Battling algorithmic bias in education

OECD Digital Education Outlook 2023

Stéphan Vincent-Lancrin, PhD
Deputy Head of IMEP division and Senior Analyst
Directorate for Education & Skills

XX/YY/ZZ
• What are the current frontiers of AI and other technologies in education?

• What are the upcoming challenges?

• Watch key experts and policy makers talk about it:

https://oecd-events.org/digital-education
Teacher feedback for self-regulation

Showing teachers where they spend time in the classroom
Preventing dropout through early warning systems

### Advisory Dashboard

#### Advisory Dashboard - Teacher's View

<table>
<thead>
<tr>
<th>Student Name</th>
<th># of F's</th>
<th>Discipline</th>
<th>Attendance</th>
<th>Enrichment</th>
<th>Community Service Hours</th>
<th>GPA Simple Current</th>
<th>GPA Simple Cumulative</th>
<th>Suspension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akins, Tamasha</td>
<td>2</td>
<td>40</td>
<td>91.92%</td>
<td>0</td>
<td>3</td>
<td>1.84</td>
<td>1.96</td>
<td>4</td>
</tr>
<tr>
<td>Jibe, Montrell</td>
<td>7</td>
<td>24</td>
<td>97.98%</td>
<td>0</td>
<td>0</td>
<td>0.41</td>
<td>0.76</td>
<td>0</td>
</tr>
<tr>
<td>Anderson, Asia</td>
<td>0</td>
<td>18</td>
<td>92.93%</td>
<td>0</td>
<td>44</td>
<td>1.70</td>
<td>2.28</td>
<td>3</td>
</tr>
<tr>
<td>Andrews, Kiara</td>
<td>0</td>
<td>3</td>
<td>91.92%</td>
<td>0</td>
<td>5</td>
<td>2.88</td>
<td>3.28</td>
<td>3</td>
</tr>
<tr>
<td>Angueta, Meyerhoftzin E</td>
<td>0</td>
<td>1</td>
<td>94.95%</td>
<td>0</td>
<td>48</td>
<td>3.56</td>
<td>3.96</td>
<td>0</td>
</tr>
<tr>
<td>Amistad, Adrienne</td>
<td>1</td>
<td>29</td>
<td>73.74%</td>
<td>0</td>
<td>2</td>
<td>1.60</td>
<td>2.26</td>
<td>9</td>
</tr>
<tr>
<td>Amistad, Sean A</td>
<td>0</td>
<td>65</td>
<td>92.73%</td>
<td>0</td>
<td>0</td>
<td>0.90</td>
<td>0.93</td>
<td>6</td>
</tr>
<tr>
<td>Bannes, Mario</td>
<td>0</td>
<td>53</td>
<td>74.75%</td>
<td>0</td>
<td>3</td>
<td>2.02</td>
<td>1.55</td>
<td>1</td>
</tr>
<tr>
<td>Banks, Devon</td>
<td>0</td>
<td>4</td>
<td>97.98%</td>
<td>0</td>
<td>10</td>
<td>3.25</td>
<td>2.26</td>
<td>0</td>
</tr>
<tr>
<td>Banks, Mlack</td>
<td>1</td>
<td>28</td>
<td>92.93%</td>
<td>0</td>
<td>10</td>
<td>8.78</td>
<td>2.99</td>
<td>0</td>
</tr>
<tr>
<td>Bar, Dejah</td>
<td>3</td>
<td>1</td>
<td>96.67%</td>
<td>0</td>
<td>7</td>
<td>0.20</td>
<td>2.34</td>
<td>8</td>
</tr>
<tr>
<td>Beck, Tekyath</td>
<td>0</td>
<td>3</td>
<td>76.70%</td>
<td>0</td>
<td>20</td>
<td>3.96</td>
<td>1.56</td>
<td>0</td>
</tr>
<tr>
<td>Bell, Maurice</td>
<td>2</td>
<td>1</td>
<td>91.92%</td>
<td>0</td>
<td>0</td>
<td>1.68</td>
<td>1.64</td>
<td>0</td>
</tr>
<tr>
<td>Bon, Tashema</td>
<td>0</td>
<td>0</td>
<td>95.96%</td>
<td>0</td>
<td>0</td>
<td>3.62</td>
<td>3.41</td>
<td>0</td>
</tr>
<tr>
<td>Booker, Isaac</td>
<td>0</td>
<td>18</td>
<td>82.93%</td>
<td>0</td>
<td>3</td>
<td>3.62</td>
<td>3.21</td>
<td>0</td>
</tr>
<tr>
<td>Booker, Kendayin H</td>
<td>0</td>
<td>18</td>
<td>92.93%</td>
<td>0</td>
<td>40</td>
<td>2.62</td>
<td>3.29</td>
<td>0</td>
</tr>
<tr>
<td>Briddon, Glen A</td>
<td>2</td>
<td>1</td>
<td>91.92%</td>
<td>0</td>
<td>3</td>
<td>1.05</td>
<td>0.88</td>
<td>1</td>
</tr>
<tr>
<td>Bond, Fonter</td>
<td>0</td>
<td>11</td>
<td>91.92%</td>
<td>0</td>
<td>0</td>
<td>2.53</td>
<td>2.92</td>
<td>0</td>
</tr>
</tbody>
</table>
But some risks as pointed out by an extensive (US) literature
Understanding, researching and mitigating algorithmic bias
Strong knowledge base about countries’ practices and policies

Opportunities, guidelines and guardrails for effective and equitable use of AI in education

Unclassified - Non classifié
Defining algorithmic bias

Bias = a systematically better or lower AI algorithmic performance leading to some harm against one person or sub-population group.

Sources of biases

- Historical bias
- Representation bias
- Measurement bias
- Aggregation bias
- (Machine) Learning bias
- Evaluation bias
- Deployment bias

Harms from biases

- Allocative harms: withholding of or unfair distribution of some opportunity across sub-population groups
- Representational harms: representation in a negative light of some group (or withholding of positive representation of some group)

Suresh and Guttag, 2021, Baker et al. 2023
Algorithmic bias: the state of the situation and policy recommendations

- Researched bias in education (mainly in the US) = where model performance is substantially better or worse across mutually exclusive groups

- Areas: Dropout/failure/academic achievement prediction, automated essay scoring, speech evaluation, student affect, etc.

- Race (US): Usually less effective for Blacks and Hispanics (and also higher rates of false positives)

- Nationality (EU, rest of the world)

- Gender: Inconsistent results

- Few studies for many other sub-categories: Indigenous, rural/urban, non-native language speakers, special needs, military-connected, etc.

Unclassified - Non classifié
Recent example of education research demonstrating possible algorithmic bias by race in predicting student success (AERA Open, 10 July 2024)

Inside the Black Box: Detecting and Mitigating Algorithmic Bias Across Racialized Groups in College Student-Success Prediction

Denisa Gándara, Hadis Anahideh, and Lorenzo Picchiarini

Abstract

Colleges and universities are increasingly turning to algorithms that predict college-student success to inform various decisions, including those related to admissions, budgeting, and student-success interventions. Because predictive algorithms rely on historical data, they capture societal injustices, including racism. In this study, we examine how the accuracy of college student success predictions differs
Recent example of education research demonstrating algorithmic bias by race

AERA Open, 10 July 2024

- Student success prediction by Machine-Learning algorithms could be used for admissions or support service allocations
- They underpredict success (and overpredict failure) with less accuracy for Black and Hispanic students than for White and Asian
- Bias was difficult to eliminate with the different mitigation techniques used

Gandara et al., 2024
From unknown bias to known bias, from fairness to equity

1. Consider **algorithmic bias** in privacy policy and mandates so that privacy requirements do not prevent researchers/developers from identifying and addressing algorithmic bias.

2. Require **algorithmic bias analyses**, and thus related necessary data collection.

3. Guide algorithmic bias analysis based on **local context and local equity concerns**.

4. Fund **development of toolkits** for algorithmic bias in education.

5. Fund **research into unknown biases** around the world

---

Baker et al., 2023 (OECD DEO 2023)
OECD-Education International (EI) Opportunities, Guidelines and Guardrails
"Privacy and data protection must be balanced against other important educational objectives such as equity or effectiveness, which may require the collection of personal data, including sensitive ones."

- Better to avoid demographic characteristics in AI algorithms, when possible, BUT the collection of personal data is crucial to identify and address algorithmic bias and thus improve fairness.

- Countries should ensure that new digital tools are tested to avoid possible biases.

- Even in the absence of biases, as AI effectiveness is largely based on detecting “profiles”, the risk of human stigmatisation of students (or teachers) in different categories should be addressed.
Read the OECD Digital Education Outlook 2023