https://doi.org/10.1590/ES.289323

AS DUALIDADES ENTRE O USO DA INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO E OS RISCOS DE VIESES ALGORÍTMICOS

João Marcos Heggler[1]
Romeu Miqueias Szmoski[1]
Awdry Feisser Miquelin[1]

RESUMO: O estudo busca identificar as dualidades entre o uso da inteligência artificial (IA) na educação e os riscos de vieses algorítmicos. A pesquisa se insere como qualitativa e uma revisão sistemática de literatura com apoio bibliométrico. A coleta de dados ocorreu nas bases *Scopus*, *Web of Science* e *ScienceDirect*, e ao final foram selecionados 16 artigos para análise e interpretação. A hipótese levantada é de que os vieses algorítmicos na educação podem comprometer a equidade e a eficácia dos processos educativos. Os vieses foram apontados desde a codificação dos algoritmos até o processamento automatizado, o que pode impactar o desempenho dos alunos e ampliar as desigualdades. Medidas sugeridas para mitigação incluem os cuidados nas fases de treinamento e implementação dos algoritmos, desenvolvimento de práticas de reparação algorítmica, supervisão das plataformas e a necessidade de tecnologias justas e confiáveis. Como resposta à hipótese, conclui-se que a eficácia da IA na educação depende de um treinamento contínuo e inclusivo dos algoritmos e da conscientização de seus usuários sobre os riscos dos vieses algorítmicos, de forma que os sistemas possam ser aperfeiçoados para evitar a reprodução de injustiças sociais. Recomenda-se uma abordagem ética e colaborativa para garantir que a IA contribua para uma educação mais justa. A interação entre os alunos e a IA pode suscitar uma reflexão crítica sobre os *feedbacks* recebidos, pois eles precisam saber identificar se os resultados são confiáveis, atualizados e não tendenciosos, uma vez que a IA pode demorar a incorporar novos dados de treinamento nos algoritmos.

Palavras-chave: Viés algorítmico. Algoritmo. Inteligência artificial. Educação.

THE DUALITIES BETWEEN THE USE OF ARTIFICIAL INTELLIGENCE IN EDUCATION AND THE RISKS OF ALGORITHMIC BIASES

ABSTRACT: The study seeks to identify the dualities between the use of artificial intelligence (AI) in education and the risks of algorithmic biases. This qualitative

1. Universidade Tecnológica Federal do Paraná – Departamento Acadêmico de Física – Ponta Grossa (PR), Brasil. E-mails: joaoheggler@utfpr.edu.br; rmszmoski@utfpr.edu.br; awdry@utfpr.edu.br

Editor de Seção: Antônio Zuin

research is a systematic literature review with bibliometric support. Data collection took place across the Scopus, Web of Science, and ScienceDirect databases, ultimately selecting 16 articles for analysis and interpretation. The hypothesis posits that algorithmic biases in education may compromise the equity and effectiveness of educational processes. Biases were observed from the coding of algorithms to automated processing, potentially impacting student performance and widening inequalities. Suggested mitigation measures include care in the training and implementation phases of algorithms, the development of algorithmic repair practices, platform supervision, and the need for fair and reliable technologies. In addressing the hypothesis, the study concludes that the effectiveness of AI in education depends on continuous, inclusive training of algorithms and user awareness of algorithmic bias risks, enabling systems to improve in ways that prevent the reproduction of social injustices. An ethical and collaborative approach is recommended to ensure AI contributes to a fairer education system. Interaction between students and AI can foster critical reflection on received feedback, as students need to identify whether results are reliable, up-to-date, and unbiased, considering the time required for algorithms to incorporate new training data.

Keywords: Algorithmic bias. Algorithm. Artificial intelligence. Education.

LAS DUALIDADES ENTRE EL USO DE LA INTELIGENCIA ARTIFICIAL EN EDUCACIÓN Y LOS RIESGOS DE SESGOS ALGORÍTMICOS

RESUMEN: El estudio busca identificar las dualidades entre el uso de la inteligencia artificial (IA) en la educación y los riesgos de sesgos algorítmicos. Esta investigación es cualitativa y se llevó a cabo una revisión sistemática de la literatura con apoyo bibliométrico. La recolección de datos se realizó en las bases de datos Scopus, Web of Science y ScienceDirect, y finalmente se seleccionaron 16 artículos para su análisis e interpretación. La hipótesis planteada es que los sesgos algorítmicos en la educación pueden comprometer la equidad y la eficacia de los procesos educativos. Se destacaron sesgos desde la codificación de los algoritmos hasta el procesamiento automatizado, lo cual puede afectar el rendimiento estudiantil y ampliar las desigualdades. Las medidas de mitigación sugeridas incluyen cuidados en las fases de capacitación e implementación de los algoritmos, el desarrollo de prácticas de reparación algorítmica, la supervisión de plataformas y la necesidad de tecnologías justas y confiables. En respuesta a la hipótesis, se concluye que la efectividad de la IA en la educación depende de un entrenamiento continuo e inclusivo de los algoritmos y de la concienciación de sus usuarios sobre los riesgos de sesgos algorítmicos, de modo que los sistemas puedan mejorarse para evitar la reproducción de injusticias sociales. Se recomienda un enfoque ético y colaborativo para asegurar que la IA contribuya a una educación más justa. La interacción entre los estudiantes y la IA puede impulsar una reflexión crítica sobre el feedback recibido, ya que necesitan saber identificar si los resultados son fiables, actualizados y libres de sesgos, dado que la incorporación de nuevos datos de formación en los algoritmos puede llevar tiempo.

Palabras clave: Sesgo algorítmico. Algoritmo. Inteligencia artificial. Educación.

Introdução

O uso de inovações tecnológicas, da internet e da inteligência artificial (IA) está transformando o cotidiano da sociedade; o mundo está cada vez mais conectado, e a dependência dos usuários nessas ferramentas está cada vez mais presente. Nesse sentido, a programação e a codificação de algoritmos; o compartilhamento e tráfego de dados, de informações e de comunicações precisam atender certos cuidados éticos, de vigilância, de equidade e de controle, pois o ambiente virtual e tecnológico, além de possibilitar resultados positivos e vantajosos, também pode trazer muitos riscos e danos aos usuários.

Esses riscos podem se referir a muitas questões que podem envolver previsões e decisões baseadas em algoritmos que influenciam o cotidiano das pessoas. Apesar de frequentemente parecerem objetivos e justos, há uma perspectiva de que esses algoritmos possam refletir vieses, preconceitos, discriminações e desigualdades, podendo afetar e causar danos a grupos específicos da sociedade, inclusive na educação (Baker; Hawn, 2022).

Um algoritmo na perspectiva do aprendizado de máquina pode ser conceituado como um conjunto de regras para a realização de tarefas (Davis; Williams; Yang, 2021). Em relação aos algoritmos ditos inteligentes, eles possibilitam a criação de modelos capazes de prever e aprender com os dados constantemente modificados e treinados (Gomede; Barros; Mendes, 2021).

Além disso, os algoritmos ocasionalmente podem assumir vieses e preconceitos de seus desenvolvedores e encontrados na sociedade em razão de determinados contextos – com riscos de produzirem resultados e ocasionarem situações discriminatórias envolvendo determinados grupos da sociedade. Também, o preconceito conhecido como estatístico se refere aos dados e às amostras erradas, incompletas, questionáveis, enquanto o preconceito social representa os dados sobre estruturas específicas e determinados contextos da sociedade (Baker; Hawn, 2022).

O viés algoritmo reflete uma ideia de injustiça em ambientes relacionados a sistemas automatizados, programações, codificações e treinamento de algoritmos; também classificado, em algumas situações, como tendencioso em razão de determinados contextos fáticos ou em relação ao grupo social analisado e implicações aos usuários. O viés algoritmo também pode surgir fora do ambiente de treinamento, resultante de vieses preexistentes ou estatísticos, ou de coleta indevida de dados (Baker; Hawn, 2022).

Nesses ambientes automatizados, para aprender e fazer previsões, os algoritmos dependem dos dados de treinamento recebidos. Se os projetos tecnológicos, as decisões tomadas e os dados refletirem tendências, os resultados produzidos poderão apresentar vieses algorítmicos. Isso é extremamente preocupante para os usuários, em especial na educação, uma vez que tais distorções podem trazer prejuízos para o desempenho dos alunos, aprofundar as desigualdades e comprometer as iniciativas e estratégias educacionais nos processos de ensino e de aprendizagem.

Essa ideia é corroborada por Baker e Hawn (2022) ao mencionarem que os vieses algoritmos podem influenciar os sistemas de aprendizagem e tecnologias educacionais, por exemplo, as distorções nas decisões importantes relacionadas às admissões e avaliações de desempenhos de alunos nas escolas. Esses autores destacam a importância de identificar e mitigar o viés presente nos algoritmos utilizados na educação, visando promover a justiça e a igualdade de oportunidades para todos os alunos.

Em se tratando de mitigação dos vieses e dos respectivos danos, a inexistência ou demora na busca de soluções preventivas ou corretivas para as causas e ocorrências dos vieses algorítmicos tem o potencial de criar ou reforçar as desigualdades sistêmicas, que são extremamente prejudiciais aos indivíduos – como exemplos, desigualdades estruturais, abalos psicológicos e morais, avaliações injustas, aconselhamento inadequado e acesso desigual a recursos educacionais. Essas situações podem prejudicar

determinados grupos sociais, reforçando as disparidades existentes e criando barreiras para a obtenção de bons resultados na aprendizagem.

De um lado, pode-se encontrar a criação e implementação de novas tecnologias baseadas em IA e no aprendizado de máquina, resultando na transformação de diversas áreas do conhecimento, inclusive a educação, uma vez que essas tecnologias prometem revolucionar os processos de ensino-aprendizagem, propondo soluções personalizadas e mais eficientes para alunos e professores. Por outro lado, juntamente com os benefícios, surgem algumas preocupações sobre os riscos da inserção de vieses durante o desenvolvimento de algoritmos. Tal ideia é confirmada por Tang e Su (2024, p. 32) quando eles mencionam que, "[...] embora os modelos de IA proporcionem uma aprendizagem personalizada aos alunos, também enfrentam o problema de nem sempre serem capazes de compreender as necessidades ilimitadas dos alunos" e de garantirem certa proteção aos usuários.

Diante de tudo isso, esta pesquisa pode ser justificada em razão da IA estar em um momento de transformação da educação, possibilitando uma diversidade de oportunidades que poderão influenciar como os alunos e os professores compartilham significados e conhecimentos. Essa inteligência tem a capacidade de personalizar o processo de ensino e de aprendizagem e, também, automatizar processos e *feedbacks* mais rápidos de informações. No entanto, apesar das inovações, dos recursos e das funcionalidades da IA terem a capacidade de potencializar e melhorar os processos educativos, surgem os riscos da inserção de vieses algorítmicos que podem comprometer todos esses processos. A hipótese levantada é de que os vieses algorítmicos na educação podem comprometer a equidade e a eficácia dos processos educativos.

Em razão dessas constatações, o presente estudo tem como objetivo identificar, com base na literatura atual, as dualidades entre o uso da IA na educação e os riscos de inserção e incorporação de vieses algorítmicos nos processos educativos.

Metodologia

A presente pesquisa qualitativa caracteriza-se como uma revisão sistemática de literatura. Na construção do *corpus* de pesquisa, foi utilizado o *software* Bibliometrix (Aria; Cuccurullo, 2017). A aplicação ocorreu através das seguintes etapas, descritas no Quadro 1:

A Fig. 1 apresenta um fluxograma das etapas de utilização do Bibliometrix e *softwares* como uma maneira de simplificação das informações sobre a pesquisa.

De acordo com a etapa 9 mencionada no Quadro l e no fluxograma acima, foram utilizados a inferência e a interpretação para análise sistemática do *corpus* da pesquisa, composto por 16 artigos listados na Tabela 1. Esta investigação não demanda a exigência de aprovação do Comitê de Ética, pois não se refere à pesquisa com seres humanos.

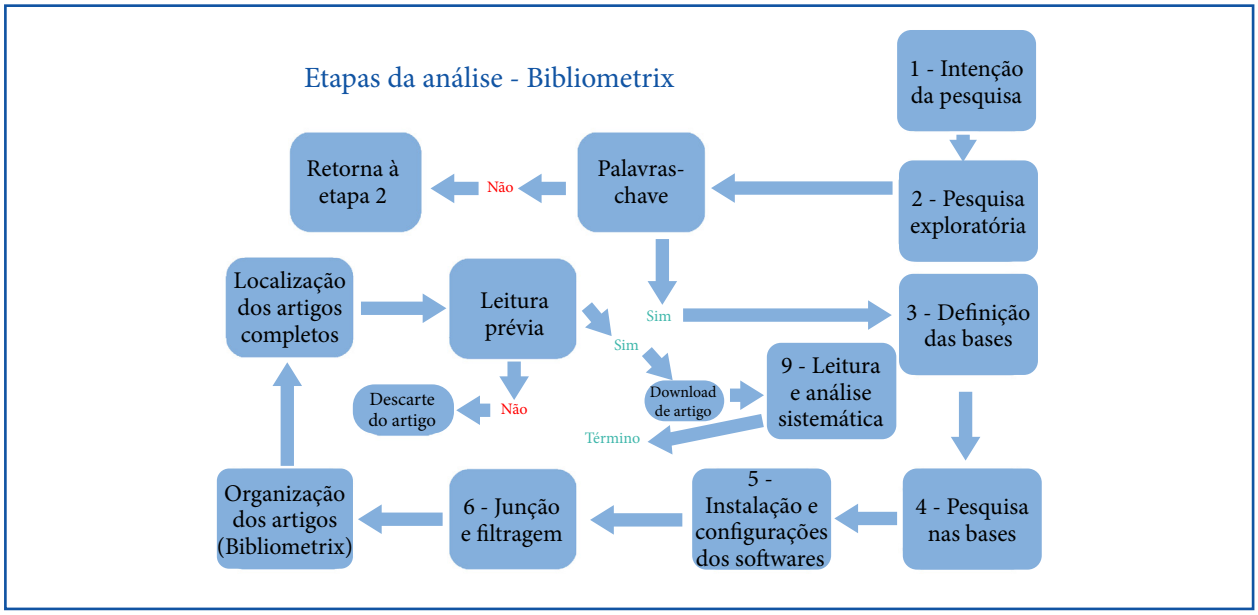
Sistematização dos Dados

A Tabela 1 apresenta uma relação dos 16 artigos selecionados para o *corpus* da pesquisa. Os trabalhos são listados considerando-se o primeiro autor; total de citações; a média de citações por ano; as revistas/fontes das publicações e o respectivo ano, ordenando-se em razão da quantidade total de citações, inferindo-se a relevância de cada trabalho para a temática pesquisada nas bases. Possibilita uma visão geral das publicações selecionadas, permitindo uma análise comparativa dos artigos ao longo do tempo.

Quadro 1. Etapas da utilização do Bibliometrix e *softwares* dependentes.

n.º	Etapas	Procedimentos
1	Estabelecimento da intenção de pesquisa	Nessa etapa foram identificados os descritores de busca e as combinações mais adequadas para responder à pergunta: É possível *identificar, com base na literatura atual, as dualidades entre o uso da IA na educação e os riscos de inserção e incorporação de vieses algorítmicos nos processos educativos?*
2	Pesquisa exploratória	A pesquisa exploratória seguiu as combinações dos descritores de busca: "*Algorithmic bias*" *OR* "*Algorithm bias*" *AND Education*. Foi considerado na busca o período entre janeiro de 2020 e maio de 2024.
3	Definição das bases de dados	Bases escolhidas: *Scopus, Web of Science e ScienceDirect*
4	Pesquisa nas bases	A localização dos trabalhos foi feita no Portal de Periódicos da CAPES, no sítio de periódicos, com o acesso ao "CAFé".
5	Procedimentos de instalação e configuração dos *softwares*	Foram instalados, em sequência, os *softwares* R e *RStudio,* para configuração do ambiente e construção do portfólio. Após a aplicação dos métodos de preparação, configurou-se o pacote Bibliometrix.
6	Procedimentos de junção e filtragem	A pesquisa resultou em um total de 98 artigos, sendo: *Scopus* – 17; *Web of Science* – 56; *ScienceDirect* – 25; e posterior eliminação de 9 artigos duplicados.
7	Organização dos artigos (Bibliometria)	Foram utilizados os *softwares RStudio*, Bibliometrix e o *Biblioshiny* para sistematização e gerenciamento das referências.
8	Localização dos artigos completos e leitura prévia	Foram pré-selecionados 89 artigos completos para a leitura prévia, ocorrendo o descarte daqueles que não apresentaram elementos para responder à pergunta desta investigação. Ao final, foram selecionados os 16 trabalhos para a leitura, interpretação e análise sistemática (etapa 9).
9	Leitura e análise sistemática dos artigos	Nessa etapa, realizou-se uma análise mais aprofundada, sistemática e interpretativa desses 16 artigos, na busca de respostas para a problemática desta pesquisa.

Fonte: Autoria própria.



Fonte: Autoria própria.

Figura 1. Fluxograma das etapas de utilização do Bibliometrix e *softwares*.

O Quadro 2 elenca os objetivos dos trabalhos selecionados, fornecendo uma síntese das metas e os propósitos de cada estudo em relação ao viés algorítmico na educação. Através desse quadro, é possível compreender a diversidade de abordagens e focos dos estudos selecionados, enriquecendo a análise e a compreensão do tema em discussão.

Tabela 1. Relação dos 16 artigos com o total de citações/por ano.

Ord.	1º Autor	Total Citações	TC/ ano	Revistas/Fontes	Ano
1	Gomede	126	31.50	Comput Educ	2021
2	Baker Rs	85	28.33	Int J Art Int Ed	2022
3	Davis Jl	48	12.00	Big Data Soc	2021
4	Sun W	36	7.20	Plos One	2020
5	Johnson Gm	29	7.25	Synthese	2020
6	Zajko M,	15	3.75	Ai Soc	2021
7	Gaskins N	4	2	Techtrends	2023
8	Yoder-Himes Dr	4	1.33	Front Educ	2022
9	Archambault Sg	2	1	Commun Inf Lit	2023
10	Williams Tr	0	0	Front Educ	2024
11	Tang L	0	0	Unir La Univ	2024
12	Solyst J	0	0	Proc Acm Hum	2023
13	Bird Ka	0	0	J Policy A Manage	2024
14	Ito-Jaeger S	0	0	Int J Qual Meth	2023
15	Sanusi It	0	0	Comp Ed Art Int	2023
16	Dakakni D	0	0	Comp Ed Art Int	2023

Fonte: Aria; Cuccurullo (2017).

Quadro 2. Apresenta os objetivos dos 16 trabalhos selecionados na pesquisa.

Autores	Objetivos
Gomede, Barros e Mendes (2021)	[...] contribuir com um sistema adaptativo de recomendação de *e-learning*, com o qual são investigadas quatro abordagens para o desenvolvimento de modelos de recomendação.
Baker e Hawn (2022)	[...] revisar o preconceito algorítmico na educação, discutindo as causas desse preconceito e revisando a literatura empírica sobre as maneiras específicas pelas quais o preconceito algorítmico se manifestou na educação.
Davis, Williams e Yang (2021)	Propõe a reparação algorítmica como base para construir, avaliar, ajustar e, quando necessário, omitir e erradicar sistemas de aprendizado de máquina.
Sun, Nasraoui e Shafto (2020)	[...] estudar duas fontes de preconceito que interagem: o processo pelo qual as pessoas selecionam informações para rotular (ação humana); e o processo pelo qual um algoritmo seleciona o subconjunto de informações para apresentar às pessoas (modo de viés algorítmico repetido).
Johnson (2020)	[...] explorar mais a fundo as conexões entre o algoritmo e os humanos, e argumentar que na etiologia, operação, avaliação e mitigação os preconceitos humanos e os preconceitos algorítmicos compartilham semelhanças importantes.
Zajko (2021)	[...] apresenta a visão de um sociólogo sobre o problema do preconceito algorítmico e da reprodução do preconceito social.
Gaskins (2023)	[...] analisa o preconceito algorítmico ou da inteligência artificial (IA) na tecnologia educacional, especialmente através das lentes da ficção especulativa, do *design* especulativo e libertador. Ele discute as causas do preconceito e revisa a literatura sobre as várias maneiras pelas quais o preconceito algorítmico/IA se manifesta na educação e em comunidades que estão sub-representadas no desenvolvimento de *software EdTech*.
Yoder-Himes et al. (2022)	Este estudo é novo, pois é o primeiro a examinar quantitativamente os preconceitos no *software* de detecção facial na interseção entre raça e sexo e tem (sic) impactos potenciais em muitas áreas da educação, justiça social, equidade e diversidade educacional e psicologia.
Archambault (2023)	[...] propõe acréscimos à atual Estrutura da ACRL para Alfabetização Informacional para o Ensino Superior que está relacionado à alfabetização algorítmica.
Williams (2024)	[...] apresenta brevemente as implicações éticas do uso de plataformas como o ChatGPT na educação. O autor enfoca a privacidade dos dados, o viés algorítmico, a autonomia na aprendizagem e a questão do plágio.
Tang e Su (2024)	[...] busca oferecer uma visão geral da pesquisa sobre as implicações éticas, os princípios éticos e as futuras direções e práticas de pesquisa do uso de modelos de IA em sala de aula por meio de uma revisão sistemática da literatura.
Solyst et al. (2023)	[...] como os jovens podem participar na identificação de preconceitos algorítmicos e na concepção de sistemas futuros para abordar comportamentos tecnológicos problemáticos.

Continua...

Quadro 2. Continuação

Autores	Objetivos
Bird, Castleman e Song (2024)	Para contextualizar nossa análise do viés algorítmico no ensino superior e para motivar os modelos específicos que investigamos, fornecemos uma breve visão geral de como a análise preditiva é usada pelas faculdades e universidades – tanto como elas desenvolvem os modelos de previsão quanto como as faculdades e universidades normalmente aproveitam as previsões resultantes na divulgação e no apoio aos alunos.
Ito-Jaeger et al. (2023)	[...] apresenta uma nova abordagem metodológica, *TrustScapes*, uma ferramenta de acesso aberto projetada para identificar e visualizar as preocupações e recomendações políticas das partes interessadas sobre proteção de dados, preconceito algorítmico e segurança *online* para um mundo *online* mais justo e confiável.
Sanusi et al. (2023)	[...] discernir como os alunos se envolveram com atividades de *Machine Learning* (ML). Além disso, exploramos se a participação em nosso programa de intervenção provoca mudanças na compreensão e nas percepções de ML dos alunos.
Dakakni e Safa (2023)	[...] investigar as atitudes de alunos e professores em relação às ferramentas de IA na sala de aula [...].

Fonte: Autoria própria.

Esses objetivos representam a diversidade de propostas e a complexidade do tema, abordando desde a investigação das causas dos vieses até a proposição de soluções e intervenções para promover a equidade e a justiça na educação. Ao analisar o Quadro 2, é possível perceber as possibilidades de discussões em torno dos vieses e dos preconceitos algorítmicos, destacando-se a necessidade contínua de reflexão crítica e de ações efetivas para a mitigação dos impactos negativos desses eventos na educação.

Inteligência Artificial na Educação

A IA está presente nos diversos segmentos da sociedade e em determinadas situações do cotidiano das pessoas, sendo indiscutível a dependência nessa tecnologia. Essa inteligência trabalha com algoritmos presentes nos *softwares* e programas; o que antes poderia ser analisado e resolvido pelo pensamento crítico e reflexivo do ser humano, agora os algoritmos poderão decidir conforme os comandos e codificações recebidas, com processamento de informações geralmente de maneira automática e repetida para situações semelhantes (Dakakni; Safa, 2023).

A expressão inteligência artificial foi cunhada em 1956 por J. McCarthy (Senai, 2018). Como um ramo da ciência da computação, essa inteligência se caracteriza por utilizar algoritmos que buscam simular o raciocínio humano, facilitando o processo de ensino e aprendizagem e auxiliando na resolução de problemas. Além disso, oferece soluções e caminhos pré-definidos, baseados em um banco de dados que precisa ser continuamente alimentado, atualizado e treinado com novas informações e dados, tanto externos quanto gerados pelo próprio sistema (Melo, 2019).

Em se tratando da IA nas escolas, a Base Nacional Comum Curricular (BNCC) ressalta, entre suas competências, a integração da cultura digital nas salas de aula. A IA pode desempenhar um papel fundamental nesse processo, destacando-se a necessidade de que os alunos tenham acesso livre a essas tecnologias para estimular o senso crítico e o desenvolvimento de habilidades e técnicas essenciais. Isso contribui para o aprimoramento de seu conhecimento e favorece o crescimento pessoal, escolar, acadêmico, profissional e sociocultural dos estudantes (Góes; Porto, 2023).

Acrescenta-se, ainda, que o uso da IA pode oferecer diversas vantagens para a educação, como a agilidade nos processos e o fornecimento de *feedbacks* mais rápidos e eficientes tanto para a gestão das instituições de ensino quanto para o trabalho pedagógico, o processo de ensino-aprendizagem – pesquisas,

modelos de textos, listas de exercícios, sugestões de atividades, avaliações – e a personalização da educação. Essa tecnologia desempenha um papel importante nos processos e planejamentos, tornando o trabalho mais eficiente e adaptável, além de proporcionar experiências criativas e inovadoras. Como uma aliada, a IA tem o potencial de transformar positivamente o cotidiano escolar e acadêmico dos estudantes e dos professores (Ferreira et al., 2023; Figueirôa, 2023).

Além disso, em relação ao rápido desenvolvimento da IA e do aprendizado de máquina, as discussões e pesquisas estão enfrentando um impasse em relação ao desenvolvimento das tecnologias de IA, entre comemorar ou ter prudência sobre os resultados desse desenvolvimento. Nessa perspectiva, a investigação de Ito-Jaeger et al. (2023, p. 2) aponta a necessidade de coparticipação de determinados setores da sociedade, em especial dos envolvidos em educação, no sentido de "incorporar as suas ideias nas tecnologias para torná-las mais seguras e mais confiáveis".

Em síntese, este tópico apresenta alguns aspectos importantes e benefícios da IA na educação:

- Facilitação do ensino e da aprendizagem;
- Auxílio na resolução de problemas;
- Estímulo ao senso crítico e ao desenvolvimento de habilidades e técnicas essenciais dos alunos;
- Agilização dos processos;
- *Feedbacks* mais rápidos e eficientes para a gestão das instituições de ensino, para o trabalho pedagógico e para o processo de ensino-aprendizagem;
- Plataformas e ferramentas tecnológicas facilitadoras dos procedimentos para a realização de pesquisas, obtenção de modelos de textos e de listas de exercícios, sugestões de atividades, sistemas de avaliações etc.;
- A personalização da educação;
- Tornar o trabalho mais eficiente e adaptável;
- Proporcionar experiências criativas e inovadoras;
- Transformar positivamente o cotidiano escolar e acadêmico dos estudantes e dos professores;
- Rápido e constante desenvolvimento da IA e do aprendizado de máquina.

Por fim, além dos inúmeros benefícios da IA para a educação, essa inteligência pode apresentar grandes desafios para os processos educativos, a exemplo do surgimento e da incorporação dos vieses algorítmicos, quando os mesmos potencializam as injustiças e desigualdades sociais. São preocupações que podem surgir quando o treinamento de algoritmos e o aprendizado de máquina se utilizam de um grande volume de dados fornecidos pelos usuários, com potencial de incorporar indevidamente nesse treinamento os preconceitos sociais, que podem ser reproduzidos pelos sistemas tecnológicos automatizados ou não e acabarem sendo utilizados na educação (Williams, 2024; Tang; Su, 2024).

Treinamento e Desenvolvimento de Algoritmos

Em relação ao treinamento de algoritmos, esse procedimento pode ocorrer através da inserção ou repetição de dados no momento da busca de informações e pesquisas, em geral presentes em serviços na internet, plataformas, *sites* e programas. É importante notar que, cada vez mais, esses algoritmos estão sendo treinados com dados fornecidos pela população em geral. Esse treinamento pode incluir perguntas; textos; rótulos; anotações e dados completos ou incompletos, verdadeiros ou não, e que muitas vezes não são verificados ou validados – o que pode introduzir preconceitos e vieses na IA, no aprendizado de máquina e nos demais sistemas computacionais (Sun; Nasraoui; Shafto, 2020).

Esses vieses algorítmicos não surgem apenas durante o treinamento e desenvolvimento da programação; podem ocorrer, também, fora dessas etapas, pois os modelos algorítmicos são treinados para resolver determinados problemas do mundo real e das interações sociais, mas podem apresentar problemas quando da análise de seus resultados em razão da interação entre os seus dados (Baker; Hawn, 2022).

Fora do ambiente de treinamento, mas na etapa envolvendo o processamento automatizado de informações, Johnson (2020) menciona que muitos resultados obtidos são imprevisíveis e aparentemente não causam nenhum tipo de dano, mas quando reagem com os fenômenos sociais, podem refletir características preconceituosas da cognição humana, como pesquisar genericamente na internet sobre o filho talentoso em detrimento de fazer a pergunta correta sobre a filha superdotada, atribuindo automaticamente uma inteligência masculina inerente; ou *softwares* que sinalizam falsamente uma maior probabilidade de futuros criminosos serem negros em detrimento dos brancos.

Em se tratando dessa automatização, no aprendizado de máquina, os vieses algoritmos podem se correlacionar com dados sensíveis da sociedade. Por isso, os programadores e codificadores também enfrentam uma preocupação com os dados pessoais dos usuários, com programas e códigos voltados à proteção de determinadas operações na internet, principalmente na proteção de grupos sociais vulneráveis, visando impedir que os algoritmos classifiquem e realizem rotinas que utilizem características sociais sensíveis – por exemplo, origem racial – de maneira discriminatória (Johnson, 2020).

Essas questões puderam apresentar um breve panorama de como os dados de treinamento e práticas de desenvolvimento de algoritmos podem introduzir vieses da educação. O treinamento algorítmico se ocupa da inserção de dados e de aperfeiçoamento de procedimentos, ferramentas, tecnologias e rotinas utilizadas, inclusive no cotidiano da aprendizagem; todavia, é necessário considerar a possibilidade de ocorrência de vieses nesses processos, e que afetam negativamente os resultados esperados, causando alguns prejuízos na educação, especialmente para os alunos e para os professores, fortalecendo os preconceitos algorítmicos e potencializando as disparidades – indicando a necessidade urgente de propostas preventivas e reparadoras desses vieses (Gaskins, 2023).

Vieses Algorítmicos e suas Consequências

Inicialmente convém apresentar o que pode ser entendido sobre a expressão viés algoritmo. De acordo com Gaskins (2023), são resultados indesejados e distorcidos de processos automatizados que ocorrem durante o aprendizado de máquina ou nos procedimentos de codificação e desenvolvimento de *softwares* e programas. Eles podem surgir de preconceitos preexistentes dos indivíduos responsáveis por criar e treinar os sistemas de IA e as tecnologias inteligentes, ou, em determinados casos, da operacionalização e da inserção de dados incompletos, falhos ou indevidos na fase de treinamento ou fora dela.

Corroborando isso, de acordo com a investigação de Baker e Hawn (2022), alguns dados chamaram a atenção em razão do aumento significativo do uso da IA e da aprendizagem de máquina em determinados casos envolvendo vieses algoritmos nas tecnologias utilizadas nos processos educativos, acarretando em algumas situações prejuízos e impactos danosos aos envolvidos.

Reforçando essa ideia, a investigação na literatura atual identificou algumas hipóteses de vieses e preconceitos algorítmicos, de acordo com o Quadro 3, que fornece um cenário das suas ocorrências nos trabalhos analisados. Cada ocorrência apresenta alguns aspectos importantes e fundamentais para entender algumas discussões de cada trabalho investigado. Através destes excertos é possível, ao final, compreender as possibilidades de criação de estratégias e práticas preventivas e de combate desses vieses e preconceitos algorítmicos.

Quadro 3. Ocorrências de vieses e preconceitos algorítmicos.

Autores	Vieses e Preconceitos	Ocorrências de vieses e preconceitos nos textos
Baker e Hawn (2022, p. 1.065).	Viés do algoritmo – classificação (racismo)	[...] se os estudantes negros são mais propensos a serem rotulados com o envolvimento em violência escolar do que os estudantes brancos, mesmo pelo mesmo comportamento - então é difícil determinar se um algoritmo funciona igualmente bem para ambos os grupos, ou mesmo encontrar alguma maneira de ter certeza de que o funcionamento do algoritmo não é tendencioso.
Davis, Williams e Yang (2021, p. 5).	Viés do algoritmo – paridade de classificação (gênero)	[...] os trabalhadores da [...], como uma força de trabalho mais ampla em tecnologia, era composta desproporcionalmente por homens. Consequentemente, usando os dados existentes o sistema de contratação aprendeu que os homens eram candidatos preferidos.
Davis, Williams e Yang (2021, p. 6).	Viés do algoritmo – calibração (gênero)	Os negros são mal avaliados com pontuações de risco excessivamente fortes e os brancos são mal avaliados com pontuações de risco excessivamente brandas. Concretamente, isto significa que mais negros acabam na prisão e mais brancos permanecem livres.
Sun, Nasraoui e Shafto (2020, p. 18).	Viés do algoritmo – classificação (novos usuários)	O desequilíbrio de classes também pode surgir do algoritmo que pede intencionalmente aos usuários que avaliem apenas os itens mais populares, uma estratégia comum usada para coletar classificações iniciais para novos usuários.
Johnson (2020, p. 9.952).	Viés do algoritmo – problema de *proxy* (programação)	[...] seria moralmente problemático permitir que um programa categorizasse os candidatos como elegíveis ou inelegíveis para um empréstimo hipotecário com base na raça desses candidatos. Como resultado, os programadores tentam evitar qualquer dependência explícita da raça, incluindo filtros que impedem o programa de classificação dos indivíduos com base na raça.
Zajko (2021, p. 13).	Preconceito algorítmico (preconceito racial)	Mas, independentemente do problema para o qual a IA está a ser usada, sempre que o preconceito racial (por exemplo) for um problema, o mínimo que um desenvolvedor pode fazer é compreender o que é raça e como a desigualdade racial está estruturada na sociedade.
Gaskins (2023, p. 420).	Algoritmo tendencioso (discriminação)	Esses algoritmos são tendenciosos porque foram escritos para pesar ou descartar deliberadamente certos fatores. Por exemplo, o *software* pode rotineiramente colocar determinados alunos em cursos que não se alinham com as suas necessidades de aprendizagem, o que pode prejudicar o seu crescimento.
Gaskins (2023, p. 420).	Algoritmo opressivo (preconceito racial)	A utilização de algoritmos, o reconhecimento facial e a vigilância têm o potencial de ter impactos muito negativos sobre os estudantes étnicos sub-representados. A tecnologia de IA pode exacerbar o preconceito racial, o que afeta ainda mais os estudantes e as comunidades que são, na maioria das vezes, mal atendidas e subestimadas nas escolas.
Gaskins (2023, p. 421).	Viés do algoritmo (preconceito)	O robô classificou fotos que foram rotuladas ou marcadas com informações sobre características faciais específicas, assim a IA foi codificada com preconceitos sobre o que é ou o que define a beleza. Na educação, isto tem implicações potencialmente prejudiciais para os testes padronizados e o reconhecimento facial.
Yoder-Himes et al. (2022, p. 2).	Vieses algorítmicos (preconceitos sistêmicos)	É da maior importância que as universidades identifiquem e não utilizem tecnologias que prejudicam as mulheres ou os estudantes negros, especialmente no que diz respeito às tecnologias relacionadas com a avaliação ou o desempenho acadêmico, como o software de supervisão de testes.

Continua...

Quadro 3. Continuação.

Autores	Vieses e Preconceitos	Ocorrências de vieses e preconceitos nos textos
Archambault (2023, p. 538).	Algoritmo tendencioso (distorção de dados)	A ferramenta também fornece exemplos em que os próprios algoritmos amplificam o preconceito quando fazem previsões mais distorcidas do que os dados de treinamento, [...].
Williams (2024, p. 4).	Viés do algoritmo (treinamento de dados)	Se, historicamente, as alunas fossem menos propensas a se inscrever em cursos de Ciência da Computação, o *chatbot* [...] pode recomendar cursos de humanidades em vez de cursos de ciência da computação para estudantes do sexo feminino com base nas tendências anteriores.
Tang e Su (2024, p. 31).	Viés do algoritmo (transparência)	[...] as decisões educacionais tomadas por modelos preditivos de IA baseados em algoritmos levam a preconceitos e discriminação, mas não estão claros quais recursos e atributos algorítmicos são necessários para reduzir esse preconceito de dados.
Solyst et al. (2023, p. 16).	Viés do algoritmo (mitigação de danos)	Os participantes também acreditavam que, na maioria dos casos, os criadores da tecnologia (por exemplo, as empresas) eram responsáveis por corrigi-la, especialmente em cenários de maior dano. Para cenários mais sutis, eles acreditavam que os usuários também tinham a capacidade de contornar o comportamento problemático da IA.
Bird, Castleman e Song (2024, p. 21).	Viés do algoritmo (análise preditiva –desigualdade)	A análise preditiva tem o potencial de melhorar a capacidade das instituições de direcionar estes recursos para os estudantes que mais necessitam de assistência, mas, como mostram as nossas análises, o viés algorítmico pode fazer com que estudantes negros em risco recebam menos apoio do que estudantes brancos semelhantes.
Ito-Jaeger et al. (2023, p. 2).	Preconceito algorítmico (coparticipação dos interessados)	As questões relacionadas com a proteção de dados, segurança *online* e preconceitos algorítmicos são questões muito complexas e, portanto, as partes interessadas devem estar ativamente envolvidas nas discussões.
Sanusi et al. (2023, p. 7).	Preconceito algorítmico (treinamento de dados)	A resposta dos alunos sugere a sua compreensão do preconceito do algoritmo, na medida em que reconhecem que o preconceito pode surgir de decisões relacionadas com a forma como os dados são recolhidos e selecionados para treinar o algoritmo.
Dakakni e Safa (2023, p. 3).	Viés do algoritmo (ética e manipulação da informação)	[...] isto levanta uma questão ética relativa à manipulação e disseminação de algoritmos por poderosos senhores da tecnologia [...] e podem decidir quais algoritmos lançar, para quem, quando e onde apenas com um clique de um botão.

Fonte: Autoria própria.

Esses dados apresentados no Quadro 3 refletem a importância das pesquisas envolvendo esses temas e a relevância dos debates contínuos entre os respectivos setores da sociedade na busca de soluções viáveis e urgentes. A seguir, serão apresentadas mais algumas situações e ocorrências retiradas da literatura investigada e que confirmam a existência de vieses e preconceitos algorítmicos.

Em mais exemplo importante a ser citado, no Reino Unido, em 2020, um regulador nacional de qualificações de notas e um sistema automatizado se utilizaram de uma base interna de previsões dos professores para classificação dos alunos em razão de suas notas, onde o algoritmo do sistema atribuiu notas baixas para os alunos de escolas do estado e notas melhores – superior à previsão dos professores – para os alunos das escolas independentes (Baker; Hawn, 2022).

E mais, os vieses podem se manifestar em razão do uso de determinados *softwares*, a exemplo da tecnologia de monitoramento remoto de exames, potencialmente responsáveis pela prática de racismo sistêmico. Esses problemas surgem através da programação preconceituosa de seus algoritmos e do estabelecimento de políticas e regras internas tendenciosas para seus alunos (Yoder-Himes et al., 2022).

A respeito do uso desses *softwares*, de acordo com Bird, Castleman e Song (2024), surge uma preocupação se o preconceito racial inserido nos algoritmos de previsão envolvendo sistemas de análise preditiva poderiam prejudicar os esforços das instituições de ensino superior (IES) em garantir uma maior equidade racial.

Ainda relacionado às questões envolvendo *softwares*, Yoder-Himes et al. (2022) discutem em sua pesquisa – envolvendo uma universidade do ensino superior nos Estados Unidos – a respeito de algumas disparidades e diferenças de tratamentos em relação a eventuais preconceitos, em especial envolvendo reconhecimento facial e relacionado à análise de cor da pele, raça e sexo encontrados nos resultados do *software* de supervisão automatizado de exames para detecção de desempenho e avaliação dos alunos. Os dados revelaram vieses algorítmicos tendenciosos contra alunos com tons de pele mais escuros, alunos negros e alunas do sexo feminino com tons de pele mais escuros. Isso serve de alerta para a importância de se examinar profundamente esses vieses para garantir a equidade e a justiça no tratamento dos alunos, especialmente dos grupos potenciais.

Outros trabalhos analisados também discutem sobre sistemas tecnológicos e *softwares*, especificamente sobre os vieses e preconceitos algorítmicos na IA e no aprendizado de máquina. A IA, através de suas novas tecnologias e funcionalidades, está cada vez mais presente nos *softwares* utilizados no cotidiano da educação, como pesquisas e atividades escolares, *chatbots*[1], critérios para ingresso de alunos, aconselhamento estudantil, escolha de materiais didáticos e sistemas avaliativos com o objetivo de auxiliar toda a comunidade escolar (Gaskins, 2023).

Essas funcionalidades da IA podem trazer alguns riscos, como bem explica Gaskins (2023) quando faz um alerta sobre os perigos de uso intensivo de dispositivos, ferramentas e plataformas relacionadas à IA e ao aprendizado de máquina. Isso porque essas tecnologias podem potencializar os efeitos dos preconceitos existentes, tais como idioma, raça, rendimento individual; e causarem danos à saúde devido ao tempo excessivo de uso pelos alunos e efeitos de um eventual isolamento social.

Além de tudo isso, pode-se citar a personalização do ensino-aprendizagem, pois ela se insere em um impasse sobre os seus benefícios educacionais e os seus riscos da ocorrência de vieses; e a constatação de que os sistemas baseados em IA podem criar níveis de padronização diferenciados entre os alunos à medida que tentam adaptar cada conteúdo à capacidade cognitiva de cada um, podendo evidenciar as desigualdades tendenciosas entre os mesmos. Essa padronização indevida também pode ocorrer quando o desenvolvimento das competências digitais desses alunos não são consideradas ou avaliadas para a implementação desse ensino personalizado (Dakakni; Safa, 2023).

Todas essas possibilidades de ocorrência de vieses e preconceitos algorítmicos não descaracterizam a importância das inovações tecnológicas e da IA para o cotidiano da educação, apenas sinalizam que algo pode estar errado quando os resultados não refletem positivamente aquilo que era esperado. Por isso a importância de incentivar debates e pesquisas, do envolvimento técnico e profissional, de conscientização e de orientação na busca de soluções. Nesse sentido, o próximo tópico encaminha algumas recomendações de soluções encontradas nos trabalhos investigados e que podem responder sobre as dualidades problematizadas neste trabalho.

Discussões: Dualidades e Recomendações

Este tópico faz uma análise sobre as dualidades entre o uso da IA na educação e os riscos de incorporação de vieses algorítmicos nos processos educativos. Serão apresentadas as questões a respeito dos usos das inovações tecnológicas e da busca por um ambiente educacional justo, ético e socialmente protegido, com a apresentação de recomendações para mitigação desses vieses em razão de cada trabalho analisado.

Iniciando a discussão pela questão do treinamento e da implementação de algoritmos. Para prevenir os preconceitos e vieses algorítmicos, destaca-se a importância de se identificar e mitigar essas ocorrências nas etapas iniciais do desenvolvimento, do treinamento e da implementação dos algoritmos; nas fases de aprendizagem automática e durante o recolhimento e tratamento dos dados – destacando-se a necessidade de atender certos cuidados em relação à qualidade de coleta, ao tratamento, à inserção e à correção dos dados (Baker; Hawn, 2022).

Em se tratando desses cuidados, Davis, Williams e Yang (2021) propõem a reparação algorítmica, deixando explícita a importância de se estabelecer uma ideia de reparação, pois as tecnologias não apenas libertam, mas refletem as dinâmicas sociais, potencializando as relações hierárquicas existentes. Pensar nessas questões sociais é pensar em possibilidades de reparação para tentar sanar as desigualdades estruturais, se utilizando de conhecimentos teóricos, técnicos e da própria tecnologia.

Ou seja, mencionar sobre fenômenos e dinâmicas sociais é tratar sobre os seus reflexos, por exemplo, o preconceito social, pois é ideal que a sua possibilidade de ocorrência esteja prevista no desenvolvimento da IA e dos seus códigos, considerando vocabulários, conceitos, teorias, filtros e funcionalidades que busquem a solução para esse preconceito e projetem um contexto social que não reproduza as desigualdades, e evite os jogos de poder institucionalizados e a prevalência de grupos em detrimento de outros (Zajko, 2021).

Além disso, a reparação se apresenta como um dever em tentar sanar as estratificações históricas e estruturais que surgem durante a criação e utilização dos códigos computacionais; uma necessidade preventiva e restaurativa de tornar os algoritmos mais justos e equitativos para as pessoas e situações implicadas (Davis; Williams; Yang, 2021).

Ainda, em se tratando de reparação algorítmica, esses autores propõem uma ideia de anticlassificação dos envolvidos quando ocorrer previsões e estimativas algorítmicas, ou seja, uma reparação que não considera os aspectos relacionados à idade, ao gênero, à raça, classe social e às questões relacionadas à saúde. Essa perspectiva reforça a indiferença em relação às características dos indivíduos propensos aos resultados automatizados, evitando, dessa maneira, hierarquias sociais nos processos de seleção e coleta de dados.

A segunda proposta de reparação de Davis, Williams e Yang (2021) revela uma tentativa de solucionar a aplicação de "pontuações de risco" e de "equivalência entre grupos" que se utilizam da calibração dos resultados, os quais devem ser independentes. As particularidades relacionadas à identidade dos indivíduos só "devem ser consideradas por uma equação algorítmica se essas características tiverem efeitos empíricos demonstráveis sobre o resultado em consideração" (Davis; Williams; Yang, 2021, p. 6).

Complementando com mais uma sugestão de reparação algorítmica, Davis, Williams e Yang (2021) sugerem a curadoria de arquivos, informações, textos e dados como potencial para uma reparação, tentando, assim, afastar as possibilidades de surgimento de algoritmos injustos contidas em determinados processos relacionados à origem e à coleta de dados. Outra hipótese é chamada de poder distribuído de IA, no qual há a preocupação de reduzir as assimetrias de poder entre os criadores dessa tecnologia e os afetados por ela; uma interação entre os desenvolvedores e a comunidade usuária e potencialmente atingida.

Em outra proposta de solução encontrada, Zajko (2021) sugere a interdisciplinaridade nas fases iniciais de desenvolvimento dos algoritmos e sistemas de IA. Duas ou mais disciplinas com visões, conceitos e teorias próprias sobre o assunto, se utilizando de linguagens diferentes para resolver o mesmo problema; cada disciplina com o seu discurso e contexto de aplicação, numa interação de conhecimentos e combinação de esforços, objetivos e justiça para enfrentar as desigualdades sociais organizadas e combater as estruturas históricas de poder.

Em se tratando do desenvolvimento de *softwares* como proposta mitigadora desses vieses, é importante enfrentar os problemas e os preconceitos observados nos procedimentos envolvendo os *softwares*

de supervisão de exames dos alunos, principalmente quando envolver os grupos marginalizados e sub-representados. Além de buscar a justiça e a equidade nos testes e nas avaliações e combater eventuais barreiras encontradas no ensino básico e no ensino superior. As empresas desenvolvedoras de *softwares* de supervisão devem se atentar para as limitações de seus programas e coordenar a execução de testes prévios para confirmarem a qualidade de seus produtos, para que não haja tratamento desigual e preconceituoso de qualquer natureza entre os seus alunos (Yoder-Himes et al., 2022).

Também pode-se destacar a necessidade de envolvimento desses desenvolvedores de *softwares* visando à criação de programas próprios para registrar e incorporar o *feedback* da utilização de IA pelos jovens usuários, oportunidades de compartilhamento de informações entre os criadores e usuários dos aparatos e sistemas tecnológicos. Por fim, incentivar o uso da IA de uma maneira crítica, ética, transparente, responsável e inclusiva por parte desses jovens, mais a importância de abertura de canais de comunicação e de alertas sobre os potenciais riscos dos vieses algorítmicos (Solyst et al., 2023; Davis; Williams; Yang, 2021).

A respeito da importância desse envolvimento para a comunidade escolar, Baker e Hawn (2022) destacam a necessidade de participação de determinados profissionais e grupos da sociedade durante todo o processo de desenvolvimento e testes de uso de algoritmos na educação, pois isso pode proporcionar benefícios significativos e ajudar a alcançar interpretações mais exatas dos dados, gerar informações mais consistentes em razão dos fenômenos observados e resultar em concepções e intervenções mais favoráveis para a comunidade escolar. Eles concluem sobre a necessidade de buscar reduzir os preconceitos desconhecidos, registrar e documentar os preconceitos conhecidos e trabalhar em conjunto para reparar os vieses algorítmicos.

Relacionado à IA como proposta de conscientização de riscos, no texto de Solyst et al. (2023), os autores destacam a importância de conscientização e formação dos alunos no sentido de tentar reconhecer os riscos potenciais de determinados vieses algorítmicos. Ainda, salientam a importância da participação e da colaboração coletiva dos segmentos da sociedade envolvidos na busca de soluções para prevenir os riscos e corrigir os reflexos desses vieses.

E mais, a IA e o aprendizado de máquina surgem como preocupações de seu potencial em criar e aumentar as desigualdades e os preconceitos através de suas funcionalidades e recursos, necessitando de vigilância ética e de segurança dos dados dos usuários, pois os vieses algorítmicos na IA podem surgir a partir da inserção de dados falhos, incompletos, preconceituosos e tendenciosos durante a fase de treinamento ou fora dela. Como medidas de mitigação, destaca-se a necessidade de uma atenção especial voltada ao desenvolvimento e uso de programas, e acompanhamento das fases dos processos manuais e automatizados dos algoritmos.

Bem como é extremamente importante que as instituições de ensino possam supervisionar as suas plataformas de IA para tentar identificar os riscos, avaliar as suas aplicabilidades e, se necessário, buscar soluções para combater o viés algorítmico que pode estar presente nesses sistemas – consequentemente, contribuindo para a conquista de uma equidade e justiça racial (Bird; Castleman; Song, 2024).

A respeito da dualidade envolvendo o uso de *chatbots* e eventuais riscos dessa utilização, os alunos precisam ser orientados a adotar transparência e incentivar debates contínuos em relação à ocorrência de eventuais preconceitos surgidos em decorrência do uso e da interação com os *chatbots*, além de serem capacitados e críticos no sentido de analisarem os resultados e não aceitarem as recomendações como verdades absolutas e/ou como julgamentos já decididos e impositivos (Williams, 2024).

Além disso, é recomendável uma navegação confiável e segura em se tratando da utilização da IA e dos *chatbots* generativos na educação, pois eles apresentam um potencial significativo em transformar os processos de ensino e aprendizagem. No entanto, apesar dessa segurança e confiabilidade necessárias, esse uso pode esbarrar em questões éticas que precisam ser cuidadosamente analisadas; além dos riscos relacionados

à privacidade dos dados dos alunos e ao preconceito algorítmico, à dependência excessiva dessas tecnologias e a estar sujeito ao plágio (Williams, 2024; Tang; Su, 2024).

Em se tratando das vantagens do uso do ChatGPT[2] numa perspectiva dos *chatbots*, a IA pode utilizar de sua capacidade preditiva como auxílio nas pesquisas e também para identificar padrões dos usuários durante o seu aprendizado de máquina. Os alunos podem se beneficiar dessas tecnologias, inclusive pelo fato de complementarem os métodos educativos tradicionais, pois elas têm o potencial de aumentar a acessibilidade e proporcionar experiências de aprendizagem personalizadas. Os *chatbots*, a exemplo do ChatGPT, têm a capacidade de oferecer uma tecnologia capaz de estruturar respostas relativamente coerentes e possibilitar interações surpreendentes entre homem-máquina, e consequentemente trazer certas vantagens administrativas para a gestão escolar e para a concretização dos objetivos do processo de ensino-aprendizagem, visando favorecer os alunos e os professores em diversos aspectos (Williams, 2024; ChatGPT, 2024).

Mais uma hipótese de dualidade que pode ser considerada: a personalização do ensino e da aprendizagem. Os desenvolvedores de tecnologias educacionais que treinam os algoritmos através de dados para a personalização do aprendizado buscam desenvolver e implementar códigos para a identificação de alunos em riscos e criar estratégias tecnológicas para facilitar o cotidiano docente. No entanto, se os preconceitos potenciais que podem afetar esses dados não forem examinados pelos programas de IA, eles poderão ser amplificados, a exemplo dos preconceitos raciais numa perspectiva do reconhecimento facial. Em algumas circunstâncias, os algoritmos de autenticação, de vigilância e de reconhecimento facial podem ser tendenciosos e reforçar as desigualdades existentes, causando danos aos estudantes vulneráveis e sub-representados, pois cada vez mais eles se utilizam de sistemas de IA em suas atividades diárias da escola e de pesquisa e necessitam utilizar tecnologias justas e confiáveis (Gaskins, 2023; Yoder-Himes et al., 2022).

Além disso, o desenvolvimento de sistemas computacionais, de IA e de algoritmos de aprendizagem de máquina impulsionaram os alunos e professores a replanejarem os seus métodos educacionais tradicionais, pois essas tecnologias favorecem a criação e implementação de estratégias personalizadas de ensino e aprendizagem, mais acessíveis e eficientes (Williams, 2024).

Essa personalização pode contribuir para a identificação dos pontos fortes e fracos e do progresso dos alunos. Ela poderá ser adaptada ao que o aluno já sabe e às suas novas necessidades de aprendizagem de acordo com as respectivas habilidades de cada um. Ainda, essa personalização também poderá influenciar positivamente os resultados das avaliações dos alunos, pois os conteúdos foram repassados de maneira coerente (Dakakni; Safa, 2023).

Entretanto, não se podem negar os desafios para equilibrar as vantagens da educação personalizada com os direitos de sigilo dos dados e das informações pessoais dos alunos. Por isso, na utilização de novas tecnologias de IA e de aprendizado de máquina, esses aspectos precisam ser considerados – além de serem altamente preocupantes e fazerem parte de um trabalho constante dos desenvolvedores, visando criar filtros e proteção para esses sistemas e programas (Williams, 2024).

Apesar disso tudo, as funcionalidades da IA podem proporcionar certa autonomia de estudos para os alunos, se tornando uma alternativa para a educação personalizada e voltada aos aprendizes. No entanto, essa autonomia carece de supervisão e orientação dos responsáveis, principalmente dos professores e dos familiares, para que todos possam ser capazes de discutir propostas de mitigação para os riscos potenciais em relação aos eventuais danos pessoais e psicológicos (Williams, 2024).

Resumindo, a partir da análise e interpretação do *corpus* da pesquisa e das dualidades encontradas, os autores analisados apresentaram as seguintes sugestões para o enfrentamento e mitigação dos riscos de incorporação de vieses algorítmicos na educação: I) mitigação dos vieses nas etapas iniciais do desenvolvimento, do treinamento e da implementação dos algoritmos; II) ideia de reparação algorítmica (anticlassificação,

pontuações de risco e equivalência entre grupos e curadoria de arquivos); III) interdisciplinaridade nas fases iniciais de desenvolvimento dos algoritmos e sistemas de IA; IV) necessidade de desenvolvimento de *softwares* de supervisão de exames dos alunos; V) abertura de canais de comunicação e de alertas sobre os potenciais riscos; VI) conscientização dos alunos no sentido de tentar reconhecer os riscos potenciais; VII) desenvolvimento e implementação de programas e acompanhamento das fases dos processos manuais e automatizados dos algoritmos; VIII) necessidade das instituições de ensino supervisionarem as suas plataformas de IA; IX) incentivo aos debates contínuos; X) recomendação de uma navegação confiável e segura; XI) vantagens do uso do ChatGPT; XII) necessidade de tecnologias justas e confiáveis; e XIII) educação personalizada.

Dito isso, em sua pesquisa, Zajto (2021) constatou, sobre a reprodução de vieses e do preconceito social pelos sistemas de IA, que um algoritmo não tem muitas chances por si só de tornar o mundo um lugar melhor e que as funcionalidades automatizadas dos computadores não deixarão de causar distinções e danos aos seus usuários, reproduzindo desigualdades e injustiças sociais preexistentes e necessitando, assim, da colaboração de todos os setores da sociedade na busca de soluções para os problemas encontrados no cotidiano tecnológico educacional – alguns deles apresentados neste trabalho.

Considerações Finais

A revisão sistemática e análise da literatura realizadas neste estudo confirmaram a existência de dualidades entre o uso da IA na educação e os riscos de inserção e incorporação de vieses algorítmicos nos processos educativos. Esses vieses, que podem surgir devido a dados de treinamento enviesados e práticas de desenvolvimento inadequadas, têm implicações profundas para a equidade e eficácia do ensino. Os efeitos desses vieses são multifacetados, afetando grupos de estudantes e podendo perpetuar desigualdades e prejudicar a inclusão educacional. "Devido aos dados de treinamento, foi demonstrado que os modelos de IA apresentam preconceito e discriminação, o que reforça estereótipos inerentes" (Tang; Su, 2024, p. 29).

A presença de vieses algorítmicos na educação pode ter profundas consequências, afetar o desempenho e o comportamento dos alunos e retardar ou dificultar certos processos educativos. Os dados inapropriados ou incompletos no treinamento dos algoritmos podem resultar em decisões automatizadas que ocasionarão desigualdades ou injustiças sociais, favorecendo certos grupos em detrimento de outros.

Os estudos analisados revelam que os vieses algorítmicos podem influenciar desde a avaliação de desempenho acadêmico até as recomendações personalizadas de aprendizagem, potencialmente distorcendo as oportunidades educacionais e agravando disparidades existentes. Consequentemente, esses vieses podem ter efeitos em longo prazo, impactando o progresso acadêmico e as futuras oportunidades profissionais dos estudantes.

Por outro lado, a popularização e o avanço contínuo da inteligência artificial têm trazido avanços na democratização do treinamento desses sistemas, e à medida que mais dados diversificados e representativos são incorporados nos algoritmos, espera-se que esses vieses sejam minimizados – uma vez que o treinamento contínuo e a atualização dos modelos com dados mais inclusivos são passos cruciais para reduzir os vieses. Isso se alinha com o progresso observado em outros setores da tecnologia e da sociedade, onde a conscientização e a melhoria contínua têm ajudado a mitigar os impactos negativos dos vieses algorítmicos.

No entanto, apesar dessas melhorias decorrentes da popularização dos bancos de dados utilizados pelos algoritmos, é importante reconhecer que os vieses algorítmicos são um reflexo de preconceitos e

desigualdades existentes na sociedade e, portanto, sua completa eliminação é difícil de acontecer. Como exemplo, pode-se citar a presença de vieses também em outros setores, como saúde, justiça e mercado de trabalho, onde decisões automatizadas podem influenciar significativamente a vida das pessoas. Assim, é fundamental que educadores, desenvolvedores de tecnologia e formuladores de políticas públicas continuem a trabalhar juntos para criar sistemas mais justos e transparentes, promovendo práticas que garantam uma educação mais eficaz e mais igualitária.

Diante dos dados da literatura investigada, não foi possível descartar a existência dos riscos e dos desafios da inserção de vieses algorítmicos nos processos educativos. As sugestões apresentadas indicam um caminho possível para a mitigação dos problemas. Em suma, por um lado, as inovações tecnológicas sendo utilizadas em larga escala na educação, contribuindo para a troca de significados e de conhecimentos em sala de aula; por outro lado, a necessidade de adoção de práticas e estratégias voltadas à ética, equidade, justiça, inclusão, dignidade e proteção digital dos alunos.

Por fim, como resposta à hipótese, conclui-se que a eficácia da IA na educação depende de um treinamento contínuo e inclusivo dos algoritmos e da conscientização de seus usuários sobre os riscos dos vieses algorítmicos, de forma que os sistemas possam ser aperfeiçoados para evitar a reprodução de injustiças sociais. Por isso, as discussões e pesquisas sobre o tema não podem parar.

Contribuição dos Autores

Conceptualização: Heggler JM, Szmoski RM; **Metodologia:** Heggler JM, Szmoski RM; **Supervisão:** Szmoski RM, Miquelin AF; **Escrita – rascunho original:** Heggler JM; **Escrita – análise e edição:** Szmoski RM, Miquelin AF; **Aprovação da versão final:** Szmoski RM, Miquelin AF.

Agradecimentos

Universidade Tecnológica Federal do Paraná – Ponta Grossa-PR.

Disponibilidade e Dados da Pesquisa

Todos os dados foram gerados/analisados no presente artigo.

Notas

1. Também conhecido pelo termo *chatterbot*, "é uma junção das palavras chatter (pessoa que conversa) e bot (abreviatura de robot – robô)" (Moraes; Souza, 2015, p. 601).

2. Essa tecnologia foi disponibilizada no Brasil para uso público em novembro de 2022 pela Empresa OpenAi. Ela disponibiliza as funcionalidades de pesquisa e de consulta aos usuários através de perguntas e respostas. Essa tecnologia surgiu como uma espécie da IA para incrementar e otimizar a interação homem-máquina e transformar a sociedade (CHATGPT, 2024).

Referências

ARIA, M.; CUCCURULLO, C. Bibliometrix: An R-tool for comprehensive science mapping analysis. **Revista de Informetrics**, v. 11, n. 4, p. 959-975, 2017. https://doi.org/10.1016/j.joi.2017.08.007

ARCHAMBAULT, S. G. Expanding on the frames: Making a case for algorithmic literacy. **Communications in Information Literacy**, v. 17, n. 2, p. 530-553, 2023. https://doi.org/10.15760/comminfolit.2023.17.2.11

BAKER, R. S.; HAWN, A. Algorithmic bias in education. **International Journal of Artificial Intelligence in Education Society**, p. 1-41, 2022. https://doi.org/10.1007/s40593-021-00285-9

BIRD, K. A.; CASTLEMAN, C. L.; SONG, Y. Are algorithms biased in education? Exploring racial bias in predicting community college student success. **Journal of Policy Analysis and Management**, p. 1-24, 2024. https://doi.org/10.1002/pam.22569

CHATGPT. **Plataforma ChatGPT**. Open AI. 2024. Disponível em: https://chat.openai.com/. Acesso: em 13 mar. 2024.

DAKAKNI, D.; SAFA, N. Artificial intelligence in the L2 classroom: implications and challenges on ethics and equity in higher education: A 21st century Pandora's Box. **Computers and Education:** Artificial Intelligence, v. 5, p. 1-10, 2023. https://doi.org/10.1016/j.caeai.2023.100179

DAVIS, J. L.; WILLIAMS, A.; YANG, M. W. Algorithmic reparation. **Big Data & Sociedade**, v. 2, p. 1-12, 2021. https://doi.org/10.1177/20539517211044808

FERREIRA, J. M. et al. A inteligência artificial na educação: a tecnologia como aliada da educação a distância. **Revista Amor Mundi**, v. 4, n. 6, p. 143-157, 2023. https://doi.org/10.46550/amormundi.v4i6.282

FIGUEIRÔA, L. M. Uma revisão literária das vantagens e desvantagens da inclusão da inteligência artificial (IA) no ensino a distância. **Revista Amor Mundi**, v. 4, n. 4, p. 81-89, 2023. https://doi.org/10.46550/amormundi.v4i4.221

GASKINS, N. Interrogating Algorithmic bias: from speculative fiction to liberatory design. **TechTrends**, v. 67, n. 3, p. 417-425, 2023. https://doi.org/10.1007/s11528-022-00783-0

GÓES, D.; PORTO, C. M. O ChatGPT: a tecnologia a serviço do aluno e do professor em sala de aula. **Simpósio Internacional de Educação e Comunicação**-SIMEDUC, n. 11, 2023. Disponível em: https://eventos.set.edu.br/simeduc/article/download/16300/14705/61252. Acesso em: 14 fev. 2024.

GOMEDE, E.; BARROS, R. M.; MENDES, L. S. Deep auto encoders to adaptive E-learning recommender system. **Computers and education:** Artificial intelligence, v. 2, p. 1-24, 2021. https://doi.org/10.1016/j.caeai.2021.100009

ITO-JAEGER, S. et al. TrustScapes: a visualisation tool to capture stakeholders' concerns and recommendations about data protection, algorithmic bias, and online safety. Disponível em: **International Journal of Qualitative Methods**, v. 22, p. 1-10, 2023. https://doi.org/10.1177/16094069231186965

JOHNSON, G. M. Algorithmic bias: on the implicit biases of social technology. **Synthese**, v. 198, n. 10, p. 9.941-9.961, 2020. https://doi.org/10.1007/s11229-020-02696-y

MELO, M. A. V. **Inteligência Artificial e ensino de inglês como língua estrangeira:** inovação tecnológica e metodológica/de abordagem? 153f. Dissertação (Mestrado em Estudos Linguísticos do Instituto de Letras e Linguística) – Universidade Federal de Uberlândia. Uberlândia, Minas Gerais, 2019. Disponível em: https://repositorio.ufu.br/handle/123456789/26726. Acesso em: 13 jun. 2023.

MORAES, S. M. W.; SOUZA, L. S. de. Uma abordagem semiautomática para expansão e enriquecimento linguístico de bases AIML para chatbots. In: CONGRESSO INTERNACIONAL DE INFORMÁTICA EDUCATIVA, 2015. **Anais TISE 2015**. Santiago: Universidad de Chile, 2015. Disponível em: https://www.tise.cl/volumen11/III.html. Acesso em: 14 mar. 2024.

SANUSI, I. T. et al. Developing middle school students' understanding of machine learning in an African school. **Computers and Education:** Artificial Intelligence, v. 5, p. 1-11, 2023. https://doi.org/10.1016/j.caeai.2023.100155

SENAI. Serviço Nacional de Aprendizagem Industrial. Departamento Nacional. **Tendências em inteligência artificial na educação no período de 2017 a 2030**. SUMÁRIO EXECUTIVO/Serviço Nacional de Aprendizagem Industrial, Serviço Social da Indústria, Rosa Maria Vicari. Brasília: SENAI, 2018. 52 p. Disponível em: https://www2.fiescnet.com.br/web/uploads/recursos/d1dbf03635c1ad8ad3607190f17c9a19.pdf. Acesso em: 2 mar. 2024.

SOLYST, J. et al. The potential of diverse youth as stakeholders in identifying and mitigating algorithmic bias for a future of fairer AI. **Proceedings of the ACM on Human**-Computer Interaction, v. 7, n. CSCW2, p. 1-27, 2023. https://doi.org/10.1145/3610213

SUN, W.; NASRAOUI, O.; SHAFTO, P. Evolution and impact of bias in human and machine learning algorithm interaction. **Plos one**, v. 15, n. 8, p. 1-39, 2020. https://doi.org/10.1371/journal.pone.0235502

TANG, L.; SU, Y-S. Ethical implications and principles of using artificial intelligence models in the classroom: A systematic literature review. **Unir La Universidad En Internet** - Repositório Digital, v 8, n. 5, p. 25-36, 2024. https://doi.org/10.9781/ijimai.2024.02.010

WILLIAMS, T. R. The ethical implications of using generative chatbots in higher education. In: **Frontiers in Education**. Frontiers Media SA, 2024. Disponível em: https://www.frontiersin.org/articles/10.3389/feduc.2023.1331607/full. Acesso em: 21 maio 2024.

YODER-HIMES, D. R. et al. Racial, skin tone, and sex disparities in automated proctoring software. In: **Frontiers in Education**. Frontiers, 2022. p. 1-16. https://doi.org/10.3389/feduc.2022.881449

ZAJKO, M. Conservative AI and social inequality: conceptualizing alternatives to bias through social theory. **AI & SOCIETY**, v. 36, n. 3, p. 1.047-1.056, 2021. https://doi.org/10.1007/s00146-021-01153-9

Sobre os Autores

JOÃO MARCOS HEGGLER é doutorando no Programa de Pós-Graduação em Ensino de Ciência e Tecnologia da Universidade Tecnológica Federal do Paraná – PPGECT/UTFPR (2023-2027); pós-graduação em Ensino de Ciência e Tecnologia – PPGECT, mestrado profissional pela UTFPR (2019-2021); pós-graduação em MBA: Gestão Estratégica de Pessoas pela Faculdades Sagrada Família – FASF (2015-2016); bacharelado em Direito pela Universidade Estadual de Ponta Grossa – UEPG (2014, c/ OAB); graduação em licenciatura em Geografia pela UEPG (2005), com Láurea Acadêmica. Pesquisas nos seguintes temas: educação, sustentabilidade socioambiental, ciência e tecnologia, e inteligência artificial.

ROMEU MIQUEIAS SZMOSKI possui doutorado (2013) em Ciências - Física pela UEPG. Atualmente é professor associado da UTFPR e editor da Revista Brasileira de Física Tecnológica Aplicada. Também atua como professor permanente do PPGECT/UTFPR. Tem experiência na área de Física, com ênfase em sistemas dinâmicos e fenômenos complexos, atuando principalmente nos seguintes temas: sincronização de caos, redes complexas, teoria de informação. Trabalha ainda com pesquisas voltadas a energias renováveis e desenvolvimento sustentável e à implementação de tecnologias de informação e comunicação para interação com fenômenos naturais e experimentos físicos.

AWDRY FEISSER MIQUELIN possui graduação em licenciatura em Física pela UEPG (2000), mestrado em Educação pela Universidade Federal de Santa Maria – UFSM (2003) e doutorado em Educação Científica e Tecnológica pela Universidade Federal de Santa Catarina – UFSC (2009). Foi professor QPM do estado do Paraná de 2003 a 2009. Atualmente é professor associado IV no DAENS da UTFPR – *Campus* Ponta Grossa, professor do PPGECT – Mestrado e Doutorado. Foi coordenador institucional do projeto PIBID da UTFPR e atualmente é coordenador do doutorado do PPGECT. Trabalha em projetos interdisciplinares voltados às relações entre ensino; à abordagem sistêmica; ciência e tecnologia e sociedade; ciência e arte e tecnologia; análise de sistemas educacionais tecnológicos comunicativos com ênfase no ensino de ciências; mediação de tecnologias de informação e comunicação na prática pedagógica para o Ensino Superior e a Escola Básica envolvendo didática, ensino e aprendizagem. Coordena o Grupo de Pesquisas Em Arte, Ciência e Tecnologia: GPACT. Bolsista PQ 2 do CNPq.

Recebido: 12 ago. 2024
Aceito: 29 out. 2024