



Interrogating Algorithmic Bias: From Speculative Fiction to Liberatory Design

Nettrice Gaskins¹

Accepted: 11 September 2022 / Published online: 19 September 2022
© Association for Educational Communications & Technology 2022

Abstract

This paper reviews algorithmic or artificial intelligence (AI) bias in education technology, especially through the lenses of speculative fiction, speculative and liberatory design. It discusses the causes of the bias and reviews literature on various ways that algorithmic/AI bias manifests in education and in communities that are underrepresented in EdTech software development. While other recent work has responded to mainstream or private sector technology development, this review looks elsewhere where practitioners, artists, and activists engage underrepresented communities in brainstorming processes to identify and solve tough challenges. Their creative work includes films, toolkits, applications, prototypes and other physical artifacts, and other future-facing ideas that can provide guideposts for private sector development. Acknowledging the gaps in what has been studied, this paper proposes a different approach that includes speculative and liberatory design thinking, which can help developers better understand the educational and personal contexts of underrepresented groups. Early efforts to advocate for fairness and equity in AI and EdTech by groups such as the Algorithmic Justice League, the EdTech Equity Project, and EdSAFE AI Alliance is also explored.

Keywords Algorithmic bias · Artificial intelligence · Educational technology · Machine learning · Speculative design · And speculative fiction

Introduction

Artificial intelligence or AI consists of algorithms—instructions or code—that run repeatedly aiming to predict and complete complex tasks in a matter of seconds. AI apps rely on mountains of data to train algorithms to recognize patterns and make decisions. Much of this data is harvested from users without them realizing it. Sequenced algorithms learn from users' interactions and this process is referred to as machine learning or ML. The ML process involves learning from algorithms to detect patterns that allow machines to make predictions about the future of a given dataset. The more ML data that is collected the more precise the AI can be. This includes personally curated playlists on Spotify, trending tweets on Twitter, or image and audio information on TikTok, which are all complex tasks that would be merely

impossible for humans to do, let alone in a timely manner, hence the increased reliance on AI and ML.

Educators are beginning to see the educational cognitive affordance that AI and ML technology have on student learning. Students' online interactions feed huge data sets of ML at incredible speeds. What they watch on TikTok, search on Google, or the permissions they create when registering for apps are all examples of how ML collates and uses their data. Students spend countless hours online and on social media websites, using a variety of devices to connect to the Internet. This was especially the case during the global Covid-19 pandemic when many school districts across the United States provided Chromebooks or other laptop computers for students to use at home. Developers of *educational technology*, otherwise referred to here as 'EdTech' stepped up their use of AI and ML to create platforms, tools, and devices for student use at home and at school. However, for some researchers, AI brings both promise and danger.

Broussard (2019: 28) points to the tendency of programmers and tech enthusiasts to emphasize the benefits over the risks or downsides of machine learning and other emerging AI technology. AI-powered technology can exacerbate

✉ Nettrice Gaskins
nettrice@gmail.com

¹ Independent Researcher, Lesley University, 1815
Massachusetts Avenue, Cambridge, MA 02140, USA

and amplifying existing biases, which further intensifies the challenges that historically marginalized students face when race intersects with students' other statuses such as language, income, and disability status. Algorithms in AI that source user data have been shown to be biased such as in a case reported by ProPublica that found that an algorithm used in judicial sentencing was biased against African Americans (Angwin et al., 2016). Research by Benjamin (2020), Buolamwini and Gebru (2018), among others call for a more balanced view of AI. Their works illuminate the dangers algorithmic bias in AI poses for vulnerable groups that are often not taken into consideration when developing EdTech tools and devices.

This paper highlights and explores both real life and fictional visions of AI in education, from sava saheli singh's short, near future fiction film “#tresdancing” to other projects that highlight the potential harms of algorithmic or AI bias and invisible surveillance systems that are deeply embedded in peoples' everyday lives. “#tresdancing” features marginalized and underrepresented youth in Canada who, like their U.S. counterparts, face similar levels of racial inequality (Attewell et al., 2009). The film shows examples of AI tools and devices that are used to track, teach and assess students. Frankie, the film's protagonist, represents students in real life who are more vulnerable to bias and surveillance. These students are increasingly being required to use online assessment and distance learning software, as well as emerging technology such as virtual and augmented reality.

This paper makes a case for the need for designers and technologists to use liberatory design methods in the development of educational technologies that are used in our classrooms and that are used by black and brown youth. As a point of interest, the paper looks at projects that address algorithmic bias in AI, which are labeled as *speculative fiction* or belong in the interconnected domains of *speculative and critical design* (Dunne & Raby, 2013, 2021). Speculative fiction encompasses works in which the setting is other than the real world, involving supernatural, futuristic, or other imagined elements (Atwood, 2014). Future-facing storytelling such as “#tresdancing” and speculative design domains have the potential to assist researchers, developers, and teachers in developing language to bring schools and families into the conversation about how EdTech products are designed to ensure fair and transparent products for underserved and underestimated students.

Section 1: History of Racialization/Bias in EdTech Software Development

EdTech tools enable learning through the interactions between students and teachers, or students and devices such as computers. Technologies such as AI, games, virtual and

augmented reality or VR and AR respectively, are examples of *computer-mediated communication* or CMC, which has been noted to amplify disparities in digital communication that emerge around socio-demographics, and Internet skills and experiences (Nguyen et al., 2021). As students have more EdTech access at school and at home, developers are tasked to create more and more tools that, unfortunately, lack the necessary flexibility to handle problems that students face as they learn.

Just under the surface of the diffusion of EdTech software lies persisting digital inequalities marked by disparities across socio-demographic groups in terms of peoples' nature of access, degrees of skill, and varieties of use (Dimaggio et al., 2020; Robinson et al., 2020). These inequalities often augment racial and ethnic disparities (Fuchs & Horak, 2008). In the past, not much has been done by EdTech developers to address these gaps.

Researchers Hebbar and Jacobs (2022) note that many EdTech companies test their products in partnership with schools that have small black, brown, and often low-income populations. According to Karumbaiah and Brooks (2021), and Benjamin (2020) these products use algorithms that are built from historical education data, which often amplifies existing biases, further encoding the racist history of social and academic systems. To address this, Hebbar et al. (2020) created the *AI in Education Toolkit for Racial Equity* that provides practices EdTech developers can implement to uncover and mitigate racial bias. To add, the EdSAFE AI Alliance (2022) asserts that there are, as of now, “no unified approaches to establishing benchmarks,” which can help users discern the quality and reliability of AI, and regulatory bodies contribute to the development of more equitable AI in EdTech.

To fill the void, fictional and non-fictional media address the impact that AI bias has on underrepresented groups, giving members from these groups a voice. The film “#tresdancing” shows how digitally connected youth frequently interact with each other online (Itō et al., 2019); take part in communities of practice and affinity spaces (Gee, 2017); network, exchange, and share resources online (Baym, 2015; Srinivasan, 2018), use their knowledge to solve problems (Nielsen, 2011; Toyama, 2015); engage in distributed learning (e.g., Mawasi et al., 2020; Pinkard et al., 2017), and make use of digital storytelling (e.g., Srinivasan, 2018; Pinkard et al., 2017). Alternative frameworks and tools provided by Bettina Love's “Get to the Future” online resource (2016) and Jeremy Vincent's EdTech company AfroBrainiac (2022) approach AI with fresh questions and envision futures that represent a radical departure from current practices.

The implementation of AI in EdTech has resulted in pedagogical challenges, equity and power issues, inequality issues related to equity, social reproduction of injustice across different contexts and systems, and bias toward

certain groups of users (e.g., Dobson, 2019; Noble, 2018; Reich & Itō, 2017; Srinivasan, 2018; Watkins & Cho, 2018). Racial biases of software developers carry over into modern EdTech (Perry & Turner-Lee, 2020). Building on a synthesis of existing literature, including equity-oriented approaches, this essay addresses three related dimensions: (1) the student-as-consumer; (2) discrimination in EdTech design; (3) and oppressive algorithms.

The Student-as-Consumer Metaphor

The Covid-19 pandemic posed an unanticipated opportunity for EdTech developers to introduce an ever-increasing amount of AI-driven educational tools into the education market. Broadly speaking, there is an understanding that the use of AI for educational purposes has the potential to enhance the learning experience. However, at the same time, there are concerns of potential misuses and ensuing violations of students' rights. Market mechanisms in education systems worldwide have led to the conceptualization of students as passive learners or consumers (Naidoo & Whitty, 2013), including in educational technology (Harrison & Risler, 2015). In the past, developers saw consumerism as an opportunity for education to be simpler to access and more efficient (Brigham, 1993). When students are treated as consumers, their relationship to their educational program or school becomes defined in a particular way. They are distanced from the very educational process which is supposed to engage them (Cheney et al., 1997). This issue is compounded by the lack of diversity in the private sector which coincides with the increasing use of EdTech in classrooms (Lynch, 2018).

The creation of AI-driven tools for K-12 students is still in its nascent stage, and the sector very much decentralized. There are broad guidelines and a plethora of resources, but there is no dominating school of thought for curricula. Researchers (Ali et al., 2019) have recently created curriculum to best prepare students to flourish in the era of AI. Others have drawn attention to the increase, through adoption of AI and machine learning, of racial bias and its harmful impacts. While many studies explore the impact of this development on students in higher education, few have looked at the effects of this on African American, Latino/Latinx, and Indigenous students, who will face more inequality unless educators address the inherent biases of the (mostly white) developers of EdTech tools (Perry & Turner-Lee, 2020; Lynch, 2018). Although underrepresented communities lack access to home broadband (Atske & Perrin, 2022), they are significantly more likely than whites to use mobile devices to go online (Lenhart, 2019).

In the film “#tresdancing” education is placed entirely within the frame of market forces that make students more vulnerable to bias and surveillance. The film persuades

viewers to take a careful look at the implications and limitations of the student-as-consumer metaphor, asserting that this concept can lead us where we really don't wish to go. Youth are expected to be comfortable with technology that invades their privacy, collects their personal information, as well as their uninformed consent. AI tools can be used for ‘demographics targeting, which systematically excludes or exploits certain groups, and enhances harmful racial profiling (Raji et al., 2020). Many EdTech developers lack awareness of the impact of this technology on students who are more vulnerable to exploitation and discrimination.

Discrimination in EdTech Design

Historically, groups such as African Americans have experienced the commodification of their culture (Njee, 2016) and have had to directly, personally understand race as it operates between fictionality (as a fabrication) and materiality (as a thing). Race becomes socially real, and people learn to see themselves as white or other, are treated as white or other, and are motivated by considerations arising out of their group identity (Young, 2006:193). The imbalance of this social construction has led to bias against the less dominant group, or what Fouché (2004:316) refers to as the *problematic of vision* or the ways that dominant analysis and interpretation hinges on the “idea that value, truth, purity, and legitimacy of marginalized individuals and communities must be judged by the standards of dominant society.” Young people from underrepresented groups fall outside of the dominant group's criteria of what it means to be seen as valuable, truthful, pure, or legitimate, and this bias can be applied to the development and use of AI technology and related innovations.

According to scholar Ruha Benjamin technology innovation is often conflated with social progress, meaning society can appear to be moving forward in tech-related fields while re-entrenching old forms of social and racial inequities. Benjamin (2020:5) coined the term “New Jim Code” to describe a specific manifestation of discriminatory design in which racist values and assumptions are built into technical systems. New technologies that do not take bias into account can cause harm for some in the name of efficiency and progress. Automated systems or AI agents create inequity using training data (Koenig, 2019). This data is used to train algorithm or machine learning models to predict outcomes that data scientists design their models to predict. For example, studies show that African American students are subject to disciplinary action at rates much higher than their white counterparts (Riddle & Sinclair, 2019). If local schools and police departments combined their data to identify students who are labeled as at-risk, the resulting intervention could further harm the students, especially if the institutions have poor records of working with the target group.

As a fictional narrative “#tresdancing” takes on trends in educational technology that are happening right now, then pushes the moral and ethical boundaries further to consider what might happen in the future. The film shows how sophisticated algorithms embedded in courseware can be used to identify struggling students in a specific subject, i.e., math. However, these algorithms are only as informed as the developers who design them. These algorithms are biased because they are written to deliberately weigh or discard certain factors. For example, software can routinely place certain students in tracks that don’t align with their learning needs, which could hinder their growth. In “#tresdancing” we see a student attempt and fail to complete an online math assessment at home. In general, learning software that uses AI is mainly designed for students who are represented within the EdTech development space (Bradley, 2021).

Oppressive Algorithms

Dominant or mainstream educational technology discourses often erase opposition to the design and use of code (algorithms, data) in education, rendering the opposing views as invisible, and reducing others’ experiences to irrelevance. Tech fields have done this to underrepresented groups, particularly as it relates to the creation and use of data. Scholar Safiya Umoja Noble (2018) predicted that AI would become a major human rights issue. Noble, along with Joy Buolamwini and the Algorithmic Justice League or AJL examine algorithmic bias in online search engine and computer vision systems that can be trained to track movements of the eye, for example. Their work considers the trade-offs, risks, and benefits of these developing technologies. Noble (2018:1) coined the term “technological redlining” as a form of digital data discrimination, through a process by which peoples’ digital identities and activities are used to bolster inequality and oppression.

Data use in AI and online tools are often biased due to historical perceptions of racial minorities that influence how developers prioritize the needs of racial majorities or justify racist attitudes through biased historical data (Noble, 2018:10). Noble (2018) and Ruha Benjamin (2020) dismantle the notion that technology is neutral by explaining how data, algorithms, and AI privilege whiteness. Benjamin (2020:18) considers the ways in which software can encode inequality and oppression, by “explicitly amplifying racial hierarchies, by ignoring but thereby replicating social divisions, or by aiming to fix racial bias but ultimately doing quite the opposite.” AI technology is often enacted without our knowledge, through our digital engagements, which become part of algorithmic, automated and “(so-called) artificially intelligent sorting mechanisms that can either target or exclude us – in either case typically not to our benefit” (Bulut, 2018).

AI technologies are increasingly embedded in the software schools use for admissions, advising, courseware, and assessment. These EdTech tools hold tremendous promise to help educators expand access, overcome structural barriers, and close equity gaps. However, more consideration must be given to the risk of structural inequities informing the software’s recommendations, or algorithmic bias. The designers and implementers of these tools seldom consider how racism and inequality can be baked into learning tools if care isn’t taken to make them fairer. The use of algorithms, facial recognition, and surveillance all have the potential to have very negative impacts on underrepresented ethnic students. AI technology can exacerbate racial bias, which further impacts students and communities that are most often underserved and underestimated in schools. This makes it paramount to start with race so that both EdTech companies and schools are equipped to identify and mitigate for racial bias as AI becomes commonplace in schools.

Section 2: Algorithmic Bias in EdTech

Algorithmic bias refers to algorithms that produce results that are systemically prejudiced due to erroneous assumptions in the machine learning process. It generally stems from biases held by people who design, or train AI and machine learning systems. These designers and trainers create or work with algorithms that reflect unintended cognitive biases or real-life prejudices, or biases based on incomplete, faulty or prejudicial data sets used to train and/or validate ML systems. Educational examples include allocative harms in standardized testing that impact high stakes admission decisions (Dorans, 2010; Santelices & Wilson, 2010) and the systematic representation of some groups in a negative light, or in a lack of positive representation (Crawford, 2017). Work by Sweeney (2013) identifies representational harms of denigration and stereotyping, where the word “criminal” was more frequently returned in online ads after searches for black-identifying first names. All these biases and harms impact certain groups negatively, some more than others.

There are other ways that bias can be brought into an AI-driven system such as when the data used is either not large enough or representative enough to teach the system. According to Benjamin (2020:7) algorithms can also “act as narratives” that reaffirm existing inequalities and “operate within powerful systems of meaning that render some things visible, others invisible, and create a vast array of distortions and dangers.” Benjamin (2020:50) notes how the 2016 Beauty AI pageant, based on a machine learning algorithm, strongly preferred contestants with lighter skin, choosing only six non-white winners out of thousands of applicants and leaving its creators confused. This contest that was judged by machines was supposed to use objective

factors such as facial symmetry and wrinkles to identify the most attractive contestants. The robot sorted photos that were labeled or tagged with information on specific facial features, thus the AI was encoded with biases about what is, or what defines beauty. In education, this has potentially harmful implications for standardized testing and facial recognition.

Standardized Testing

In the U.S., many states rely on and use natural language processing (NLP) AI systems, also known as automated scoring engines, to grade standardized tests. NLP systems allow machines to learn about relationships between data and, in some cases, without direct human involvement. These AI systems or agents allow a computer to understand human inputs using algorithms that comb through billions of words of training data. However, automated grading agents also suffer from built-in biases based on the way they are taught to look for mistakes and errors. Data published by the Educational Testing Service (Winerip, 2012) highlights the results of its E-rater grading engine which found that the machine under-scored African American students and showed bias against Arabic, Spanish, and Hindi speakers. This bias can do a significant amount of damage to a student's grade—which can be essential to their opportunity to pass a class or advance to the next grade level.

Unlike human graders who can interpret information in front of them, particularly when given the subjective task of grading an essay, an algorithm only knows what to look for what it is trained to grade on. The AI scoring agent is trained by feeding the machine learning algorithm sets of data or models that have already been scored by humans. The machine processes the results, systematizing the process of figuring out what student output is passing or failing. Then, it takes the training set and applies it to new assignments, making a prediction about how the assignments should be graded using what it has learned. The scoring engine measures the metrics, but the rest is often lost in the translation: What about environmental factors that can negatively impact test taking students? This issue is briefly addressed in “#tresdancing” when a student fails because an online test requires students to be in a private solitary space.

Facial Recognition and Stereotyping

Image or facial recognition systems that use biased machine learning data sets contain inherent racial and gender biases. A recent study conducted by the U.S. National Institute of Standards and Technology or NIST confirmed that even the best facial recognition software has algorithmic bias (Grother et al., 2019). NIST quantified the accuracy of face recognition algorithms for demographic groups defined

by sex, age, and race or country of birth. They used algorithms with large datasets of photographs collected by the U.S. government, including mugshots and photos taken for immigration benefits, visas and border crossings. One such algorithm used by the AEGIS system misidentified African American men twice as often and African American women ten times as often as their white counterparts (Fergus, 2020). Some school districts have started using facial recognition software that has problems identifying the faces of African American students.

Facial recognition technologies allow for the extraction of a wide range of features from images. Researchers found that face analyzing A.I. systems work significantly better for white faces than black ones (Buolamwini & Gebru, 2018; Buell, 2018). Joy Buolamwini, a Ghanaian Canadian computer scientist and digital activist, discovered that wearing a white mask worked better than using her actual face (Lee, 2020). Buolamwini and Gebru (2018) looked at how well corporate face-scanning systems did at figuring out whether a person in a picture was a man or a woman. The study found that if the person in a photograph was white and male, then the systems guessed correctly more than 99% of the time. On the other hand, the systems failed to identify Black women 50% of the time. The reason for this is that most face-analysis programs are trained and tested using databases of hundreds of pictures, which research has found are overwhelmingly white and male (Buolamwini & Gebru, 2018).

“#tresdancing” highlights the experiences of a Black Canadian girl who is made to use AI-driven software that tracks her eye movements, even as research indicates that the algorithms embedded in this type of system has a problem with registering darker facial features, which can lead to more mistakes (i.e., mistaken identity, racial profiling). According to Perry and Turner-Lee (2020) students like Frankie will face greater inequalities if educators go too far toward digitizing education without considering how to check the inherent biases of the (mostly white) developers who create AI systems. Joy Buolamwini and Algorithmic Justice League (2020) call for developers to build better “face databases for development and testing that display more diversity across parameters such as race, gender and age.”

Section 3: Speculative Design Thinking and Liberatory Design

What is lacking in the mainstream technology sector can be found elsewhere, where practitioners, artists, and activists engage their communities in brainstorming processes to identify and solve tough challenges. They use *design thinking* to prototype ideas and learn from mistakes. Design thinking may offer some direction to developers who are

tasked to address the needs of the most vulnerable, underrepresented student populations. To avoid embedding inequity in future EdTech products, developers must intentionally avoid status quo design or designing for the “average student”, who is often portrayed as white, male, and middle- to upper-income. Design thinking for a more equitable future must include a broader range of perspectives and lived experiences. Future-facing design ideas created by people underrepresented in the EdTech field can provide guideposts for private sector development.

Researchers, artists and practitioners use speculative design to apply what is referred to as.

foresight to interrogate the development and use of AI technology. According to Buehring and Liedtka (2018:140) having foresight when designing projects can help foster inclusion and equity, as well as provide opportunities to learn through design thinking, especially prototyping and experimentation. Employing this strategy in the development of emerging technologies invites new perspectives such as the examples included in this paper that use current knowledge of AI technology to explore alternative futures that might happen, is likely to happen, and in ways that more diverse groups want to happen. Also, it gives participants from underrepresented groups a different role in the creation of AI technology solutions.

Persistent inequities in the technology field extended to mainstream science fiction where, according to writer and cultural critic Dery (1997:188), the lack of diversity was a sign for underrepresented groups to keep out. Alternative practices were established to encourage members of these groups to challenge narrow assumptions, preconceptions and givens about the role technology plays in their lives. Dery interviewed sci-fi writer Samuel Delany who made the distinction between the *white boxes of computer technology* and the *black boxes of modern street technology* (Dery, 1997:192). Whereas the white boxes are used to develop software for computer-based/mediated instruction, the black boxes such as mobile phones are popular with people from underrepresented groups. Speculative design re-casts artists, writers and cultural critics from these groups as practitioners whose works propose changes to existing systems, encouraging readers/viewers to imagine and explore a range of ideas that provoke thought about their past, present, and future engagements with technology.

Afrofuturism and Liberatory Design

Dery (1997:185) coined the term *Afrofuturism* in 1994 to describe “speculative fiction that treats African American themes and addresses African American concerns in the context of twentieth century technoculture.” This domain has grown to include other areas such as *speculative design*, which is a method used to help people address societal

problems and look towards the future—and create projects for those scenarios (Dunne & Raby, 2013). One of the earliest examples of this for African Americans is a short treatise written by Amiri Baraka circa 1971 that calls for creators, designers, and developers to learn that “western technology must not be the end of our understanding” of the technology fields. The next wave, commonly referred to as Afrofuturism 2.0, has applied a “liberatory design lens” to enable designers and software developers to engage a Black ethos in seeing – as articulated by Baraka – “everything fresh and ‘without form’ to “then make forms that will express us <Blacks/African-Americans> truthfully and totally” (Winchester, 2019; Baraka, 1971).

Through Afrofuturism and speculative design, Black and indigenous storytellers, artists, and makers address algorithmic bias in AI by applying a liberatory design lens to technology development such as Jeremy Vincent’s EdTech venture AfroBrainiac (2022). The Iyapo Repository facilitates community-driven, speculative design thinking workshops, to collect materials such as field notes to be made into physical artifacts (Gaskins, 2021:25; Okunseinde, 2020:92; Sinclair, 2017). Ayodamola Okunseinde’s *Incantation* critiques AI algorithms as “languages of exclusion” and asks us to re-assess the development and predictive power of these systems (Fortunato, 2020). Stephanie Dinkins recruited coders and creative technologists from underrepresented groups to create a storytelling AI robot. The AI’s storytelling algorithm is driven by a training dataset that repurposes text from Toni Morrison’s *Sula* and the artist’s family interviews (Gaskins, 2021:28).

EdTech developers can learn from projects such as AfroBrainiac, *Incantation*, and Dinkins’ storytelling AI, which are at the intersection of Afrofuturism and speculative design. These works often examine the darker side of technology, wherein developers can anticipate and find potential issues that could be used to find better solutions. Anatola Araba Pabst (2021) created an animated short film titled “Afro Algorithms” that imagines a distant future where AI technology and machine learning are a part of everyday life. The film’s main protagonist, Aero, is an AI-driven robot who becomes a world leader. At some point, Aero realizes that important voices and worldviews are missing from her databank, including the experiences of the historically marginalized and oppressed. Aero resolved to fix this problem by seeking out and finding the missing data to add to her dataset. The intention of this film is to spark “conversations about race, technology, and where humanity is driving the future of this planet” (Pabst, 2021).

Stephanie Dinkins (see Civin, 2022) believes that AI has the potential to become a “democratic survey tool, taking in many ideas, analyzing them, and sketching exciting frameworks for action”, which aligns with EdTech researchers’ calls for addressing the threat algorithmic bias poses to AI in

the EdTech industry. Referencing Afrofuturism and speculative design at varying stages of the EdTech development process can detect insights that address risks and harms in AI and lay out the constraints of current technological capabilities and resources, and what would need to change in order to create a more equitable future (Buehring & Liedtka, 2018:141). Educational opportunities in underrepresented and underserved communities can help to establish a future-focused process that benefits these groups to get at what might happen, what is likely to happen, and what they want to see happen. Combining these elements and practices in an iterative, culturally relevant way creates a space for moving from speculative design thinking towards a liberatory practice and mindset in EdTech AI-driven software development.

Conclusion: Challenges and Recommendations

EdTech developers use data to train algorithms that promise to personalize learning, identify at-risk students, and save teachers' time. However, without examining the biases that influence this data, companies using AI can amplify existing racial biases, along with their own assumptions, into the products created for students to use. Programmers, who are mostly white and male, lack the necessary knowledge and flexibility to address persisting digital inequalities marked by disparities across socio-demographic groups in terms of access, skills, and different uses of the technology perhaps in ways that were not intended. Layers of bias and racism can move from one system to the next, especially when diverse sources of information and data are not considered at the development phase. The use of biased algorithms in facial recognition, and surveillance have the potential to have very negative impacts on historically marginalized, underrepresented ethnic students that rely more and more on AI technology at home and school. This paper looked at a variety of projects (e.g., films, toolkits, applications) that hold tremendous promise to help educators, developers, and researchers address bias and racial inequity.

Algorithms can reaffirm existing inequalities such as in standardized testing that impact high stakes admission decisions, the systematic misrepresentation of the historically marginalized and the representational harms of denigration and stereotyping. AI in standardized testing has been shown to under-score black students and show bias against other groups. Automated AI scoring agents are not designed to consider environmental factors out of students' control that can negatively impact their test taking performances. Facial recognition software, used in some EdTech software, struggles with registering women and darker faces, which leads to mistaken identities and racial profiling. To counter these issues, researchers are calling for developers to be more

intentional about addressing bias when creating AI-powered technology.

EdTech solutions need to balance empowering student users as critical digital agents of change while ensuring that the solutions provide a safe space for encouraging them to learn about AI and algorithmic bias. EdTech developers can learn from Afrofuturism and speculative design projects that examine the darker side of AI, providing fictional and non-fictional scenarios that address the current needs of the historically marginalized and underserved. EdTech company AfroBrainiac uses Afrofuturism and 'digital play' to engage youth in AI. Films such as "#tresdancing" and "Afro Algorithms" can provide EdTech developers with design archetypes of users that can help them better understand the educational and personal contexts of underrepresented groups. These works (tools, films, art) channel alternative frameworks that addresses equity through the generation of new ideas and prototypes that counter bias and other negative effects of AI in underrepresented and under resourced groups.

New avenues of research and creative work are urgently needed because oppressive algorithms are increasingly being used in systems that more and more students are required to use. These algorithms have the potential to amplify existing inequalities. Some work has already been done such as The AI in Education Toolkit for Racial Equity, the EdSAFE AI Alliance, and Digital Promise that assesses the utility and fairness of AI in EdTech. The Algorithmic Justice League combines art and research to illuminate the social implications and harms of AI. There is much at stake: AI systems can be trained to promote or discriminate, approve or reject, or render visible or invisible. Thus, they must be interrogated, and we must have wider public discussions about their consequences. By learning from past missteps, we can instead move towards a future where more people feel liberated to reimagine constructions of race (class, gender, etc.), but also how this could lead to more liberatory constructions of AI in the future.

Declarations

Conflicts of Interests/Competing Interests The author declares that they have no conflicts of interest.

References

- Ali, S., Williams, R., Payne B., Park H., Breazeal C. (2019). Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education. Presented in the *International Workshop on Education in Artificial Intelligence K-12* (EDUAI '19). Palo Alto, CA, USA. Proceedings of IJCAI 2019.

- Angwin, J., J. Larson, S. Mattu, & L. Kirchner. (2016). Machine bias. Retrieved September 3, 2022, from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Atske, S., & Perrin, A. (2022). Home broadband adoption, computer ownership vary by race, ethnicity in the U.S. Retrieved April 5, 2022, from <https://www.pewresearch.org/fact-tank/2021/07/16/home-broadband-adoption-computer-ownership-vary-by-race-ethnicity-in-the-u-s>
- Attewell, P., Kasinitz, P., & Dunn, K. (2010). Black Canadians and black Americans: Racial income inequality in comparative perspective. *Ethnic and Racial Studies*, 33(3), 473–495. <https://doi.org/10.1080/01419870903085883>
- Atwood, M. (2014). *In other worlds: SF and the human imagination*. Virago.
- Baraka, I. A. (1971). Technology & Ethos. Raise, race, rays, raze: Essays since 1965. *Amistad*, 2, 155–158.
- Baym, N. K. (2015). Social media and the struggle for society. *Social Media + Society*, 1(1), 205630511558047. <https://doi.org/10.1177/2056305115580477>
- Benjamin, R. (2020). *Race after technology: Abolitionist tools for the new jim code*. Polity.
- Bradley, B. (2021). Tool aims to track disparities in Ed-tech usage by race, poverty level. Retrieved April 5, 2022, from <https://marketbrief.edweek.org/marketplace-k-12/tool-aims-track-disparities-ed-tech-usage-race-poverty-level>
- Brigham, S. E. (1993). TQM: Lessons we can learn from industry. *Change*, 42–48.
- Broussard, M. (2019). *Artificial unintelligence: How computers misunderstand the world*. The MIT Press.
- Buehring, J. H., & Liedtka, J. (2018). Embracing systematic futures thinking at the intersection of strategic planning, foresight and design. *Journal of Innovation Management*, 6(3), 134. https://doi.org/10.24840/2183-0606_006.003_0006
- Bulut, E. (2018). Interview with Safiya Umoja Noble: Algorithms of oppression, gender and race. *Journal of Cultural Studies, Faculty of Communication*, Hacettepe University, 5(2): 294–301.
- Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in Proceedings of Machine Learning Research* (Vol. 81, pp. 77–91).
- Cheney, G., McMillan, J. J., & Schwartzman, R. (1997). Should we buy the ‘student-as-consumer’ metaphor? *The Montana Professor*, 7(3), 8–11.
- Civin, M. (2022). Stephanie dinkins: On love and data. Retrieved September 3, 2022, from <https://brooklynrail.org/2022/04/artseen/Stephanie-Dinkins-On-Love-and-Data>
- Crawford, K. (2017). The trouble with bias - NIPS 2017 keynote . YouTube. Retrieved September 14, 2022, from https://youtu.be/fMym_BKWQzk
- Dery, M. (1997). *Flame wars: The discourse of cyberculture*. Duke Univ. Press.
- Dimaggio, P., Harittai, E., Celeste, C., & Shafer, S. (2020). Excerpts from “digital inequality: From unequal access to differentiated use”. *Inequality in the United States: A Reader*, 98–113. <https://doi.org/10.4324/9781003071709-13>.
- Dobson, J. E. (2019). *Critical digital humanities: The search for a methodology*. University of Illinois Press.
- Dorans, N. (2010). Misrepresentations in unfair treatment by Santelices and Wilson. *Harvard Educational Review*, 80(3), 404–413. <https://doi.org/10.17763/haer.80.3.1253473353686748>
- Dunne, A., & Raby, F. (2013). *Speculative everything: Design, fiction, and social dreaming*. The MIT Press.
- Dunne, A., & Raby, F. (2021). *Design noir: The secret life of electronic objects*. Bloomsbury Visual Arts.
- EdSAFE AI (2022). Retrieved September 3, 2022, from <https://www.edsafeai.org>
- Fergus, J. (2020). A facial recognition company lied to a school about how racist its algorithm is. Retrieved March 29, 2022, from <https://www.inputmag.com/tech/a-facial-recognition-company-lied-to-a-school-about-how-racist-its-algorithm-is>
- Fortunato, S. (2020). Straight talk: Ayodamola Okunseinde. Retrieved April 1, 2022, from <https://www.sciartmagazine.com/straight-talk-ayodamola-okunseinde.html>
- Fouché, R. (2004). Not made for black history month: Lewis Latimer and technological assimilation. *Appropriating technology: Vernacular science and social power* (pp. 315–330). University of Minnesota Press.
- Fuchs, C., & Horak, E. (2008). Africa and the digital divide. *Telematics and Informatics*, 25(2), 99–116.
- Gaskins, N. R. (2021). *Techno-vernacular creativity and innovation: Culturally relevant making inside and outside of the classroom*. The MIT Press.
- Gee, J. P. (2017). *Teaching, learning, literacy in our high-risk high-tech world: A framework for becoming human*. Teachers College Press.
- Grother, P, Ngan, M., & Hanaoka, K. (2019). Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects. *National Institute of Standards and Technology*.
- Harrison, L. M., & Risler, L. (2015). The role consumerism plays in student learning. *Act Learn High Educ*, 16(1), 67–76. <https://doi.org/10.1177/1469787415573356>
- Hebbar, N., & M. Jacobs. (2022) What do Edtech and AI have to do with racial bias? Retrieved September 3, 2022, from <https://digitalpromise.org/2021/10/28/what-do-edtech-and-ai-have-to-do-with-race-bias/>
- Hebbar, N., Jacobs, M., Volk, M., & Allendorf, E. (2020). AI in Education Toolkit for Racial Equity. *The EdTech Equity Project*.
- Itō, M., Martin, C., Pfister, R. C., Rafalow, M. H., Salen, K., & Wortman, A. (2019). *Affinity online: How connection and shared interest fuel learning*. New York University Press.
- Karumbaiah, S., & J. Brooks. (2021). How Colonial Continuities Underlie Algorithmic Injustices in Education. *2021 Conference on Research in Equitable and Sustained Participation in Engineering, Computing, and Technology (RESPECT)*. <https://doi.org/10.1109/respect51740.2021.9620605>.
- Koenig, R. (2019). The new Jim code? race and discriminatory design - edsurge news. Retrieved April 5, 2022, from <https://www.edsurge.com/news/2019-08-20-the-new-jim-code-race-and-discriminatory-design>
- Lee, J. (2020). When bias is coded into our technology. NPR. Retrieved September 14, 2022, from <https://www.npr.org/sections/codeswitch/2020/02/08/770174171/when-bias-is-coded-into-our-technology>
- Lenhart, A. (2019). Mobile access shifts social media use and other online activities. Retrieved April 5, 2022, from <https://www.pewresearch.org/internet/2015/04/09/mobile-access-shifts-social-media-use-and-other-online-activities>
- Love, B. (2016). Get to the future. Retrieved September 3, 2022, from <http://getfreehiphopicivics.com/get-to-the-future/>
- Lynch, M. (2018). Analyzing Edtech's diversity problem. The Tech Advocate. Retrieved September 14, 2022, from <https://www.thetechadvocate.org/analyzing-EdTechs-diversity-problem>
- Mawasi, A., Ahmed, I., Walker, E., Wang, S., Marasli, Z., Whitehurst, A., Wylie, R. (2020). Using Design-Based Research to Improve Help Giving in Middle School Math Classroom. *International Conference of the Learning Sciences*.
- Naidoo, R., & Whitty, G. (2013). Students as consumers: Commodifying or democratising learning? *International Journal of Chinese Education*, 2(2), 212–240. <https://doi.org/10.1163/22125868-12340022>

- Nguyen, M. H., Hargittai, E., & Marler, W. (2021). Digital inequality in human communication during a time of physical distancing: The case of covid-19. *Computers in Human Behavior*, *120*, 106717. <https://doi.org/10.1016/j.chb.2021.106717>
- Nielsen, M. (2011). *Reinventing discovery: The new era of networked science*. Princeton University Press.
- Njee, N. (2016). Share cropping blackness: White supremacy and the hyper-consumption of black popular culture. *Sociology*.
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
- Okunseinde, A. T. (2020). Iyapo repository: Constructing and archiving alternate futures. In V. Bradbury & S. O'Hara (Authors), *Art hack practice: Critical intersections of art, innovation and the maker movement* (pp. 87–98). : Routledge.
- Pabst, A. A. (2021). Afro algorithms: Imagining new possibilities for race, technology, and the future through animated storytelling. *Patterns*, *2*(8). <https://doi.org/10.1016/j.patter.2021.100327>.
- Perry, A., & Turner-Lee, N. (2020). AI can disrupt racial inequity in schools, or make it much worse. Retrieved April 1, 2022, from <https://hechingerreport.org/ai-can-disrupt-racial-inequity-in-schools-or-make-it-much-worse/>
- Pinkard, N., Erete, S., Martin, C. K., & McKinney de Royston, M. (2017). Digital youth divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities. *Journal of the Learning Sciences*, *26*(3), 477–516. <https://doi.org/10.1080/10508406.2017.1307199>
- Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., & Denton, E. (2020). Saving face. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3375627.3375820>
- Reich, J., & Itō, M. (2017). From good intentions to real outcomes: Equity by design in learning technologies. *Digital Media and Learning Research Hub*.
- Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences*, *116*(17), 8255–8260. <https://doi.org/10.1073/pnas.1808307116>
- Robinson, L., Schulz, J., Khilnani, A., Ono, H., Cotten, S. R., McClain, N., ... Tolentino, N. (2020). Digital inequalities in time of pandemic: Covid-19 exposure risk profiles and new forms of vulnerability. *First Monday*. <https://doi.org/10.5210/fm.v25i7.10845>
- Santelices, M. V., & Wilson, M. (2010). Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning. *Harvard Educational Review*, *80*(1), 106–134. <https://doi.org/10.17763/haer.80.1.j94675w001329270>
- Sinclair, K. (2017). We have the power to code a new operating system for humanity. Retrieved September 3, 2022, from <https://www.weforum.org/agenda/2017/06/artificial-intelligence-code-better-world/>
- Srinivasan, R. (2018). *Whose global village? Rethinking how technology shapes our world*. NYU Press.
- Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, *56*(5), 44–54. <https://doi.org/10.1145/2447976.2447990>
- Toyama, K. (2015). *Geek heresy: Rescuing social change from the cult of technology*. PublicAffairs.
- Vincent, J. (2022). Afrobrainiac. Retrieved September 3, 2022, from <https://www.afrobrainiac.com/>.
- Watkins, S. C., & Cho, A. (2018). *The digital edge: How black and Latino youth navigate digital inequality*. NYU Press.
- Winchester III, W. W. (2019). Engaging the black ethos: Afrofuturism as a design lens for inclusive technological innovation. *Journal of Futures Studies*, *24*(2), 55–62.
- Winerip, M. (2012). Facing a robo-grader? just keep obfuscating mellifluously. Retrieved March 29, 2022, from <https://www.nytimes.com/2012/04/23/education/robo-readers-used-to-grade-test-essays.html>
- Young, J. K. (2006). *Black writers, white publishers: Marketplace politics in twentieth-century African American literature*. University of Mississippi.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.