Project Update 1-Bias in Online Education Platforms

Ayush Dhoundiyal
ayushdh96@gmail.com

Drashti Patel
drashti202@gmail.com

1 Problem Statement

AI-driven recommendation mechanisms in web-based learning platforms such as Coursera, Khan Academy, and edX personalize learning experiences through course and resource suggestions to learners based on their past interaction and demographic details. These recommendation mechanisms, nonetheless, tend to introduce implicit gender-based, ethnicity-based, region-based, and socio-economic status-based biases, thus making the learning opportunity not level for certain groups of students.

The project aims to investigate the bias in such AI systems, determine the impact on student learning and engagement, and suggest ways to reduce bias so that equal and fair access to learning opportunities can be provided to all students.

2 Motivation

Bias detection is critical in AI-driven systems, such as online learning websites, where computerized recommendations can have the potential to play a significant role in students' educational outcomes. Because these sites employ demographic data (gender, ethnicity, socio-economic status, etc.) as the foundation of their course offerings, there are opportunities that entrenched bias in the accessible data can tint the content selection for certain kinds of students. This can result in unequal educational opportunities and act to perpetuate entrenched educational disparities.

To find and address such biases is paramount to making fairer and more inclusive learning spaces. Through a determination of where AI systems tend to unknowingly favor some populations, this work seeks to help all students, no matter what their background may be, equally access personalized pathways of learning so that eventually diverse student groups might have better education outcomes.

3 Ethical Question to address?

The main moral question your project will be working on is:

"How do we discover and combat biases in AI-powered recommendation platforms on online schooling sites to ensure that these recommendation platforms provide even, inclusive, and unbiased schooling opportunities to students regardless of their demographic background?"

We address the main ethical challenges of: Detecting different bias in education systems dataset affecting certain groups .

4 Current Results

Upto now, we have worked with the first dataset, OULAD, conducting thorough analysis to spot biases in learning platforms online. Our approach was a two-part process: Exploratory Data Analysis (EDA) first in order to keep track of demographic distributions and student performance to manually check and get a first hand look at the data, then statistical bias detection with fairness metrics. This helped us systematically detect discrepancies in rates of student achievement among different groups of demographics.

Our bias detection framework derived from the OULAD dataset revealed differences ranging from trivial to substantial levels with respect to various demographic characteristics. Gender bias was trivial, with a Disparate Impact (DI) of 1.0487 and a Statistical Parity Difference (SPD) of 0.0225, indicating that male and female students were nearly equal in terms of their success. Regional bias (DI: 0.9599, SPD: -0.0196) was present, suggesting that students from less represented regions had marginally lower success rates than students from more common regions. The largest disparities were observed in disability status and education level. Lower prior education (DI: 0.8369, SPD: -0.0849) and disability (DI: 0.7918, SPD: -0.1003) students were highly disadvantaged, reflecting systemic inequality in AI-based educational attainment. Age bias (DI: 1.1629, SPD: 0.0734) showed that students in the most common age category performed better than all other groups, reflecting a potential bias in learning opportunities.

Intersectional bias detection also underscored compounded disparities. The region + education bias (DI: 1.1936, SPD: 0.0771) shows that lower education level, underrepresented region students are more disadvantaged. Similarly, the age + disability bias (DI: 1.0841, SPD: 0.0385) reveals that younger, non-disabled students have a better performance, highlighting issues of accessibility gaps. These results, as shown in Table 1, highlight the most prominent biases detected in the dataset.

Table 1: Bias Metrics for Different Attributes For OULAD Dataset

Bias Type	Disparate Impact	Statistical Parity Difference
Gender Bias	1.0487	0.0225
Region Bias	0.9599	-0.0196
Education Bias	0.8369	-0.0849
Disability Bias	0.7918	-0.1003
Age Bias	1.1629	0.0734
Gender + Disability Bias	0.9977	-0.0011
Region + Education Bias	1.1936	0.0771
Age + Disability Bias	1.0841	0.0385

The link to our github repository is as follows : Github Repo

5 Upcoming Results

Next, we will expand our analysis to the other two datasets, ASSISTments and EdX, to make a comparison of biases across different online learning platforms. This will allow us to determine whether similar demographic differences occur across datasets or if specific platforms exhibit unique patterns of bias based on the type of their user base and models of learning.

Once individual bias tests are complete, we will proceed to compare the results from all three datasets, conducting a comparative study to investigate how different platforms handle learning recommendations for students with different demographics. Cross-dataset comparison will provide more insights into whether patterns of bias are platform-specific or widespread across AI-based education systems.

6 Analysis Done

Current Analysis: On analysing the results we can draw down certain conclusions. While gender bias was minimal, students with less prior education and disabilities were disadvantaged significantly, raising issues of equitable access to learning opportunities.

AI models may favor academically gifted pupils, unintentionally reinforcing educational inequalities rather than providing individualized assistance for pupils needing intervention. Regional bias means pupils from underrepresented regions receive less precise recommendations, likely due to data imbalances favoring dominant regions.

Intersectional biases also compound disadvantages—students who belong to underrepresented locations and have lower education levels suffer from double disadvantages, while age and disability bias suggest accessibility problems in AI-assisted learning pathways. These trends foreshadow that AI models are not fully inclusive, limiting personalized learning for marginalized groups.

7 Upcoming Analysis

1.We will analyze different sets of dataset attributes to determine the factors that most significantly affect bias in AI-driven educational results.

2.By considering various attribute interactions, we will find out how various demographic attributes contribute toward generating differences in learning recommendations.

3.Carry out statistical comparision of our findings across the three datasets, examining whether bias patterns are consistent across platforms or dataset-specific.

4.Develop understanding, through our comparative analysis, what traits make the greatest contribution towards bias and how they vary across different learning arrangements.

Based on these conclusions, we will further elaborate on the discussion, including the broader implications of these biases on student participation, success rates, and access to learning.This deep analysis in education dataset shall help developers to effectively develop bias reduction strategies, rendering AI-based education systems equitable and inclusive.

8 Problems Encountered

The problems we encountered were not major, but we had to conduct extensive EDA to understand the dataset, identify potential biases, and determine whether bias existed in the first place. The main challenge we faced was combining different columns to analyze bias within their interactions and extracting meaningful insights from these combinations.

9 Conclusion

We reveal biases by education level, disability status, and geographic representation across AI-powered learning platforms. We hope our study and analysis for comparing bias trends, and explore mitigation techniques leads to conclusions for improving AI-delivered education systems for all students.