EDUCATION
& SKILLS

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

OECD
BETTER POLICIES FOR BETTER LIVES

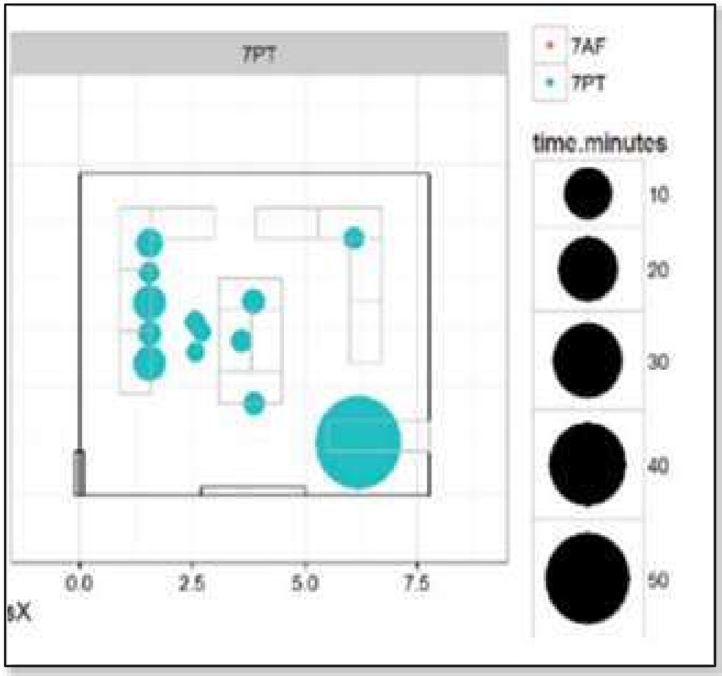
OECD Digital Education Outlook 2021

- 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

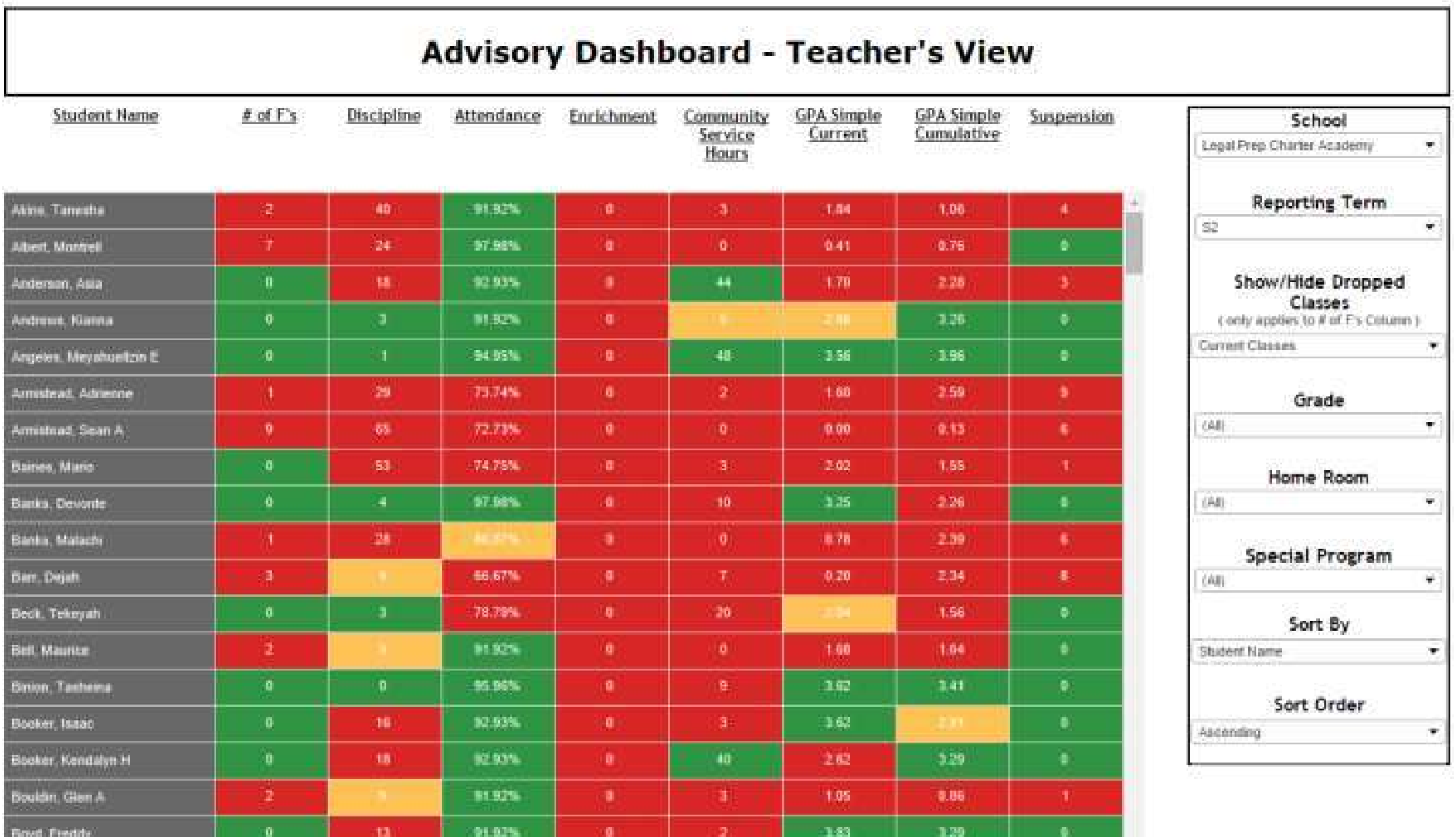
Teacher feedback for self-regulation

Showing teachers where they spend time in the classroom



Preventing dropout through early warning systems

Advisory Dashboard



Advisory Dashboard - Teacher's View

Student Name	# of F's	Discipline	Attendance	Enrichment	Community Service Hours	GPA Simple Current	GPA Simple Cumulative	Suspension
Akins, Tanesha	2	40	91.92%	0	3	1.84	1.06	4
Albert, Montrell	7	24	97.98%	0	0	0.41	0.76	0
Anderson, Asia	0	18	92.93%	0	44	1.70	2.28	3
Andrews, Kianna	0	3	91.92%	0	[illegible]	[illegible]	3.26	0
Angeles, Meyahueltzin E	0	1	94.95%	0	48	3.56	3.96	0
Armstead, Adrienne	1	29	73.74%	0	2	1.60	2.59	9
Armstead, Sean A	9	65	72.73%	0	0	0.00	0.13	6
Baines, Mario	0	53	74.75%	0	3	2.02	1.55	1
Banks, Devonte	0	4	97.98%	0	10	3.25	2.26	0
Banks, Malachi	1	28	[illegible]	0	0	0.78	2.39	6
Barr, Dejah	3	[illegible]	66.67%	0	7	0.20	2.34	8
Beck, Tekeyah	0	3	78.79%	0	20	[illegible]	1.56	0
Bell, Maurice	2	[illegible]	91.92%	0	0	1.68	1.04	0
Binion, Tasheina	0	0	95.96%	0	9	3.62	3.41	0
Booker, Isaac	0	16	92.93%	0	3	3.62	[illegible]	0
Booker, Kendalyn H	0	18	92.93%	0	40	2.62	3.29	0
Bouldin, Glen A	2	[illegible]	91.92%	0	3	1.05	0.86	1
Boyd, Freddy	0	13	91.92%	0	2	3.83	3.29	0

School
Legal Prep Charter Academy

Reporting Term
S2

Show/Hide Dropped Classes
(only applies to # of F's Column)
Current Classes

Grade
(All)

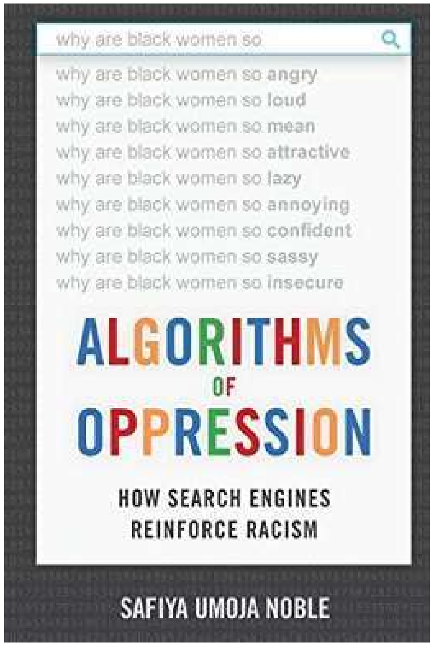
Home Room
(All)

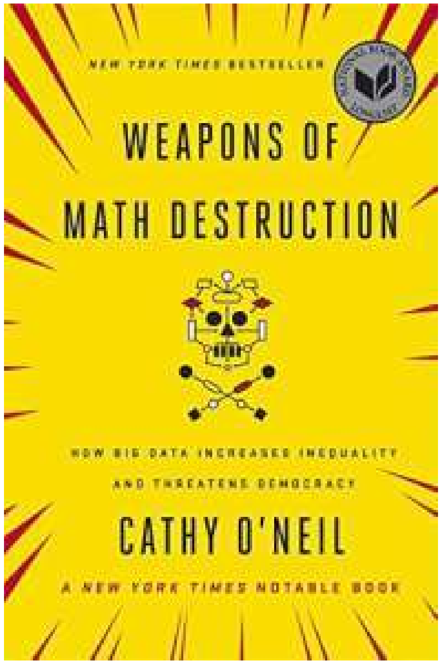
Special Program
(All)

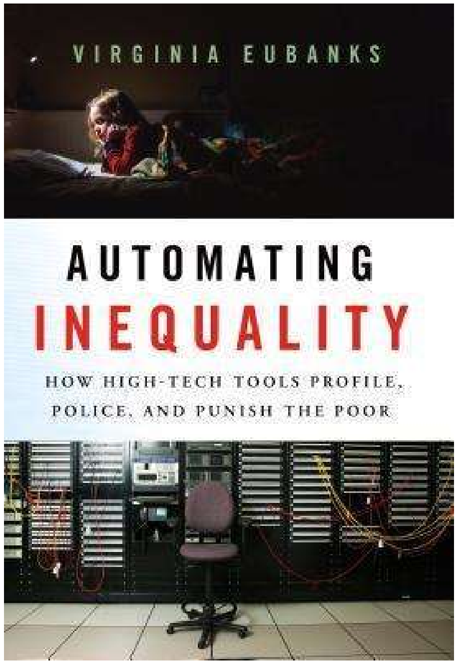
Sort By
Student Name

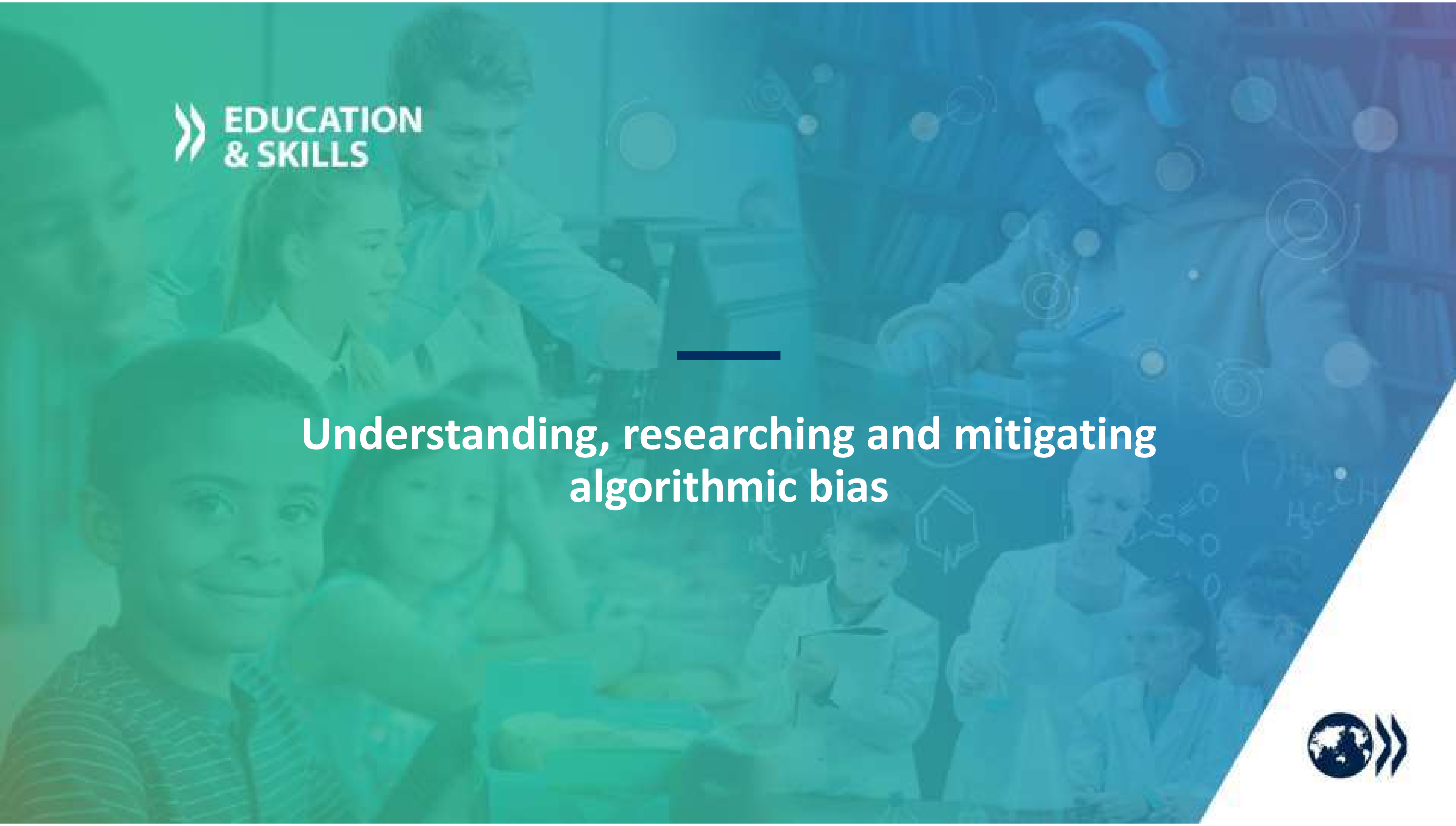
Sort Order
Ascending

But some risks as pointed out by an extensive (US) literature





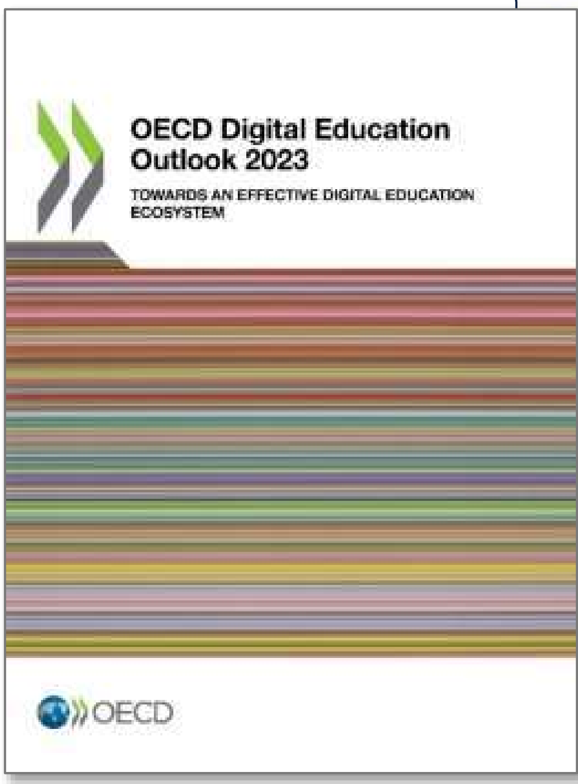


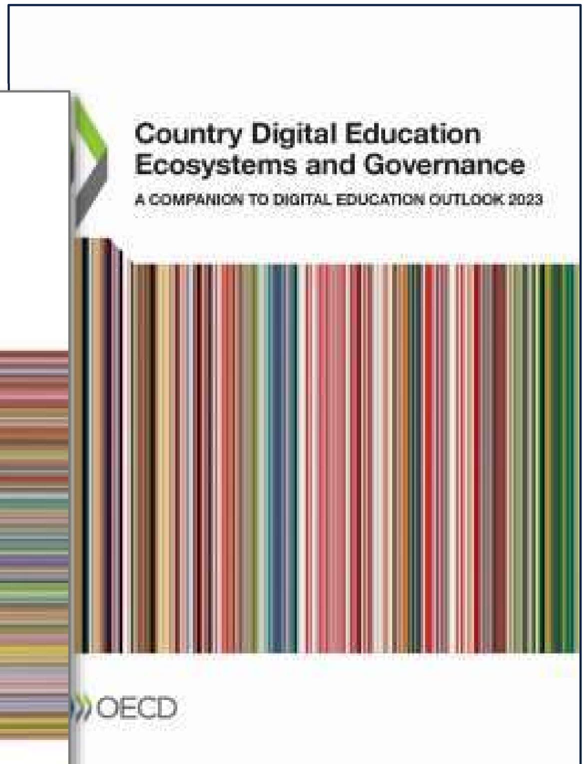


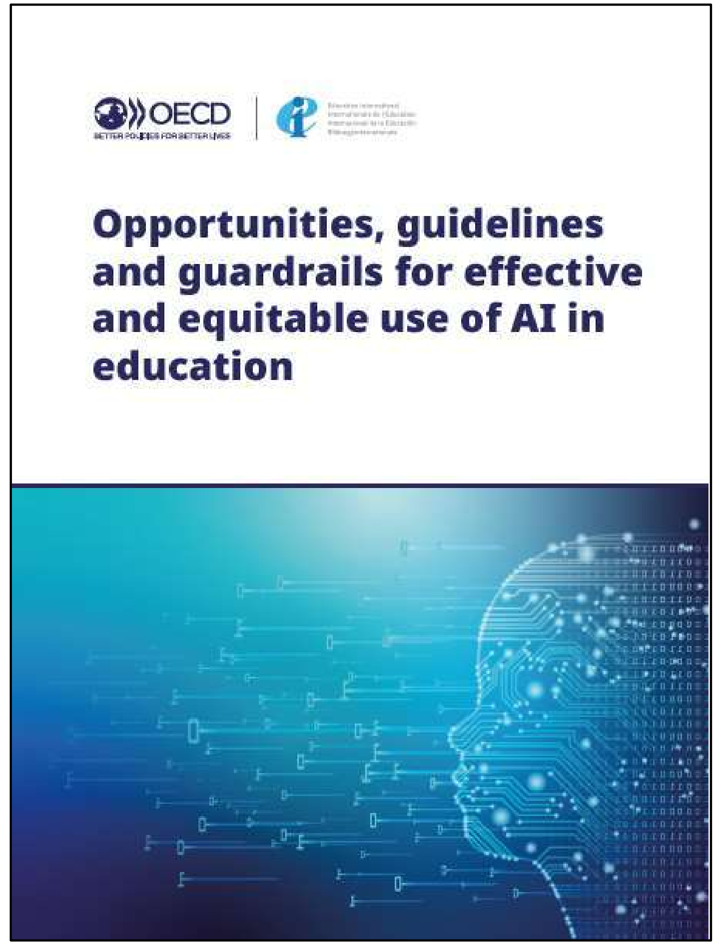
Understanding, researching and mitigating algorithmic bias

OECD Digital Education Outlook 2023

Strong knowledge base about countries' practices and policies







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 » by Ryan Baker, Aaron Hawn and Seiyon Lee

242 |

9 Algorithmic bias: the state of the situation and policy recommendations

Ryan S. Baker, M. Aaron Hawn and Seiyon Lee

University of Pennsylvania

This chapter discusses the current state of the evidence on algorithmic bias in education. After defining algorithmic bias and its possible origins, it reviews the existing international evidence about algorithmic bias in education, which has focused on gender and race, but has also involved some other demographic categories. The chapter concludes with a few recommendations, notably to ensure that privacy requirements do not prevent researchers and developers from identifying bias, so that it can be addressed.

OECD DIGITAL EDUCATION OUTLOOK 2023 © OECD 2023

- 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.

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

AERA Open

Impact Factor: 3.5
5-Year Impact Factor: 3.8

JOURNAL HOMEPAGE SUBMIT PAPER

Open access | Research article | First published online July 10, 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 +1 View all authors and affiliations

All Articles | https://doi.org/10.1177/23328584241258741

Contents | PDF / ePub | Cite article | Share options | Information, rights and permissions | Metrics and citations | Figures and tables

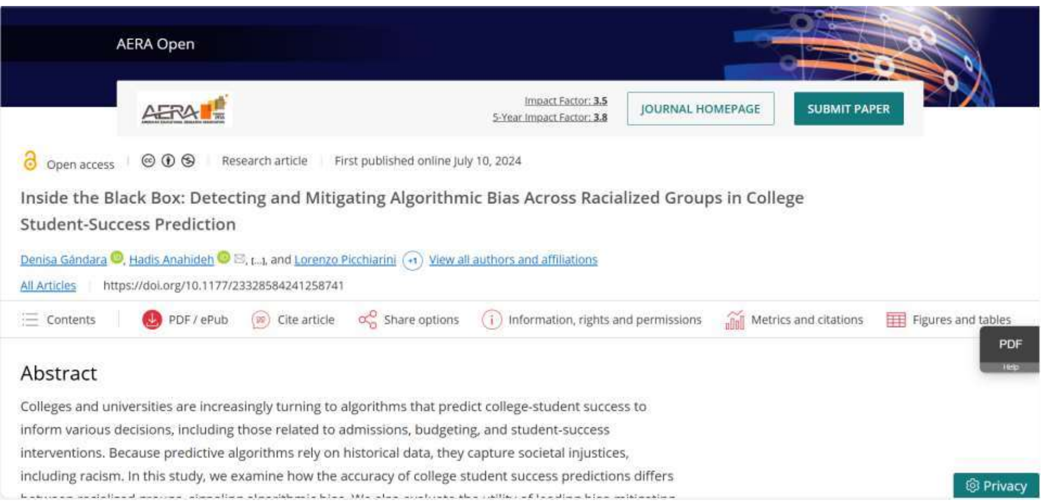
PDF
Help

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

Privacy

Recent example of education research demonstrating algorithmic bias by race
AERA Open, 10 July 2024


AERA Open

Impact Factor: 3.5
5-Year Impact Factor: 3.8
JOURNAL HOMEPAGE
SUBMIT PAPER

Open access | Research article | First published online July 10, 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 (+1) View all authors and affiliations

All Articles | https://doi.org/10.1177/23328584241258741

Contents | PDF / ePub | Cite article | Share options | Information, rights and permissions | Metrics and citations | Figures and tables

PDF

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

Privacy

- Student success prediction by Machine-Learning algorithms could be used for admissions or support service allocations
- They underpredict success (and overpredict failture) with less accuracy for Black and Hipanic 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

Unknown bias → Known bias → Fairness → Equity

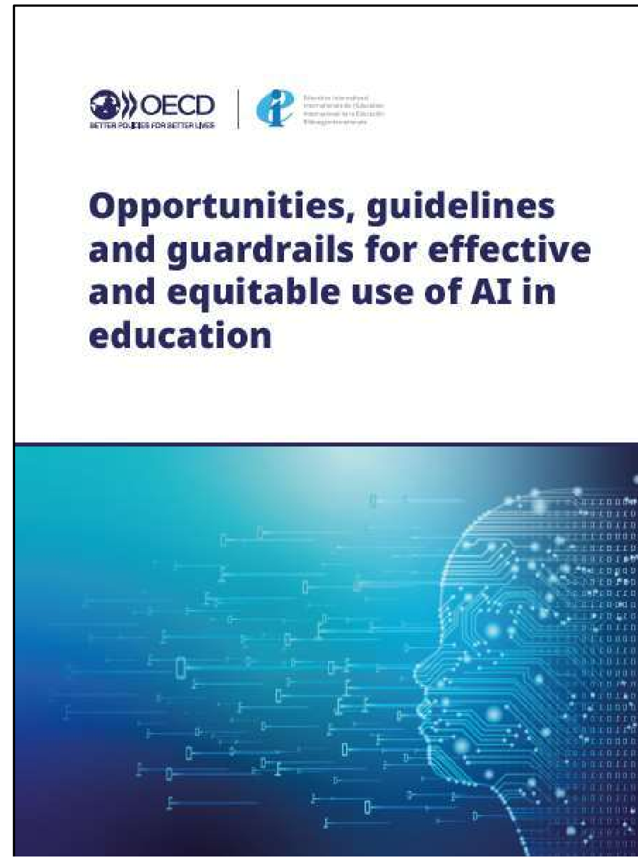
Baker et al., 2023 (OECD DEO 2023)

EDUCATION
& SKILLS

OECD-Education International (EI) Opportunities, Guidelines and Guardrails

7: Ethics, safety and data protection

"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.

EDUCATION
& SKILLS

Read the OECD Digital Education Outlook 2023