborda o tema, desde a década de 90 a atual, sempre em seu início se trata sobre este assunto (Teixeira, 1990; Gomes, 2010; Gonsales & Kaufman, 2023). Para tanto, há de maneira bem conhecida e difundida certas publicações literárias em forma de livro (Russel & Norving, 2013; Lee, 2019; Cozman, Plonski & Neri, 2021) que condensam um maior volume de informações sobre definição do conceito de IA, de maneira mais ampla e bem fundamentada a origem Dela enquanto campo de estudo, derivando-se dos estudos de Filosofia da Mente, Teoria do Conhecimento e, mais recentemente, das Ciências Cognitivas, bem como do avanço nas áreas de Neurociência, Computação, Tecnologia, Engenharia e Robótica.

O livro intitulado "Inteligência Artificial: uma abordagem moderna" de Stuart Russell e Peter Norving (2013), é bem abrangente sobre muitos assuntos que englobam a IA. Ele aborda o estudo em IA e suas fundamentações de maneira ampla e basilar, reconhecendo-a como um verdadeiro campo universal no qual

> "[...] a IA abrange uma enorme variedade de subcampos, do geral (aprendizagem e percepção) até tarefas específicas, como jogos de xadrez, demonstração de teoremas matemáticas, criação de poesia, direção de um carro em estrada movimentada e diagnóstico de doenças." (p. 22)

Assim, o conceito de IA pode ser abordado a partir de diferentes perspectivas, que se dividem em duas dimensões principais: o foco em decodificar processos do pensamento ou do seu comportamento e funcionalidade racional; e o foco de simular ou replicar tais funções humanas em máquinas.

Um passo seguinte desses autores, Russel e Norving (2013), foi, então, apresentar oito definições clássicas sobre IA disponíveis a seu tempo em forma de tabela, que trazem a visão das duas dimensões. As (i) definições que enfatizam o pensamento humano ou racional encontram-se no topo da classificação, enquanto (ii) as que se concentram em replicar funções humanas estão localizadas na parte inferior.

(i) Pensando como um humano	(i) Pensando racionalmente
"O novo e interessante esforço para fazer os computadores pensarem (...) *máquinas com mentes*, no sentido total e literal." (Haugeland, 1985) "[Automatização de] atividades que associamos ao pensamento humano, atividades como a tomada de decisões, a resolução de problemas, o aprendizado..." (Bellman, 1978)	"O estudo das faculdades mentais pelo uso de modelos computacionais." (Charniak e McDermott, 1985) "O estudo das computações que tornam possível perceber, raciocinar e agir." (Winston, 1992)
(ii) Agindo como seres humanos	(ii) Agindo Racionalmente
"A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas."	"Inteligência Computacional é o estudo do projeto de agentes inteligentes." (Poole et al., 1998)

(Kurzweil, 1990) "O estudo de como os computadores podem fazer tarefas que hoje são melhor desempenhadas pelas pessoas." (Rich and Knight, 1991)	"AI... está relacionada a um desempenho inteligente de artefatos." (Nilsson, 1998)

Russel e Norving (2013), adaptada pelo autor

Mais recentemente o ex-presidente da empresa *Google* na China, Kai-Fu Lee, elaborou um livro em que aprimora e traz conceitos e consolidações ainda melhores sobre IA. Com títulos de capítulos instigantes e uma abordagem histórica da IA mais voltada em seu país, seus escritos servem bem para apresentar a tese das 4 ondas da IA, que auxilia na compreensão de como a IA caminhou, caminha e caminhará.

A Primeira Onda da IA é a "IA da Internet". Esta fase refere-se ao uso de IA para analisar e entender os comportamentos dos usuários na internet, permitindo que empresas forneçam serviços mais personalizados. É uma etapa em que os algoritmos aprendem a partir das interações dos usuários online, refinando constantemente suas capacidades com base no grande volume de dados disponíveis. (LEE, 2019, p. 122)

Segunda Onda: IA Empresarial. Nesta fase, a IA é aplicada a processos internos das empresas, analisando dados estruturados que melhoram a tomada de decisões empresariais. A IA ajuda a otimizar operações, sugerir estratégias e antecipar resultados a partir de padrões presentes nos dados históricos das empresas. (LEE, 2019, p. 126)

A Terceira Onda é a "IA da Percepção". Aqui, a IA começa a integrar dados do mundo real, como imagens, sons e movimentos, permitindo que as máquinas tenham uma percepção do ambiente ao seu redor. Isso envolve o uso de câmeras, sensores e outras tecnologias que captam informações do mundo físico e as integram ao processamento da IA. (LEE, 2019, p. 133)

Na Quarta Onda, em relação a independência humana, temos a "IA Autônoma" que é a fase mais avançada, na qual a IA assume um papel de tomada de decisões independente, controlando veículos autônomos, robôs industriais e outras tecnologias que não dependem da intervenção humana constante. Essa onda representa o auge da integração entre a IA e o mundo real, com sistemas capazes de agir por conta própria. (LEE, 2019, p. 145)

Ainda assim, sob este prisma das áreas de atuação da IA, Cozman, Plonski & Neri (2021) argumentam que não se pode confundir toda e qualquer atividade que envolva aparelhos digitais como provenientes do uso de IA em seu funcionamento. Para os autores

> Muitas inovações recentes creditadas à inteligência artificial decorrem simplesmente da automatização de tarefas quotidianas ou do uso de tecnologias já dominadas há algum tempo. Por exemplo, temos dispositivos como *smart* câmeras, por exemplo, em que técnicas sofisticadas de processamento de imagens produzem efeitos surpreendentes mas que dificilmente podem ser confundidos com inteligência. Outras notícias nos alertam sobre aparelhos de ar-condicionado inteligentes e mesmo sorvetes baseados em IA... (COZMAN, PLONSKI & NERI, 2021, p. 22)

Por isso, conforme os autores e para aprofundar nossa reflexão, tomamos como pano de fundo as duas últimas ondas que a IA passou ou tem passado: IA de percepção e IA autônoma. É razoável ter em mente que a IA aqui refletida estará relacionada com computadores digitais e seus periféricos, como citado na segunda onda e noções de ética e responsabilidade para as ações autônomas da IA, pois hoje as IA's "tem" "[...] conhecimento e crenças, tomam decisões e aprendem, e interagem com seu ambiente, realizando todas essas atividades ou pelos algumas com nível alto de sofisticação [...]" (COZMAN, PLONSKI & NERI, 2021, p. 23), sendo essa última sentença uma definição razoavelmente clara sobre o escopo e conceito de IA, que será adotada para este trabalho. Por fim, a IA tem ampliado seu alcance para além dos círculos acadêmicos e técnicos, passando a influenciar políticas públicas, legislação e debates éticos, sobretudo no Brasil, que busca equilibrar o progresso tecnológico com a responsabilidade social e a proteção dos direitos humanos.

2 INTELIGÊNCIA ARTIFICIAL NO BRASIL E A LEGISLAÇÃO NACIONAL

A trajetória da Inteligência Artificial (IA) no Brasil começou a ganhar força na década de 1980, apesar de o campo já estar em desenvolvimento no exterior. Até então, a participação de pesquisadores brasileiros em conferências internacionais como a IJCAI (International Joint Conference on Artificial Intelligence) era praticamente inexistente, e publicações nacionais sobre o tema eram escassas. O interesse em IA foi despertado principalmente pela ampla divulgação e sucesso dos sistemas especialistas, que estavam revolucionando a tecnologia em nível global. Esse cenário motivou a formação de grupos

de pesquisa em diferentes regiões do Brasil, com destaque para os esforços pioneiros do pesquisador Emmanuel P. Lopes Passos, que, em 1971, apresentou a primeira dissertação de mestrado em IA no Brasil, com o título “Introdução à Prova Automática de Teoremas” na Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ) (COSTA et al., 2021).

Esse movimento inicial culminou na realização do primeiro grande encontro nacional sobre o tema, o "Simpósio Brasileiro de Inteligência Artificial" (SBIA), em 1984, organizado por Philippe Navaux da Universidade Federal do Rio Grande do Sul (UFRGS). O evento reuniu 35 pesquisadores de 10 instituições de ensino, proporcionando um espaço para troca de experiências e debate sobre o desenvolvimento da IA no país. Nomes como Rosa Maria Vicari, que mais tarde se tornaria uma referência na área de sistemas tutores e processamento de linguagem natural, foram fundamentais nesse processo de estruturação da comunidade de IA brasileira (COSTA et al., 2021). Esse primeiro simpósio foi um marco que consolidou a IA como um campo de pesquisa sério e em ascensão no Brasil.

Com o passar dos anos, a pesquisa em IA no Brasil se expandiu e ganhou projeção internacional. Durante a década de 1990, o SBIA passou por um processo de internacionalização, adotando o inglês como língua oficial e convidando pesquisadores estrangeiros para participar. Essa evolução levou à criação de novos eventos e à maior participação de pesquisadores brasileiros em conferências internacionais. A década de 2000 marcou um período de consolidação e diversificação da IA no Brasil, com um aumento significativo na quantidade de pesquisas e publicações, cobrindo áreas como aprendizado de máquina, sistemas multiagentes, processamento de linguagem natural e redes neurais artificiais. O SBIA se transformou em 2012 no "Brazilian Conference on Intelligent Systems" (Bracis), que continua sendo um dos principais eventos de IA no país (COSTA et al., 2021).

Nos anos 2010, o Brasil testemunhou um crescimento exponencial em pesquisas de IA, levando o país a se posicionar como um dos principais centros de pesquisa da América Latina. Em 2020, a produção acadêmica brasileira em IA se destacou no cenário mundial, ocupando a 12ª posição global, com cerca de 1.236 publicações em 2018. Instituições como a Universidade de São Paulo (USP), a Universidade Estadual de Campinas (Unicamp) e a Universidade Federal de Pernambuco (UFPE) tornaram-se referências em pesquisa e inovação em IA, contribuindo para o reconhecimento internacional do Brasil na área (COSTA et al., 2021). Esse crescimento reflete o

amadurecimento da IA no Brasil e seu potencial de inovação em diversas áreas do conhecimento.

No cenário econômico e empresarial, a IA tem se consolidado como um dos pilares da transformação digital no Brasil. Setores como o agronegócio, saúde, finanças e varejo têm adotado cada vez mais tecnologias de IA para otimizar processos, reduzir custos e oferecer soluções inovadoras. A Empresa Brasileira de Pesquisa Agropecuária (Embrapa), por exemplo, destaca-se pelo uso de IA em pesquisas voltadas para o agronegócio, consolidando o Brasil como um dos principais atores no desenvolvimento de tecnologias de IA aplicadas a esse setor. Além disso, cerca de 15% das médias e grandes empresas brasileiras já utilizam IA em diversas frentes de trabalho, como automação de processos, atendimento ao cliente e análise de fraudes, conforme relatórios de consultorias especializadas (COSTA et al., 2021).

A trajetória da Inteligência Artificial no Brasil é marcada por um crescimento progressivo que vai da formação de uma comunidade acadêmica e de pesquisa à integração da IA em setores estratégicos da economia. Esse desenvolvimento, porém, não ocorre sem desafios, especialmente no que diz respeito à regulamentação e ao estabelecimento de diretrizes éticas para o uso da IA. Com a crescente influência da IA na sociedade e nas atividades econômicas, o Brasil se encontra em um momento crucial para definir políticas e legislações que orientem a evolução dessa tecnologia de forma ética e responsável.

Por isso, o desenvolvimento da regulamentação da IA no Brasil tem se intensificado nos últimos anos, com a preocupação crescente em equilibrar os avanços tecnológicos e a proteção dos direitos fundamentais.

Com isso, no corpo do próprio Projeto de Lei 2338/2023 (PL) do chefe do Senado Brasileiro, Rodrigo Pacheco, ainda em trâmite, mas o mais atual que trata do assunto, apresenta-se um breve histórico da demanda legal e jurídica que tramitou e ainda tramita no Congresso brasileiro. A primeira proposta de lei significativa que tratou da IA no Brasil foi o Projeto de Lei 5051/2019, de autoria do Senador Styvenson Valentim. Esse projeto estabeleceu os princípios para o uso da IA no país e deu início ao debate sobre a necessidade de uma regulamentação que assegurasse o uso ético e responsável da tecnologia (BRASIL, 2023, p. 28).

Em seguida, o PL 21/2020, apresentado pelo Deputado Federal Eduardo Bismarck, trouxe contribuições importantes ao estabelecer fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da IA no Brasil. Esse projeto foi aprovado

pela Câmara dos Deputados e consolidou-se como uma das primeiras tentativas de definir um marco regulatório que orientasse o uso da IA em diferentes setores da sociedade.

Outro marco importante foi o PL 872/2021, do Senador Veneziano Vital do Rêgo, que abordou de maneira mais detalhada o uso da IA e a necessidade de criar um ambiente regulatório que favorecesse a inovação, mas também protegesse os direitos dos cidadãos. Em 2022, esses três projetos passaram a tramitar em conjunto no Senado Federal. Para assegurar a elaboração de uma lei que atendesse às demandas tecnológicas e sociais, foi instituída a Comissão de Juristas, presidida pelo Ministro do Superior Tribunal de Justiça, Ricardo Villas Bôas Cueva, com a missão de elaborar uma minuta de substitutivo a esses projetos, resultando, enfim, na apresentação do Projeto de Lei 2338/2023, de autoria do Senador Rodrigo Pacheco.

O PL 2338/2023 representa o ponto culminante dessa trajetória, estabelecendo princípios fundamentais para o uso da IA, como a transparência, a não discriminação, a proteção de dados e a necessidade de supervisão humana em sistemas de alto risco. Esse projeto destaca a busca do Brasil por um marco legal que não apenas promova o desenvolvimento da IA, mas que também garanta a proteção dos direitos humanos e da dignidade da pessoa.

Com isso, a trajetória legislativa da IA no Brasil reflete um esforço contínuo para construir um ambiente regulatório que seja capaz de acompanhar os avanços tecnológicos e assegurar que a IA seja utilizada de forma ética e responsável no país.

Porém, até aqui se entende a IA como artefato mecânico, muito consolidado em computadores digitais e outras formas tecnológicas atuais que auxiliam em várias áreas aos homens, mas nada se refletiu sobre a responsabilidade ética do resultado obtido no uso dessas inovações. Alguns fatos bem contraditórios apontam a possibilidade da IA estar enviesada e carregada de preconceitos:

> Tecnologias, no entanto, não operam no vácuo, nem determinam a história de povos e países apenas com o desdobrar de sua própria lógica. São, por natureza, artefatos sociais, criados por agentes morais submetidos às tensões da política, das contradições econômicas, das imposições de poder, das desigualdades, de viés cognitivo, de hábitos arraigados e de todos os constrangimentos que alicerçam as sociedades modernas. Esse reconhecimento não diminui o peso específico da tecnologia. Pelo contrário, a identificação de seus traços humanos ajuda a visualizar seus limites, ainda que nem sempre estejam ao alcance de nossa compreensão. (Arbix *in*. Cozman, Plonski e Neri, 2021, p. 264)

Assim, seguindo estes e outros autores (Silva, 2022; Shen, 2023; Rodrigues & Chai, 2023), é preciso apresentar fatos e reflexões sobre o viés algorítmico, reprodutor muitas vezes de racismo algorítmico.

3 VIESES ALGORÍTMICOS: A REPRODUÇÃO DED PRECONCEITOS RACIAIS A PARTIR DA INTELIGÊNCIA ARTIFICIAL

O debate atual sobre o uso da IA e seus limites éticos, que levou a reprodução nacional de mecanismos legais para, principalmente, definir à responsabilização de autoria e identificação para possível penalização judicial do desenvolvedor humano das respostas obtidas pelas máquinas, vem a ser a questão fundante sobre o viés algorítmico. É sabido que a reprodução dos *outpus* (cada resultado obtido no uso de alguma ferramenta com IA), conforme argumentado anteriormente, reflete o coletivo humano, pois foi o homem que alimentou e alimenta a base de dados da qual ela obtém suas produções, ou seja, se o ser humano age eticamente ou não, a IA também reproduzirá sua conduta moral, não sendo possível responsabilizar o aglomerado de plástico, aço, fios ou qualquer outro material físico utilizado neste ambiente tecnológico como detentor de uma ação moral boa ou ruim, conforme a ética, pois eles não possuem elementos o que um agente moral deve ter para responsabilidade por suas ações: voluntariedade, liberdade e – mesmo com os avanços da quarta onda de IA de automação – autonomia.

De acordo com Shen (2023), a discussão sobre a Ética a Nicômaco de Aristóteles revela a complexidade dos princípios morais, destacando as abordagens da ética intuicionista, que se baseia na intuição inata, e da ética empírica, que considera a experiência como fundamental na formação do sentido moral. Essa reflexão levanta questões sobre a possibilidade de codificar princípios morais em máquinas. Shen argumenta que, apesar de a IA operar com princípios éticos determinados por humanos, ela não possui a capacidade de compreender a moralidade, o que a impede de tomar decisões adequadas em dilemas morais complexos, como o "Dilema do Trolley"[4]. Assim, conclui-se que, por sua natureza projetada, a IA não pode ser responsabilizada por suas ações, pois opera apenas segundo procedimentos predefinidos e não entende os valores humanos

Embora existam variações na compreensão acadêmica sobre a autonomia da IA, é

[4] Mais conhecido como "Dilema do Bonde" onde, suscintamente: o agente moral deve tomar uma decisão sobre puxar ou não uma alavanca para mudar a direção de um trem desgovernado. Se puxar a alavanca, somente uma pessoa morre. Se não puxar, cinco pessoas acabariam morrendo. Trindade (2015) revela que originalmente este experimento de pensamento foi elaborado pela filosofa britânica Phillippa Foot (1920 – 2010). TRINDADE, Gabriel G. Resenha de: Cathcart, Thomas. The trolley problem, or would you throw the fat guy off the bridge?A philosophical conundrum. New York: Workman, 2013. Peri, v. 07, n. 02, 2015. p. 2012

amplamente aceito que ainda estamos na fase da IA fraca, com um nível de autonomia muito restrito – IA fraca refere-se a sistemas que realizam tarefas específicas e limitadas, sem consciência ou entendimento próprio, funcionando apenas dentro dos parâmetros programados, conforme Searle (2011). A maior parte dos robôs contemporâneos opera sob significativa supervisão humana e não atinge uma autonomia plena. A IA atual baseia-se em algoritmos, onde especialistas definem princípios morais como parâmetros, permitindo que as máquinas façam julgamentos e tomem ações baseadas nesses princípios.

Sendo assim, outra reflexão vai até a aparente neutralidade moral apresentada para os algoritmos, que são os responsáveis pelas conexões de assuntos, tanto do modo máquina-máquina, quanto, máquina-homem, homem-máquina.

Segundo Shen (2023) "as questões éticas da IA giram principalmente em torno de três aspectos: vulnerabilidade no aprendizado de máquina, vulnerabilidade humana e alienação da relação homem-máquina" (p. 62), onde para este trabalho se destacam dois desses termos: a vulnerabilidade no aprendizado de máquina está se referindo à susceptível exposição dos sistemas de IA a erros e preconceitos, onde tal fato é causado pela complexidade dos sistemas ao qual a IA acessa gerando uma espécie de "caixa preta da IA" como comparação dos dados seguros de um avião, significando ser difícil delimitar as bases de dados confiáveis do qual a IA se nutre, sendo possível a mesma ser nutrida de locais escassos de regulamentação legal ou até onde o alcance judicial não possa ir ou encontre barreiras cibernéticas rígidas para legislar. E também o termo alienação da relação homem-máquina: referindo-se à transformação da relação entre humanos e máquinas, onde, com o aumento da autonomia dos sistemas de IA, as pessoas passam de utilizadores a meros objetos de controle pelos algoritmos. Isso ameaça a posição dominante dos humanos na tomada de decisões, gerando um deslocamento de poder da humanidade para as máquinas

Tudo isso tem gerado impactos negativos e riscos que está afetando a vida das pessoas e suas relações sociais que agora, segundo a teoria da Sociedade da Informação, como bem aborda alguns autores (Teixeira, 1994; Moraes, 2014), está digitalizada, a ponto de ser questionado se os algoritmos são neutros de forma a não reproduzir preconceitos humanos. Isso é denominado IA enviesada, que reproduz preconceitos

humanos e teve, nos casos a serem apresentados, repercussão significativa no debate atual.

O trabalho de Tarcízio Silva é basilar e fundamenta muitos assuntos sobre racismo a serem discutidos. É de sua autoria também o portal "Linha do Tempo do Racismo Algorítmico: casos, dados e reações", mais adiante apresentado. Em seu livro Racismo Algorítmico: Inteligência artificial e discriminação nas redes digitais (2022) é abordado de maneira ampla toda esse dilema. Trabalharemos dois conceitos e fatos que reforçam cada um deles conforme Silva (2022).

O conceito de "racismo algorítmico", tal como apresentado por Tarcízio Silva, revela como sistemas de inteligência artificial podem reforçar, reproduzir e amplificar preconceitos raciais. Esse fenômeno ocorre quando algoritmos, que se pretendem neutros, são alimentados com dados históricos que refletem desigualdades sociais, resultando em discriminações sistêmicas. Essa ideia de que os algoritmos "aprendem a ser racistas" não é um exagero, mas uma constatação preocupante dos vieses ocultos na infraestrutura digital que governa nossas interações cotidianas. Como o autor demonstra, a coleta massiva de dados e a formação de perfis algorítmicos reproduzem padrões de segregação racial, muitas vezes de forma invisível. Esse processo não apenas naturaliza a exclusão social de grupos racializados, mas também questiona a premissa de imparcialidade frequentemente atribuída à tecnologia. Portanto, como definição mais própria de racismo algorítmico vem a ser

> O modo pelo qual a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados. Tal ordenação pode ser vista como uma camada adicional do racismo estrutural, que, além do mais, molda o futuro e os horizontes de relações de poder, adicionando mais opacidade sobre a exploração e a opressão global que já ocorriam desde o projeto colonial do século XVI. (SILVA, 2022, p. 66).

Fatos que corroboraram a estes argumentos são refletidos por Rodrigues e Chai (2023) ao relatarem que em 2015, por meio de métricas aplicadas por IA num aparelho de visão computacional, o programador Jacky Alciné teve suas fotos e de sua namorada sinalizadas com o termo "gorilas" no álbum de fotos eletrônico *Google* Photos. Outro fato é a indevida colocação de "tag's" em fotos de cabelos negros e crespos os sinalizando como perucas. (RODRIGUES & CHAI, 2015)

Em seu artigo, ainda, refletem sobre a branquitude que vem sendo usada para agravar a falta de representatividade de pessoas negras nos postos de comando de grandes empresas detentoras das *Big Datas* (grandes dados) que armazenam os dados que geram influência para a segregação racial digital, pois "é evidente que a natureza enviesada da tecnologia, regida e concentrada nos pólos computacionais, como o Vale do Silício, que possui pouca diversidade cultural e étnica em seus grupos de desenvolvedores". (RODRIGUES & CHAI, 2015, p. 101)

Dois outros fatos colhidos por relatos das próprias vítimas via *Instagram*[5] corroboram com o tema. A deputada estadual do Rio de Janeiro, Renata Souza (42), mulher autodeclarada negra, apresentou em seu perfil na rede social em 25 de outubro de 2023 que um *input* (entrada de informações) na "trend" de geração de imagens por IA estilo princesa da Disney, com as seguintes determinações "gere uma imagem de uma mulher negra com uma favela ao fundo e vestida de um blazer estilo africano". A imagem obtida foi a de uma mulher com uma favela no fundo portando uma arma de fogo, que não fora em nenhum momento pedido para a ferramenta. A deputada encaminhou medidas legais a produtora do aplicativo, identificada como a *big tech* Microsofot Bing. Na mesma esteira, em 11 de julho de 2024, uma produtora do Instituto Conhecimento Liberta, na mesma rede social[6], explanou que ao utilizar o aplicativo de edição de fotos *Canva* e solicitar que ele aplicasse a ferramenta de expansão de imagem, uma capacidade desse tipo de IA que sugere um aumento do cenário em volta uma foto. A foto era de Lélia Gonzalez, escritora e ativista negra brasileira. O inusitado foi que a sugestão proposta trazia uma porção de cocaína em cima de uma mesa. Com a mesma ferramenta uma imagem com um copo de leite ao lado foi sugerida ao submeter uma foto de um homem branco.

A "necropolítica algorítmica" é outro conceito chave abordado por Tarcízio Silva e que merece uma atenção especial no contexto brasileiro. Inspirada na teoria de Achille Mbembe, essa ideia relaciona a tecnologia de IA com a gestão da vida e da morte, especialmente no que concerne ao uso de sistemas de reconhecimento facial pela polícia. Segundo Silva, essas tecnologias operam de maneira seletiva, exacerbando a criminalização de corpos negros e racializados. A questão ética aqui se intensifica, já que

[5] Disponível em: <https://www.instagram.com/p/Cy1p6EQpwXB/>. Acesso em 01 de out. 2024.
[6] Disponível em: <https://www.instagram.com/p/C9SxRorvJyg/>. Acesso em 01 de out. 2024.

esses sistemas, supostamente criados para aumentar a segurança, tornam-se, na verdade, mecanismos de perpetuação de violência contra minorias.

Sobre este conceito, os fatos envolvem os erros consistentes que tem ocorrido com a IA de uso na segurança pública para ajudar no reconhecimento facial de criminosos. Silva (2021)[7] relata o caso de uma abordagem equivocada a um adolescente feita pela polícia da Bahia, identificando-o como traficante por câmeras de segurança no metrô de Salvador e por isso detido. A mãe do jovem, em entrevista à TV Itapoan, relatou que o filho voltou para casa profundamente abalado, precisando de tempo para se acalmar e contar o ocorrido. Após o incidente, ele desenvolveu medo de frequentar a escola e utilizar transporte público, devido ao trauma e à violência policial que enfrentou.

A discussão sobre racismo algorítmico destaca a urgência de regulamentação que responsabilize empresas de tecnologia (*big techs*) e instituições públicas. O Projeto de Lei 2338/2023, em tramitação, busca abordar essas questões, mas ainda precisa de dispositivos claros de penalização, que ainda não ocorreu. Com isso, encaminho para as considerações finais.

Considerações finais

A presente análise das implicações éticas e jurídicas da Inteligência Artificial (IA) evidencia um campo de investigação repleto de complexidades e desafios intrínsecos, especialmente no que tange ao viés algorítmico e suas repercussões sociais. Ao longo deste trabalho, foi possível constatar que, embora a IA tenha o potencial de promover avanços substanciais em diversos domínios, sua implementação pode igualmente reproduzir e amplificar preconceitos arraigados, como o racismo algorítmico. Tais vieses não se restringem a problemas técnicos, mas são, antes, reflexos das desigualdades sociais e da homogeneidade nas equipes responsáveis pelo desenvolvimento dessas tecnologias. A reprodução de preconceitos, especialmente em casos de racismo algorítmico, exemplifica como a IA pode perpetuar desigualdades estruturais já presentes na sociedade. Casos de discriminação racial, como os citados no uso de ferramentas de reconhecimento facial e nos exemplos de algoritmos que reforçam estereótipos, evidenciam que os sistemas de IA, longe de serem neutros, refletem os valores e vícios da humanidade que os alimenta.

[7] Disponível em: <https://tarciziosilva.com.br/blog/reconhecimento-facial-na-bahia-mais-erros-policiais-contra-negros-e-pobres/>. Acesso em 30 de set. 2024.

A discussão em torno do Projeto de Lei 2338/2023, que busca regulamentar o uso da IA no Brasil, representa um passo significativo rumo à construção de um marco legal que assegure a responsabilidade social e a salvaguarda dos direitos humanos. Contudo, a eficácia dessa regulamentação estará intrinsecamente ligada à sua capacidade de abordar de maneira abrangente os desafios que a tecnologia impõe, promovendo um ambiente inclusivo que busque mitigar os riscos associados ao uso da IA.

É imperativo que acadêmicos, desenvolvedores e legisladores atuem em colaboração, a fim de garantir que as inovações tecnológicas sejam concebidas e implementadas de forma ética e responsável. A promoção de uma maior diversidade nas equipes de desenvolvimento, aliada à criação de mecanismos robustos de transparência e responsabilização, revela-se fundamental para a mitigação dos riscos de discriminação algorítmica.

Finalmente, é essencial que a sociedade civil se mantenha engajada neste debate, exigindo a construção de sistemas de IA que não apenas promovam a eficiência e a inovação, mas que também respeitem a dignidade humana e fomentem a equidade social. Somente por meio de um esforço coletivo e multidisciplinar será possível assegurar que a IA se configure como uma ferramenta de progresso para todos, evitando que se converta em um meio de perpetuação das desigualdades.

Referências

BRASIL. **Projeto de Lei nº 2338/2023**. Dispõe sobre o uso da Inteligência Artificial. Autoria: Senador Rodrigo Pacheco (PSD/MG). Brasília-DF, 2023;

CLAUDIO, Alexandre Aparecido. **Decodificando vieses sociais**: a intermediação decisiva da inteligência artificial e sua própria tendência aos vieses. SciELO Preprints, 2024. Disponível em: https://doi.org/10.1590/SciELOPreprints.7939. Acesso em: 1 out. 2024;

DESCARTES, René. **Meditações** (Pensadores). Editora Abril Cultural. São Paulo-SP, 1983;

GOMES, Dennis dos Santos. **Inteligência Artificial**: conceitos e aplicações. Revista Olhar Científico, Faculdades Associadas de Ariquemes. V. 1, n. 2. Ariquemes-RO, Ago. de 2010;

GONSALES, P.; KAUFMAN, D. **IA na educação**: Da programação à alfabetização em dados. ETD - Educação Temática Digital, v. 25, p. 1-22. Campinas-SP, 2023;

LOPES, Isaías Lima; OLIVEIRA, Flávia Aparecida; PINHEIRO, Carlos A. Murari. **Inteligência Artificial**. Editora Elsevier, 1ª ed. Rio de Janeiro-RJ, 2014;

MORAES, João Antônio de. **Implicações éticas da "virada informacional na filosofia**. Editora EDUFU, Uberlândia-MG, 2014;

ROGRIGUES, Júlia C.; CHAI, Cássius G. **Inteligência Artificial e racismo algoritmo**: análise da neutralidade dos algoritmos frente aos episódios de violão de direitos nos meios digitais. Revisa Eletrônica do TRT-RP. Curitiba-PR, v. 12, n. 118. Mar. de 2023;

RUSSEL, Stuart Jonathan; NORVING, Peter. **Inteligência Artificial**: uma abordagem moderna. Tradução: Regina Célia Simille. Editora Elsevier. Rio de Janeiro-RJ, 2013;

SHEN, Ying Dong. **The ethical dilemma of artificial intelligence and its construction of moral responsibility**. The Frontiers of Society, Science and Technology, v. 5, n. 14, 2023

SILVA, Tarcízio. **Racismo algorítmico**. São Paulo: Edições Sesc São Paulo, 2022; TEIXEIRA, João de Fernandes. O que é Inteligência Artificial? Editora Brasiliense. Brasília-DF, 1990;

TRINDADE, Gabriel Garmendia. Resenha de: Cathcart, Thomas. **The trolley problem or would you throw the fat gu y off the bridge?** A philosophical conundrum. New York: Workman, 2013.

Recebido em: 24/08/2025
Aprovado em: 30/08/2025