Current Approaches: Empirical Evidence for Bias

Empirical Evidence on Bias in Adaptive Systems

- Differences in design and demographic differences in outcomes
    - Using female characters as pedagogical agents beneficial for female students (Arroyo et al., 2013)
    - Affective feedback delivered by the pedagogical agents was more effective for students with learning disabilities (Woolf et al., 2010)

Empirical Evidence on Bias in Adaptive Systems

- Differences in design and demographic differences in outcomes
- Cultural mismatches in student behaviors and classroom practices as possible explanations for demographic differences
  - Collaborative use of an adaptive system in Latin American and Asian countries that was developed in the United States for individual use (Ogan et al., 2012, 2015)
  - Consistently less off-task behavior in the Philippines (Rodrigo et al., 2013)

Empirical Evidence on Bias in Adaptive Systems

- Differences in design and demographic differences in outcomes
- Cultural mismatches in student behaviors and classroom practices as possible explanations for demographic differences
- Some discussion (non-empirical) on how research methods may lead to biases in adaptive systems
  - Disconnect in implementing methods developed in the western context to a low-income global south country (Andres et al., 2015)
  - Differing needs of students across globe and “limited cultural diversity” in the student population investigated in empirical work (Blanchard, 2015)

Empirical Evidence on Bias in Adaptive Systems

- Differences in design and demographic differences in outcomes
- Cultural mismatches in student behaviors and classroom practices as possible explanations for demographic differences
- Some discussion (non-empirical) on how research methods may lead to biases in adaptive systems
- Overcoming bias by adapting design to students’ needs
    - Better performance of third graders in science when the adaptive system used a similar dialect as the native tongue of the students (African American Vernacular English) (Finkelstein et al., 2013)

Empirical Evidence on Algorithmic Bias

- Mostly on race, gender, and nationality as the demographic categories
- Likely problematic to assume population validity in algorithmic systems

Study	Subgroups	Prediction Task	Finding on Bias
Hu & Rangwala, 2020	Gender, Race	At-Risk (course)	Worse for African American & male students
Yu et al., 2020	Gender, Race	College success	Demographics as predictor led to inaccurate predictions for female students & some racial groups
Lee & Kizilcec, 2020	Gender, Race	Course grade	Equity-corrected version does better with underrepresented racial/ethnic groups & male students
Anderson et al., 2019	Gender, Race	Graduation	Higher false-positive for White students & higher false-negative for Latino & male students
Kai et al., 2017	Gender, Race	Online college retention	Varies by algorithm
Bridgeman et al., 2009, 2012	Gender, Nationality	Essay scoring	More accurate for male students and higher scored for Chinese and Korean students
Ogan et al., 2015	Nationality	Learning Outcome	Models on data from same country more accurate than the data from other countries

Baker and Hawn (2021)

Empirical Evidence on Algorithmic Bias

- Mostly on race, gender, and nationality as the demographic categories
- Likely problematic to assume population validity in algorithmic systems
- A few studies on other demographic categories
    - native language and dialect, disabilities, urbanicity, parental educational background, socioeconomic status, international students, and military-connected status

Baker and Hawn (2021)

Empirical Evidence on Algorithmic Bias

- Mostly on race, gender, and nationality as the demographic categories
- Likely problematic to assume population validity in algorithmic systems
- A few studies on other demographic categories
- Much more work is needed to understand subpopulation differences
  - no studies on non-binary gender identities or any other categories of LGBTQ identities
  - only one study examined data on indigenous students (model highly unstable)
  - some categorizations maybe oversimplified or politically influenced (e.g., combining several distinct communities such as Sri Lankan, Korean, and Vietnamese into Asian; Strmic-Pawl et al., 2018)
  - research primarily from groups in the United States - lack of voice from other contexts

Baker and Hawn (2021)

Empirical Evidence on Algorithmic Bias

- Mostly on race, gender, and nationality as the demographic categories
- Likely problematic to assume population validity in algorithmic systems
- A few studies on other demographic categories
- Much more work is needed to understand subpopulation differences

What are some limitations and challenges in auditing your example model(s) for bias? Try to be as specific as you can.

Discussion Board Question