To: U.S. Department of Education

From: Chloe Jefferson

Date: Wednesday, November 20, 2024

RE: Promoting School-Level Accountability for the Responsible Deployment of AI and Related Systems in K-12 Education: Mitigating Bias and Increasing Transparency

---

Promoting School-Level Accountability for the Responsible Deployment of AI and Related Systems in K-12 Education: Mitigating Bias and Increasing Transparency[1]

Chloe Jefferson

With
Howard University & Apart Research

Abstract (Executive Summary)

This policy memorandum draws attention to the potential for bias and opaqueness in intelligent systems utilized in K–12 education, which can worsen inequality. The U.S. Department of Education is advised to put Title I and Title IV financing criteria into effect that require human oversight, AI training for teachers and students, and open communication with stakeholders. These steps are intended to encourage the responsible and transparent use of intelligent systems in education by enforcing accountability and taking reasonable action to prevent harm to students while research is conducted to identify industry best practices.

Keywords: bias, K-12 education, AI, intelligent systems, transparency, Title I, Title IV, human oversight

[1] Research conducted at the Howard University AI Safety Summit & Policy Hackathon (https://www.apartresearch.com/event/howard-university-ai-policy-hackathon), 2024

1. Introduction (Problem Analysis & Context)

Artificial Intelligence (AI) represents any system that imitates human intelligence. It has broad applications in industries ranging from healthcare, transportation, entertainment, education, and more. AI is a transformative technology that is developing rapidly–so rapidly that it is difficult for regulators to keep pace with the risks it introduces in society. AI technology uses algorithms, which are procedures used to accomplish a task following a series of steps. Both AI and complex algorithms can create a "black box" or a system in which the reasoning behind the decisions is unclear or unknowable. The combination of AI and complex algorithms will be described as intelligent systems from here on.

Intelligent systems will reflect the biases in the data they are trained on, leading to biased outcomes. A Stanford Center for Racial Justice report supports that algorithms can worsen educational disparities when the training data reflects existing biases (Pham et al., 2024). Studies show that biased algorithms can disproportionately affect marginalized groups, exacerbating disparities in educational outcomes (Pham et al., 2024). Since these technologies increasingly affect important decisions like student assessments, bias in intelligent systems employed in education is a developing worry that could exacerbate already-existing imbalances. Additionally, a lack of transparency regarding when and how these systems are being deployed and their potential limitations takes the voice away from students, parents, and teachers.

Currently, there appear to be no comprehensive federal requirements or state laws addressing school accountability for the impact of bias in intelligent systems. When the risk of bias in intelligent systems and a lack of transparency of their usage is left unaddressed, schools can continue to implement intelligent systems that may perpetuate historical bias without accountability for their role. This policy memorandum will discuss suggestions to increase accountability for the responsible deployment of intelligent systems in K-12 education focused on the short to medium term (approx. 5 years).

2. Methods (Evidence & Analysis)

2.1 Data Sources & Methodology

This analysis draws from a combination of literature review, qualitative assessment, and case studies to evaluate the deployment of AI in K-12 education. The most pressing risks were identified and prioritized, with actionable recommendations created to address these issues, focusing on the short to medium term (5 years).

There are some notable limitations to the investigation that should be highlighted. Since AI has gained traction in the public ethos in recent years (ChatGPT's revolutionary public release was only two years ago), there are few peer-reviewed articles available on the subject. Research relating to bias in education is especially limited, with many relevant articles focused on complex algorithms used in education rather than the most recently available general AI systems. This led to reliance on the inclusion of informal analysis in news articles, industry reports, and pre-published research to stay current with the pace of tech development in this area.

2.2 Key Findings

Evidence of Bias

- **Racial Bias in Dropout Early Warning System (DEWS):** A 2021 equity analysis in Wisconsin found that DEWS disproportionately flagged Black and Hispanic students as likely to drop out, with Black students having a 42% higher false alarm rate than White students (Feathers, 2023). Using risk-scoring systems can have a negative impact on teachers' perceptions and students' self-esteem, which can influence their academic potential, even though these students graduated at higher rates than anticipated (Pham et al., 2024). As additional states adopt such systems, the risk of pervasive prejudice increases in the absence of sufficient safeguards
- **ChatGPT Socioeconomic Bias in Essay Scoring**: Research showed that identical essays received significantly lower scores when the prompt revealed the "student" was from a low-achieving public school and significantly higher scores when the prompt described

the student as being from an elite private school (Warr et al., 2023). This suggests that ChatGPT's responses may be influenced by implicit biases in training data, highlighting the need for further investigation into these biases in frontier large language models (LLMs).

Evidence of Lack of Transparency

A school district using Wisconsin's Dropout Early Warning System (DEWS) sent letters to parents about their students' risk level, but there was no mention of the algorithm used to make this determination (Feathers, 2023). However, Wisconsin is not the only state with issues of transparency. A more recent parent survey by Learner revealed that "50% of parents surveyed are unaware of their school's policy regarding the use and implementation of AI tools and technologies" (Neves, 2024). It is more difficult for parents to challenge the accuracy of intelligent systems or safeguard their children's interests when they are unaware of how they are being used due to this information access gap.

2.3 Analysis of Policy Approaches

Currently, approaches to reduce the risk of bias from intelligent systems used in K-12 education are disjointed. Many states and school districts have no formal laws or policies in this area. Four approaches to the problem using influence at the federal level will be discussed below.

Option 1: Offering a Guidance-Only Approach to Support Schools/Districts/States - While providing guidance is a necessary starting point, it lacks enforceability and accountability for the safe deployment of intelligent systems. Without this, the approach falls short of meeting the critical needs of the moment–protecting students from biased systems and ensuring transparency with all relevant stakeholders.

Option 2: Focus on Influencing the Developers of Intelligent Systems - While efforts aimed at accountability for the developers of intelligent systems are critical, this approach does nothing to address the gap in accountability at the level of deployment in the near term. This leaves the

question, “How will schools be held responsible for the impact of their implementation of intelligent systems?” unanswered.

Option 3: Mandate School-Level Bias Identification and Mitigation Systems to Receive Federal Funding - This approach puts all the onus of the technical solution to bias mitigation at the local level. This approach is currently infeasible because the technology is rapidly developing and accountability for solving the problem at this level requires a high degree of technical expertise. There are simply not enough subject matter experts available to offer solutions at the individual school or district level across the nation. This field has yet to identify research-backed, standardized, industry best practices for bias mitigation in intelligent systems, and more time is needed for clear solutions to arise. In a U.S. Department of Education Listening Session to gain feedback from stakeholders on what leaders in education should do, investing in research and development was the most highly valued method to expand trust under the category of “guardrails and guidelines” (U.S. Department of Education, Office of Educational Technology, 2023, p. 58)

Option 4 (Recommended): Apply Moderate Requirements for Accountability at the School Level to Receive Federal Funding - Placing clear and reasonable restrictions on the use of intelligent systems that factor in current research and stakeholder needs creates a stopgap for bias mitigation without requiring a full-fledge technical solution. This balanced approach ensures schools are held accountable for their decisions and their impact on students. These policy recommendations will be discussed in the next section.

3. Results (Recommended Solution)

3.1 Policy Recommendation

The U.S. Department of Education should implement conditional funding criteria linked to schools' appropriate use of intelligent systems in order to reduce bias and guarantee transparency in the implementation of these systems in K–12 education. In particular, these rules should be

incorporated into Title I and Title IV financing regulations to make sure that schools are held responsible for the equitable and open use of intelligent systems.

Rationale for the Selection of Title I and Title IV

Schools with a significant proportion of low-income children can receive federal funds through Title I to enhance learning opportunities and reduce achievement inequalities. Since Title I funding is already designed to address inequalities, it is an ideal candidate for the additional requirements to mitigate algorithmic bias.

Programs including school safety, career development, and a holistic education that improves students' well-being are supported under Title IV. As AI continues to shape daily life and careers, integrating it into these programs is essential to prepare students for the modern workforce.

Both funding sources are essential for schools; Title I is one of the largest federal funding sources for public schools, impacting millions of students across the country.

Recommended Updates to Title I and Title IV Funding Requirements

- **Title I Requirements:**
    - Schools must not rely *exclusively* on intelligent systems for critical areas such as student assessment and decision-making processes that impact students' academic paths or future opportunities. Human oversight is critical as these systems have been shown to display signs of bias.
    - Training programs must be established for both educators and students on how to use and understand intelligent systems in the classroom. This training must include their limitations and the impact of bias.
    - Annually and whenever new intelligent systems are introduced, schools must provide clear communication regarding the use of these systems in a publicly available policy.
- **Title IV Requirements:**

    - This includes all Title I requirements, along with an additional requirement for educator professional development to prepare students for careers in the age of AI.

3.2 Implementation Roadmap

- **3-6 months:**
    - Add the guidelines to the U.S. Department of Education website.
    - Draft the changes to the Title I and Title IV funding requirements, outlining the conditions necessary to secure federal funding.
- **6-12 months:**
    - Finalize the updates to the funding requirements.
    - Initiate the public notice period to inform the public of the proposed funding rule changes and gather feedback.
    - Launch a comment period to allow schools, stakeholders, and the public to provide input on the proposed regulations.
- **12-18 months:**
    - Issue the final rule changes for Title I and Title IV funding, making the updated regulations official and enforceable.

4. Discussion (Impact Assessment)

4.1 Impact Assessment

The proposed policy changes are expected to bring both short-term and long-term benefits:

- **Short-term (1-2 years):**
    - Increased public awareness of bias in intelligent systems
    - Effort to improve stakeholder transparency to align with new Title I and Title IV funding requirements
- **Medium-term (3-5 years):**

    - A more informed school community, with educators trained to implement AI responsibly
    - Students understand AI's impact on their future as a student and in their career
    - Schools are held accountable through the possibility of losing federal funding for non-compliance
- **Long-term (5+ years):**
    - Stopgap measures allow research to catch up
    - Industry best practices are established for bias mitigation in AI systems
    - Policies evolve to address new developments.

4.2 Feasibility

Technical and Economic

The proposed requirements have no extensive financial impact on schools and schools are not restricted in their choice of intelligent systems. Additionally, AI training resources and guidance are available for free from respected organizations and institutions, such as the U.S. Department of Education. Any school that purchases an intelligent system should have access to specialized resources and customer support for their systems to ensure proper training of staff. Conditional funding will motivate schools to take action.

Political

Given the current political landscape, introducing requirements relating to bias may be interpreted as federally mandated diversity, equity, and inclusion education. Conservative political groups may attack these measures if the requirements are not carefully worded to be as politically neutral as possible. It must be clear that the individual schools have complete control over the implementation of the requirements. The guidance will provide options to support schools in meeting the requirements, but no specific framework will be explicitly mandated. This must be kept in mind during the drafting phase, and all viewpoints must be taken into consideration during the public notice period.

5. Conclusion

Research has demonstrated that intelligent systems can be biased, exacerbating inequalities. Currently, more research needs to be conducted to create formal bias identification and mitigation systems. In the interim, measures to increase accountability at the school level can be tied to federal funding (Title I and Title V). The proposed funding requirements include a mandate for human oversight in assessment and decision-making, new training for educators and students, and a public policy on the use of intelligent systems. The U.S. Department of Education must take actionable steps now to implement these solutions. Without them, schools will have no accountability for their part in the perpetuation of bias through intelligent systems, and unintended consequences may cause lasting harm for students in schools today whose voices may not be heard.

6. References

Feathers, T. (2023, April 27). False alarm: How Wisconsin uses race and income to label students “high risk”. The Markup. https://themarkup.org/machine-learning/2023/04/27/false-alarm-how-wisconsin-uses-race-and-income-to-label-students-high-risk

Neves, M. (2024, February 6). *“Hey Siri, will AI change my kid’s future?” What parents really think*. Learner. https://www.learner.com/blog/ai-education-survey

Pham, H., Kohli, T., Llano, E. O., Nokuri, I., & Weinstock, A. (2024, June 29). *How will AI impact racial disparities in education?* Stanford Center for Racial Justice. https://law.stanford.edu/2024/06/29/how-will-ai-impact-racial-disparities-in-education/

U.S. Department of Education, Office of Educational Technology. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations.* https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf

Warr, M., Oster, N. J., & Isaac, R. (2023). *Implicit bias in large language models: Experimental proof and implications for education.* SSRN. https://doi.org/10.2139/ssrn.4625078