

Computer Vision API Prejudice Audit

Chris Dulhanty, MSc Systems Design Engineering, Waterloo

Alexander Hinton, MA Economics, UoG

Andrew Nixon, MA Management, UoG

LOCAL
H[⚡]C^K
DAY

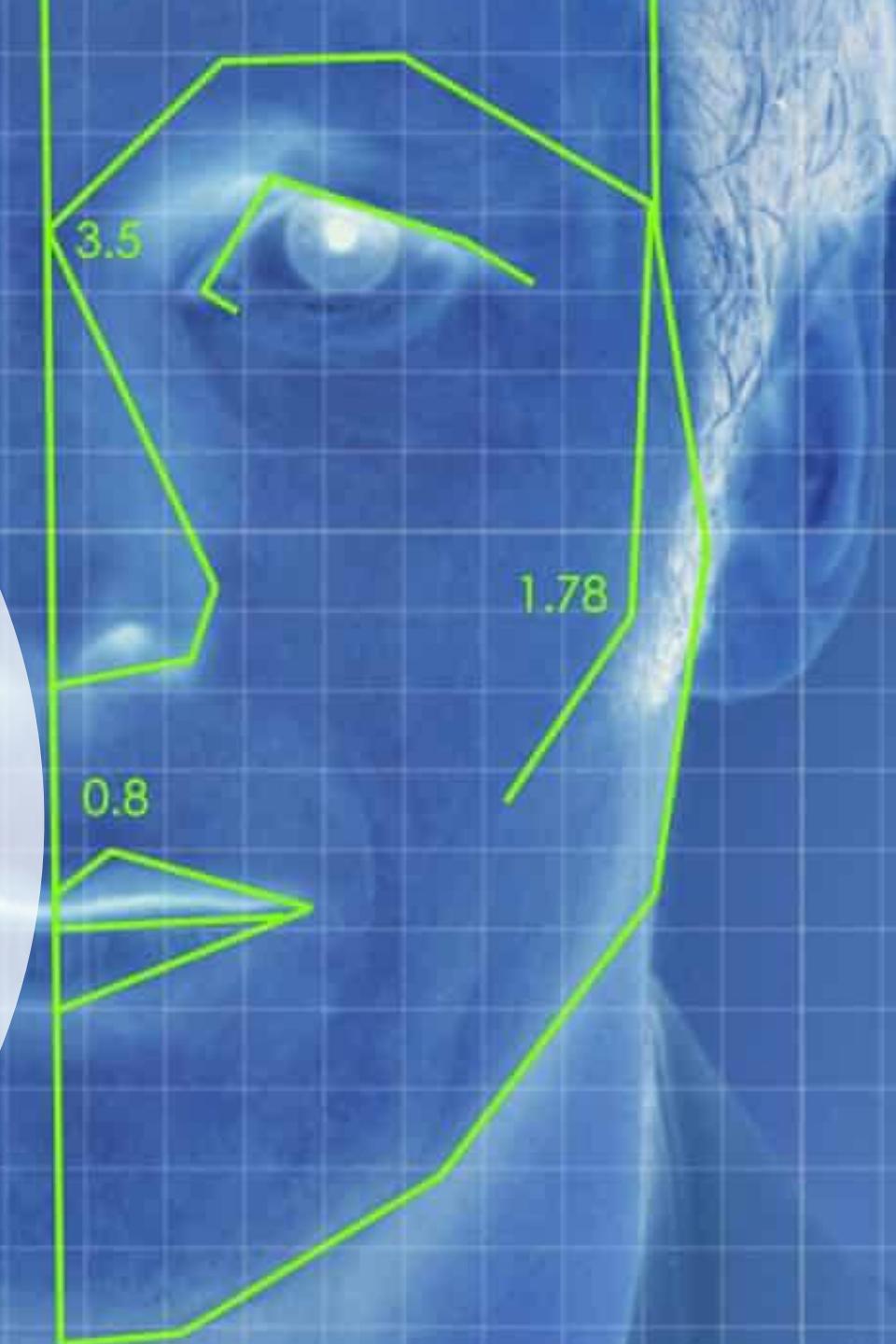
HOSTED BY GitHub

45
50
55
60
65
70
75
80
85
90
95

Is AI always the answer?

Applications of AI have run wild:

- Immigration
- Policing
- Judicial decisions
- HR
- IQ
- Probability of being a pedophile



**Amazon
Rekognition**



 **Google Cloud**
Vision API



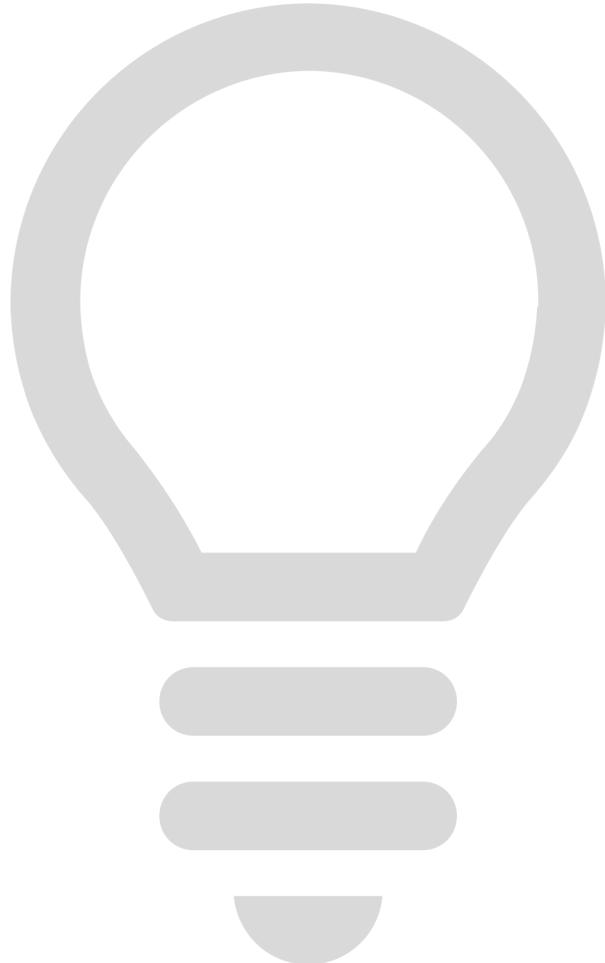
Today's Hack

- Major tech companies are rolling out enterprise solutions that leverage their computer vision APIs
- These tools are being applied to situations that could influence your career prospects



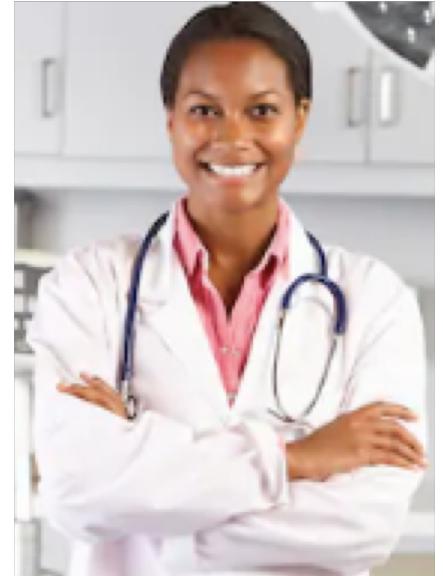
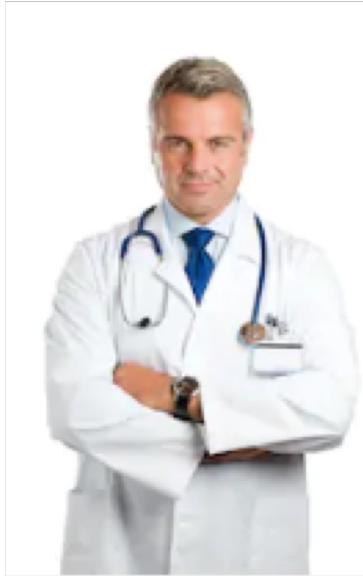
Research Question:

Will image classifiers give different predictions about stock images of doctors split by race and gender?



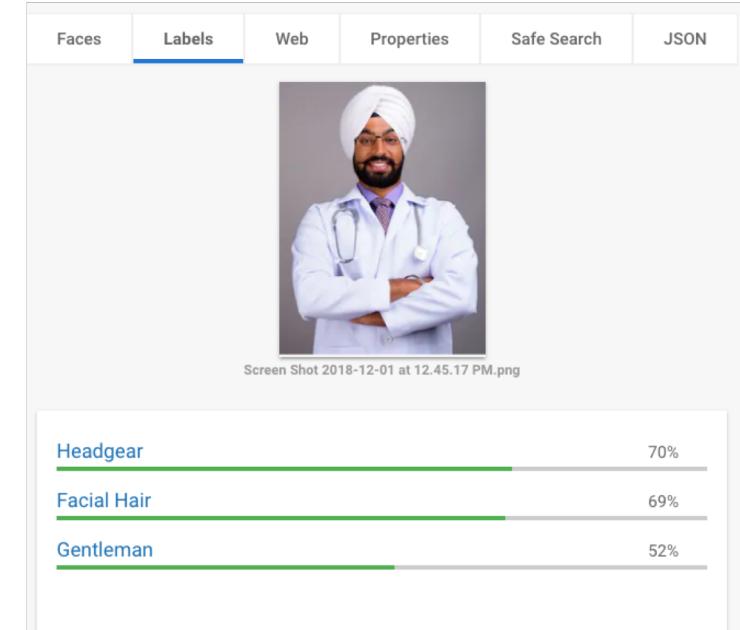
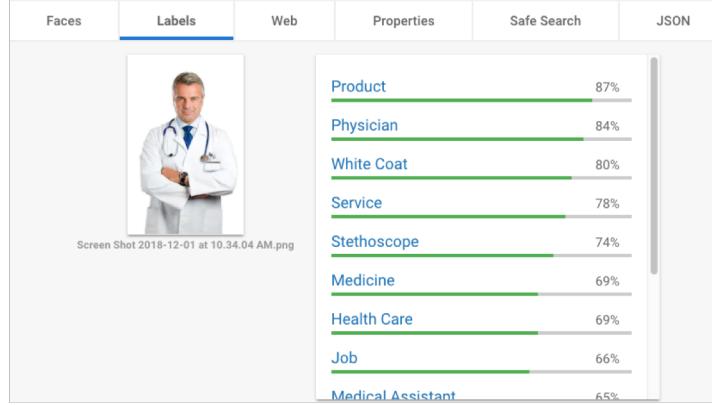
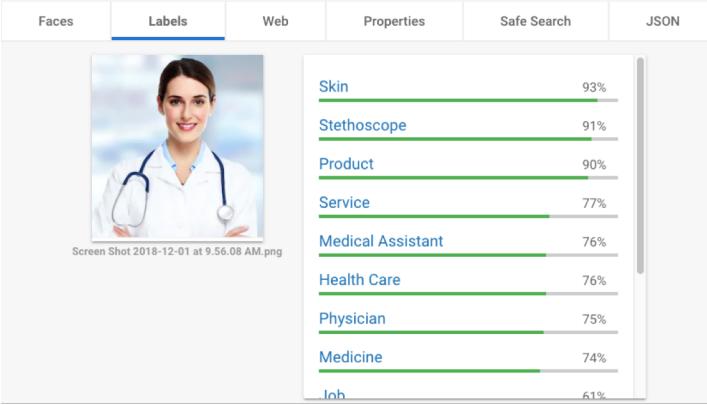
Methodology

- Collected stock images of doctors
- Male
 - White n = 31
 - Asian n = 29
 - Black n = 30
 - Brown n = 30
 - Total n = 123
- Female
 - White n = 30
 - Asian n = 33
 - Black n = 30
 - Brown n = 31
 - Total n = 124



Methodology

- Created accounts with the four CV APIs
- Ran the 247 images through 3 of the APIs and extracted the predictions of physician/doctor
- Evaluated differences across the eight separate groups
 1. Male_Asian
 2. Male_Black
 3. Male_Brown
 4. Male_White
 5. Female_Asian
 6. Female_Black
 7. Female_Brown
 8. Female_White



Google Vision API Demo

Results I

- Microsoft did not have a 'doctor' or 'physician' label, this API was removed from analysis
- APIs only returned top n labels, not all labels of all classes, so missing data was an issue. Our solution: throwing out examples that did not contain a label from one or more APIs:
 - 10 male examples removed
 - 19 female examples removed
 - *** notable ***10 South Asian examples removed
- Analyzed results in R for statistical significance

Results II

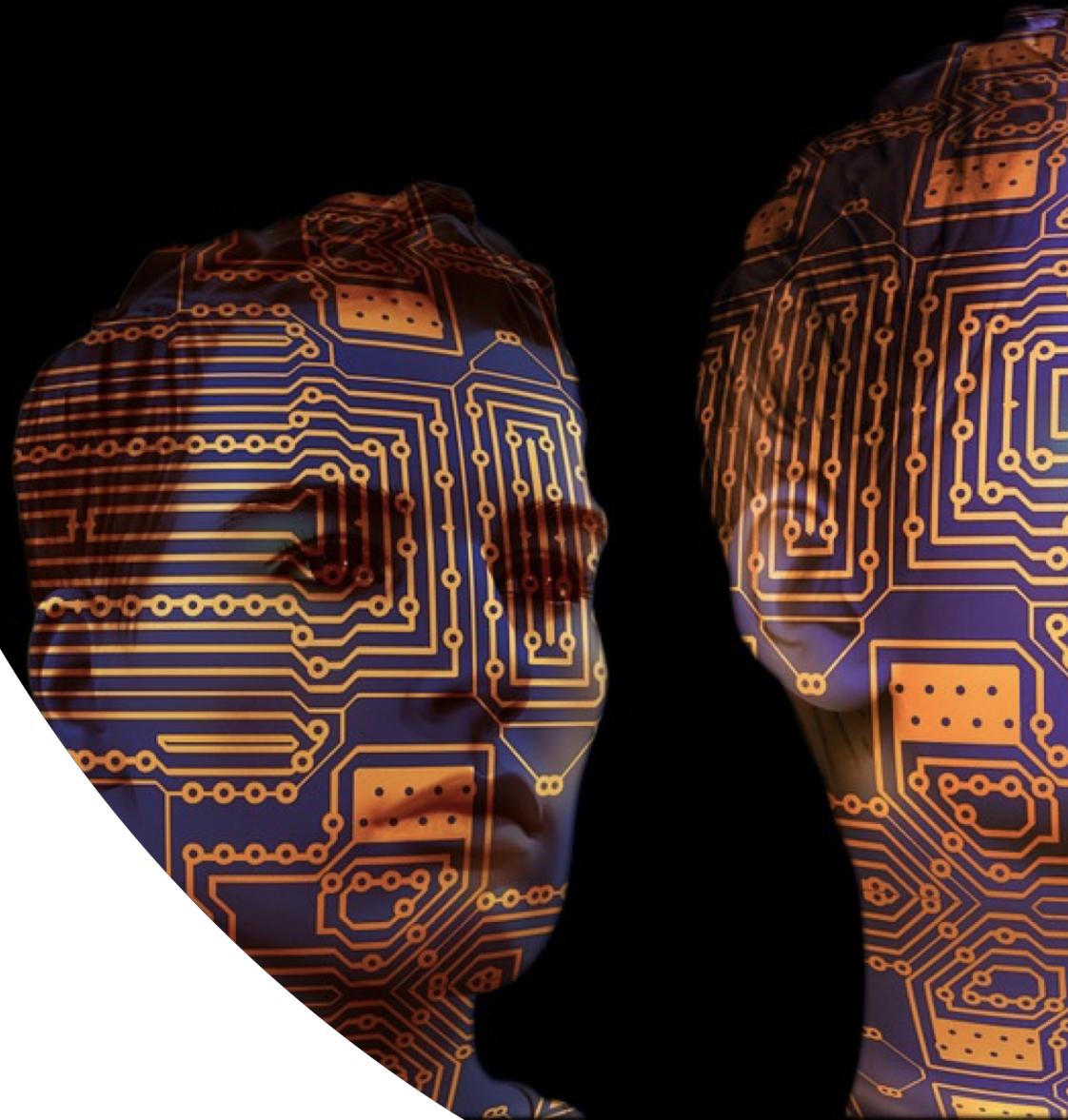
- **Google Vision API**
 - Men significantly higher than women
 - Black significantly lower than other three groups
 - Black Female lowest predicted class
- Amazon Rekognition API
 - Males significantly higher
 - No bias at race level
 - White Female lowest predicted class
- Clarifai API
 - Females significantly higher
 - No bias at race level
 - Black Male lowest predicted class

Results III

- Main Takeaway: Google Vision API exhibited the most bias in predictions, Clarifai API the least

Why Should I Care?

- Computer vision is increasingly being relied on for image labelling
- If the underlying program for determining these labels is biased, as we demonstrated, then these biases can be reinforced
- While the application we chose is not considered high-risk, these tools are being applied to areas that have life-changing impacts



Why Should I Care?

- AI in and of itself is not objective
- Algorithms automate the biases in the data they are trained on
- Researchers and industries need to be critical of how robust these tools really are
- As we build these systems, we must begin with an ethical mindset and continuously evaluate at each dev stage