



Week 6

MORPH Algorithmic Fairness



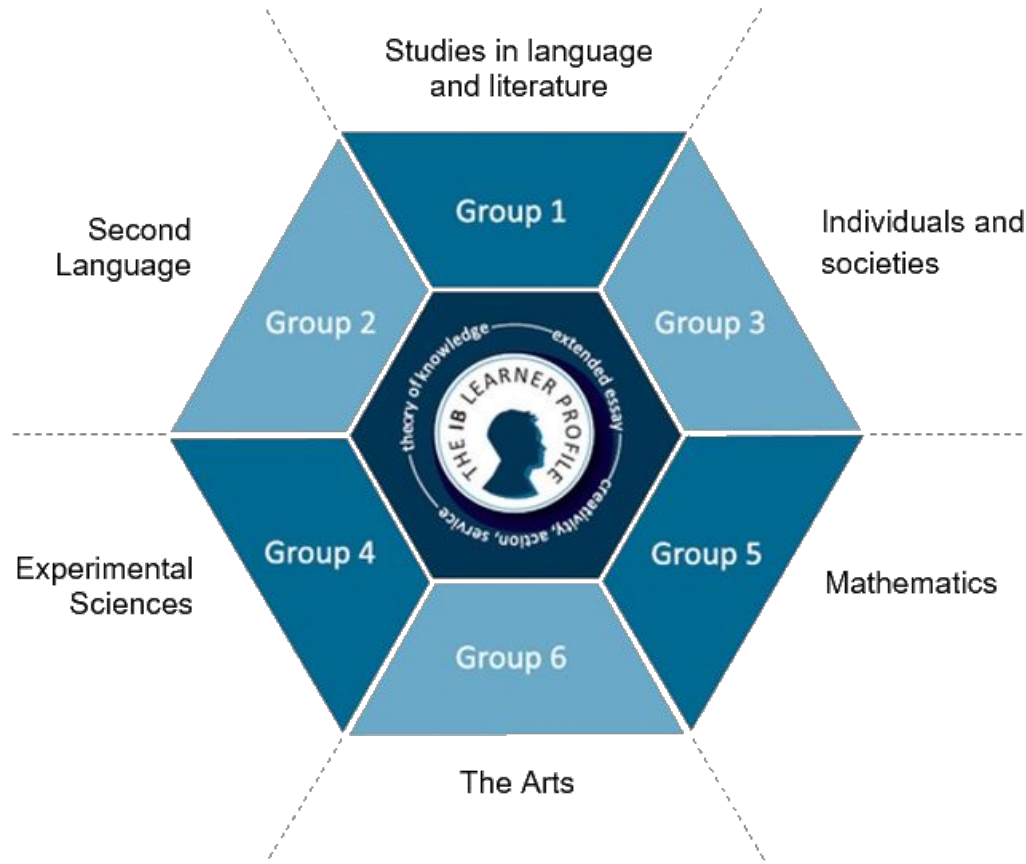
Agenda: Deepfakes

- This Week in Fairness
- Deepfakes
 - Introduction
 - Video demos from recent research
 - Discussion
- Project brainstorm



The image features a large white circle in the center, set against a black background. To the left of the white circle, there is a stack of three overlapping circles: a dark grey one in the back, a medium grey one in the middle, and a light grey one in the front. To the right of the white circle, there is a series of five concentric white circles of varying diameters, creating a ripple effect. The text "This Week in Fairness" is centered within the white circle in a bold, black, sans-serif font.

This Week in Fairness



International Baccalaureate (IB) program



International Baccalaureate (IB)

- Cancelled their spring exams due to COVID-19
- Uses a formulaic algorithm instead to “predict” scores
 - Predicted grades would count towards the final grade
- Determines what colleges you can get into





**What kind of
algorithm, you ask?
Hmm...**





The system used signals including a student's grades on assignments and grades from past grads at their school to predict what they would have scored had the pandemic not prevented in-person tests. — WIRED



International Baccalaureate (IB)

- Many students around the world received significantly lower final grades than what they were predicted
- Model = complete black box; IB has not revealed how the algorithm was **designed, tested, implemented**, etc.



What problems with fairness / bias could there be?

- Past graduates of your school = not exact indicator of individual student
- Proxies for class/income
- Results for smaller classes vs larger classes
- Classes with more assignments than fewer, etc.
- More established schools vs not





To read more:

WIRED: Meet the Secret Algorithm That's Keeping Students Out of College

<https://www.wired.com/story/algorithm-set-students-grades-altered-futures/>

Inside Higher Ed: What's Wrong with This Year's IB Scores?

<https://www.insidehighered.com/admissions/article/2020/07/13/algorithm-used-ib-scores-year-blamed-students-low-marks>







Deepfakes




Deepfakes: Potential Effects

- Disinformation – people are more inclined to share news/videos confirming our viewpoints
 - Fake sex tapes/porn can ruin someone's lives forever
 - 98% of all deepfakes are sex videos
 - Fake audio of business owners → ruin businesses
 - Undermine trust before elections, IPOs, etc.
- 
- 



Deepfakes: Potential Effects

- Court trial process impacted – hard for humans to differentiate, so cannot blindly trust audio/video anymore
 - Preserve memories and behavior of lost loved ones
 - Help those who cannot / can no longer speak
- 





Something different this week: we'll talk about papers & research, but through video demos

Go to www.potato.land
with the link I send in Zoom



Video demonstrating both pros
and cons of deepfakes:

<https://www.youtube.com/watch?v=gLoI9hAX9dw>

Audio deepfake (Google AI, 2019):

<https://www.youtube.com/watch?v=0sR1rU3gLzQ>

Making deepfakes (2020):

<https://www.youtube.com/watch?v=mUfJOQKdtAk>

AI Deepfake detectors (2019):

<https://www.youtube.com/watch?v=RoGHVI-w9bE>



Discussion

- What biases can be embedded in the deepfake creation process?
- What, if anything, should be done about deepfakes?
 - How to regulate?
- If we reach a point where video/audio are no longer considered 100% verifiable evidence, what should we use as the standard of undoubtable proof in court?





Discussion

- One advantage of deepfakes is the ability to preserve memories or personalities of someone after disability or death
 - Do you think it is ethical to do this? Especially for after death?
 - What if it extended beyond deepfake videos/audios to real robots?



The image features a large white circle centered on a black background. Inside this circle, the words "Project" and "Brainstorm" are written in a bold, black, sans-serif font, stacked vertically. To the left of the white circle, there is a cluster of overlapping circles in various shades of gray. To the right, there is a series of concentric white circles of varying diameters, creating a ripple effect.

Project Brainstorm