Logistics: Project

- Project presentation
  - Dec 6/8 during lectures
  - Everyone is expected to attend both lectures
  - 10 minutes for presentation + 1~2 minutes for QA and transition

- Project reports
  - Due: Dec 9 (no late submissions)
  - Up to 6 pages (plus additional pages for only references/citations)
  - No strict format requirements
    - You are encouraged to use standard templates, such as AAAI format or NeurIPS format
  - More details will be announced later

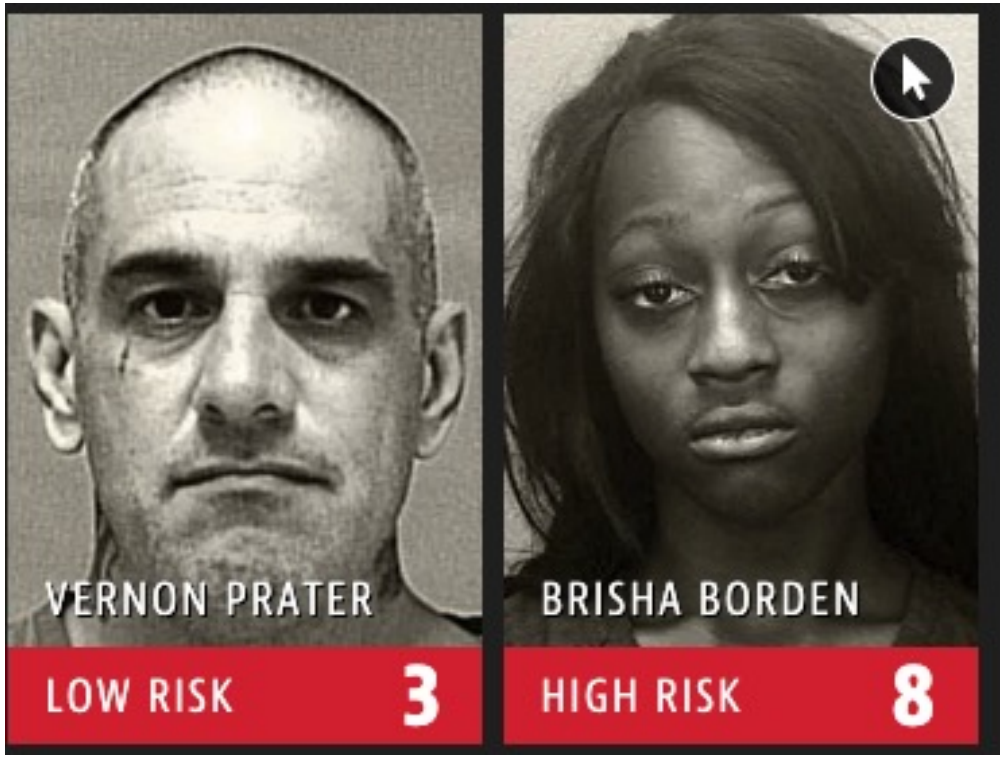
Peer Review

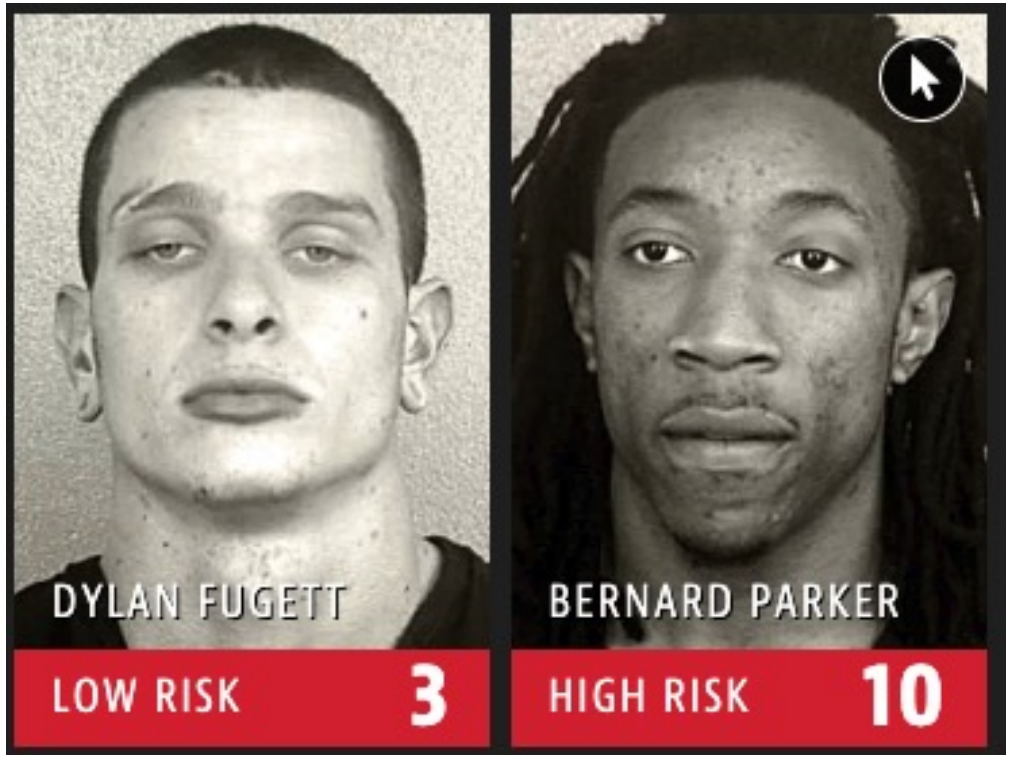
- Please submit the peer review by 6pm

Lecture 19
Human Perceptions of Fairness

Instructor: Chien-Ju (CJ) Ho

More Discussion on Recidivism Prediction





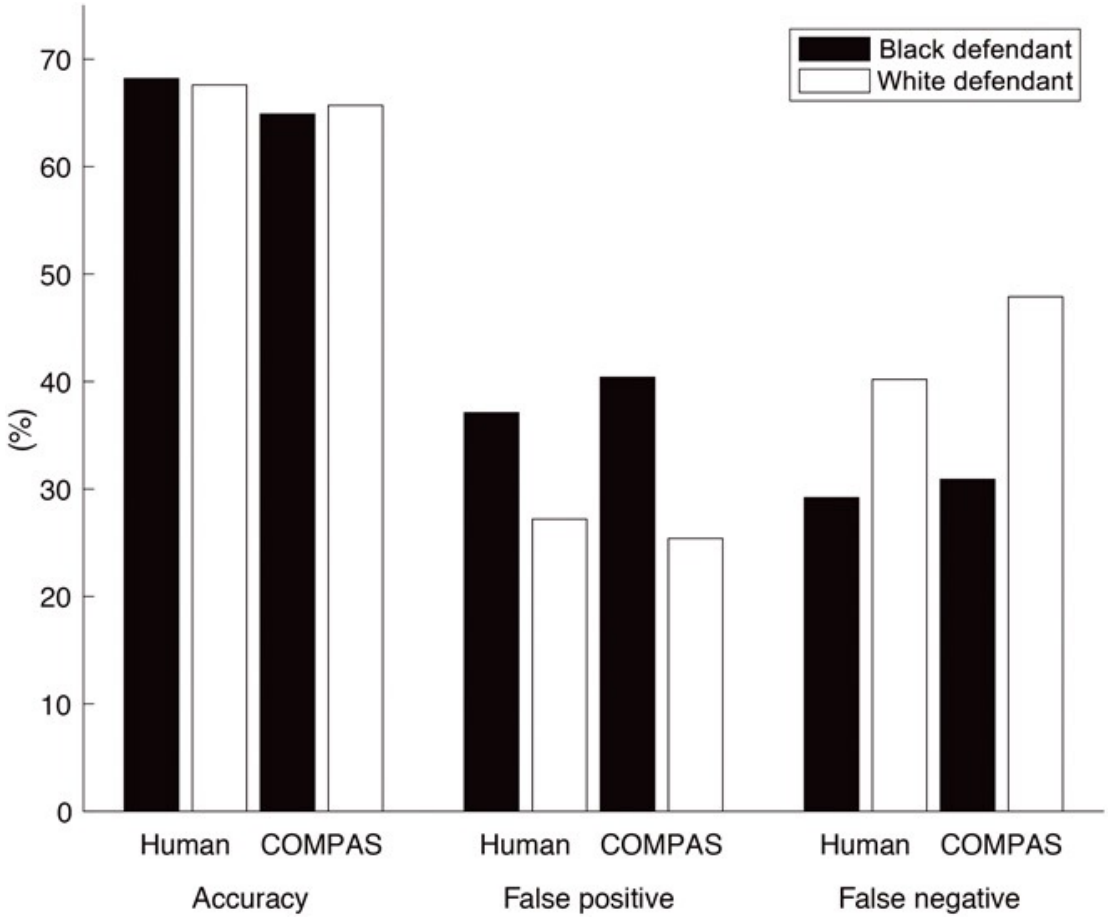
Human Compared with ML [Dressel and Farid. 2018]

The defendant is a 24-year-old female. She has had many prior arrests and convictions, starting at age 14, and has spent time in prison.

In this jurisdiction, 29% of defendants get rearrested for another crime on average.

How likely do you think this person is to be arrested for another crime within the next 2 years?

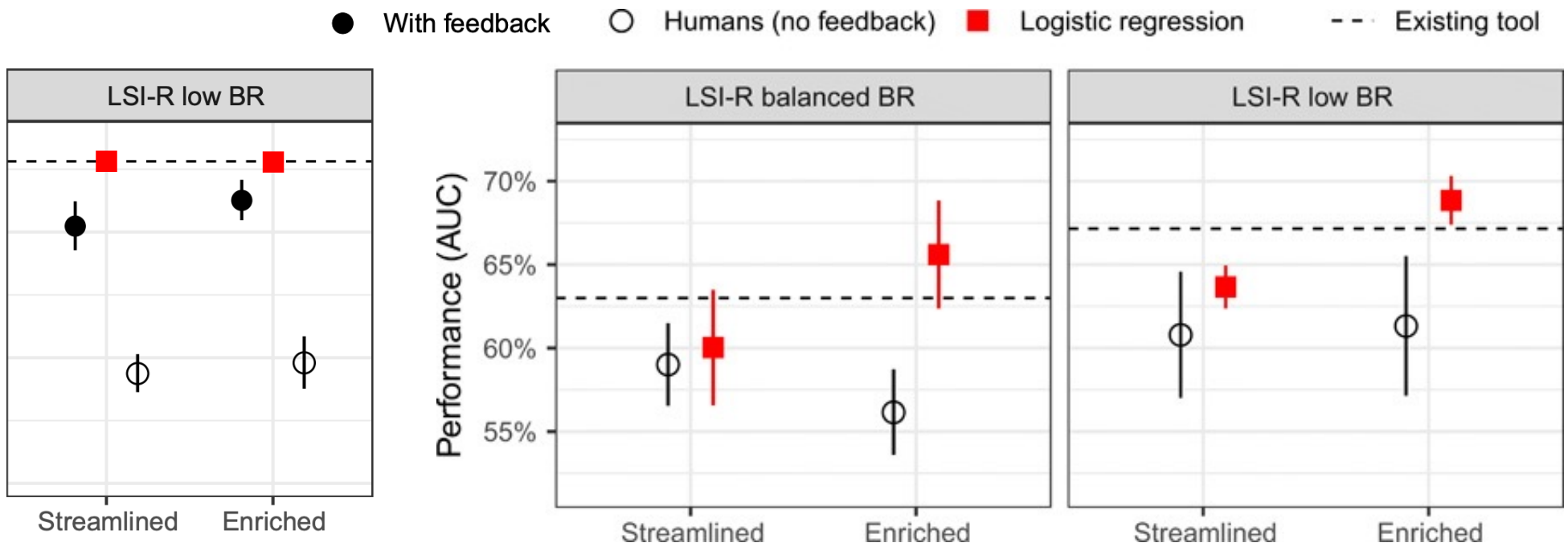
Human Compared with ML [Dressel and Farid. 2018]



- Humans are closely on par with ML for this task in both accuracy and fairness.
- Caveat:
  - Humans are learning (seeing ground truth) in this experiment

A Follow-Up Study [Lin et al. 2020]

- Change three factors
  - Whether humans obtain truth feedback
  - Base rates of the recidivism
  - Amount of information given to users



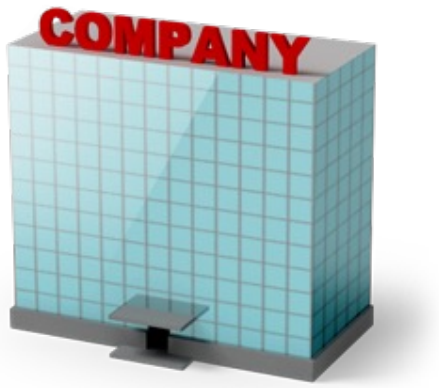
Do you think it's reasonable to treat the outcomes of the algorithm as a final decision? Or to use the algorithm outcomes as a reference and let people make the final decision. Why?

- A common theme
  - Giving the control to humans
  - The goal of AI is to **augment**, not **replace**, humans

As a human, this sounds good….

Think About Problem 3 in Assignment 3

- Company aims to hire interns
  - Only 30% students satisfy the requirement and the company knows that
  - I can make them hire 60% by manipulating the information



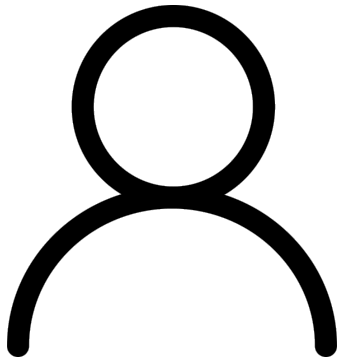
Think About Problem 3 in Assignment 3

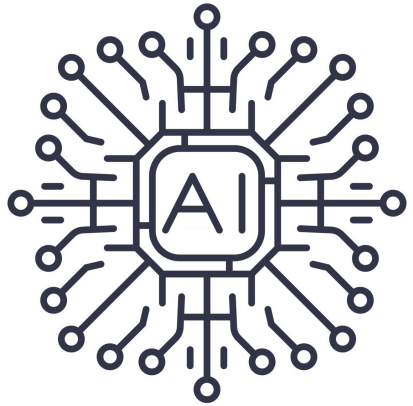
- Company aims to hire interns
  - Only 30% students satisfy the requirement and the company knows that
  - I can make them hire 60% by manipulating the information

- The example in the original paper [Kamenica and Gentzkow. 2011]
  - Company -> Judge
  - Teacher -> Prosecutor

Think About Problem 3 in Assignment 3

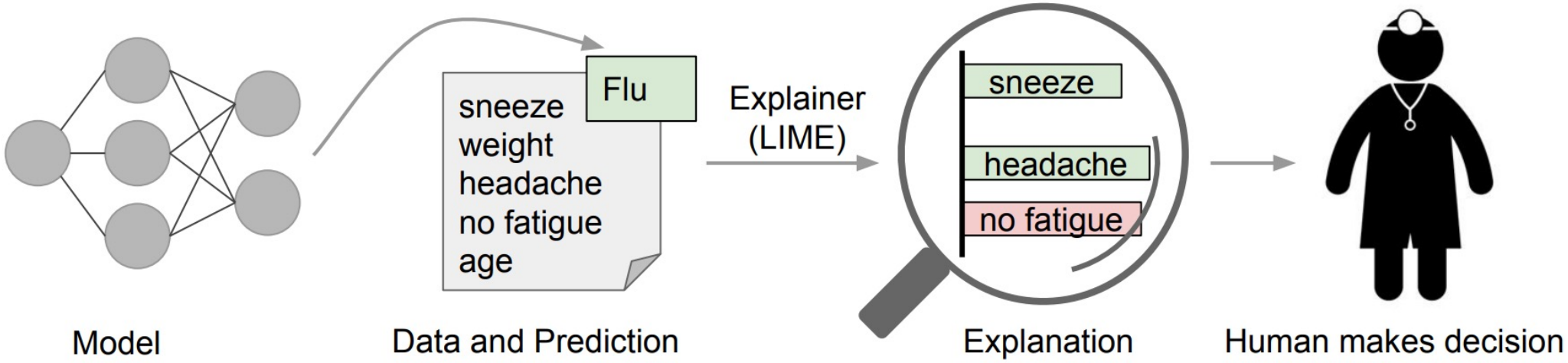
- Company aims to hire interns
  - Only 30% students satisfy the requirement
  - I can make them hire 60% by manipulating the information
- The example in the original paper [Kamenica and Gentzkow. 2011]
  - Company -> Judge
  - Teacher -> Prosecutor
- Let’s extend it a bit
  - Company -> Human
  - Teacher -> AI





AI-Assisted Decision Making

- We will discuss these at the last two lectures



AI-Assisted Decision Making

- The idea is nice => AI is helping humans to make decisions
- AI might also be able to manipulate humans to make decisions
  - Think about recommendation letter writing
- Let's just ask AI to present all information
  - Exceed the human cognitive load
- Who to decide what AI should do?

Participatory Design (more on Thursday)