OUR MEETING WILL BEGIN SHORTLY

NAIC

WELCOME TO THE INNOVATION, CYBERSECURITY, AND TECHNOLOGY (H) COMMITTEE

August 10, 2022

IN-PERSON ATTENDEES

Wi-Fi Network: NAIC2022; Password (case sensitive): Summer2022

VIRTUAL ATTENDEES

- Audio will be muted upon entry
- If virtual attendees would like to speak, please use the "Raise Hand" feature and we will let the Chair know you'd like to speak
- Enter with video on or off (your choice)
- Use the “Chat” feature for questions, comments, or assistance
- If you have joined by phone, to mute and unmute your line, press *6
- For additional help, please contact NAIC Technical Support team at MeetingTechHelp@naic.org or call (866) 874-4905

NAIC

NATIONAL MEETING

SUMMER 2022

PORTLAND

Welcome and Call to Order

Commissioner Kathleen A. Birrane (MD)

Roll Call

Kathleen A Birrane, Chair	Maryland	Adrienne A. Harris	New York
Evan G. Daniels, Co-Vice Chair	Arizona	Jon Godfread	North Dakota
Dana Popish Severinghaus, Co-Vice Chair	Illinois	Judith L. French	Ohio
Karima M. Woods	District of Columbia	Elizabeth Kelleher Dwyer	Rhode Island
John F. King	Georgia	Carter Lawrence	Tennessee
Amy L. Beard	Indiana	Kevin Gaffney	Vermont
Chlora Lindley-Myers	Missouri	Mike Kreidler	Washington
Troy Downing	Montana		

NAIC Support Staff: Denise Matthews / Scott Morris

1. Consider Adoption of its Spring National Meeting Minutes (Attachment One)

—Commissioner Kathleen A. Birrane (MD)

2. Consider Adoption of a Request for NAIC Model Law Development from the Privacy Protections (H) Working Group (Attachment Two)

—Katie Johnson (VA)

3. Consider Adoption of its Working Group Reports (Attachment Three)

—Commissioner Kathleen A. Birrane (MD)

A. Big Data and Artificial Intelligence (H) Working Group
—Superintendent Elizabeth Kelleher Dwyer (RI)

B. Cybersecurity (H) Working Group
—Cynthia Amann (MO)

C. E-Commerce (H) Working Group
—Commissioner Troy Downing (MT)

D. Innovation in Technology and Regulation (H) Working Group
—Director Evan G. Daniels (AZ)

E. Privacy Protections (H) Working Group
—Katie Johnson (VA)

4. Receive an Update on Innovation, Cybersecurity, and Technology (H) Committee Projects

—Commissioner Kathleen A. Birrane (MD)

A. ICT-Hub Concepts and Progress
B. Collaboration Forum on Algorithmic Bias Program and Kansas City Fly-In

5. Discuss Any Other Matters Brought Before the Committee

—Commissioner Kathleen A. Birrane (MD)

6. Adjournment

—Commissioner Kathleen A. Birrane (MD)

Collaboration Forum on Algorithmic Bias Session

Open Session

Collaboration Forum on Algorithmic Bias Panel: Approaches Companies Are or Can Implement to Manage and Mitigate the Risk of Unintended Bias and Illegal Discrimination When Developing and Using Artificial Intelligence (AI)/Machine Learning (ML)

—Dale Hall (Society of Actuaries—SOA)

—Tulsee Doshi (Lemonade Insurance Company)

—Daniel Schwarcz (University of Minnesota)

Ethical and Responsible Use of Data and Predictive Models

R. Dale Hall, FSA, MAAA, CFA, CERA
Managing Director of Research
Society of Actuaries Research Institute

SOCIETY OF ACTUARIES®

SOA Certificate Programs

- Designed for regulators, actuaries, data scientists and others who work in the insurance profession.
- Full program released in 2021 with executive version being developed
- Certificate for those passing an assessment
- First program introduced was on Predictive Analytics, with next on Ethical and Responsible Use of Data and Predictive Models
- Focus is on an ethical framework

Purpose of a Data Ethics Framework

- Based on a set of principles, which are then applied in practice
- Framework provides a way to organize data processes and analyze a given situation
- Framework facilitates the ability to have checkpoints and ask important questions
- Different approaches to frameworks, but many have the same core intentions

Ethical and Responsible Use of Data and Predictive Models: Principles

Fairness

Safety

Transparency and Accountability

Applying the Framework

SOCIETY OF ACTUARIES®

Applying the Framework: Gathering Data and Developing Models

- Follow applicable regulations related to fair use and privacy
- Avoiding biases in data collection
  - Selection bias; Measurement bias; Feature selection and omitted variable bias
- Avoiding biases that may influence model development
- Understand how various modeling approaches relate to the framework

Applying the Framework: Model Fairness

- Need to avoid disparate impact and disparate treatment
- Numerous measures that can be considered
  - Unawareness
  - Demographic parity
  - Positive predictive value
  - Individual fairness
  - Counterfactual fairness

Applying the Framework: Avoiding Unintended Discrimination

- Potential approaches
  - Pre-approved variables
  - Expand information used
  - Collect protected data to ensure model does not discriminate
  - Require demonstrated direct causal relationship

Applying the Framework: Model Safety and Governance

- Full understanding of the problem, the data, and the algorithm
- Model meets the intended purpose
- Avoid potential misuse or misinterpretation
- Ensure proper business operations through model development, formal control, implementation, and deployment
- Incorporate as part of the risk management and audit function
- Accountability by identifying who owns the processes, what the processes are, and why decisions were made

Applying the Framework: Model Explanation

- Important for transparency
- Models can be difficult to explain
- Technical approaches include
  - Feature importance measures
  - Partial dependence plots
  - Global surrogate models
  - Local surrogate models (e.g., LIME, Shapley values)
- BUT just as important: strong and simple verbal explanation

Feedback and Control Considerations

- Implementation
  - Feature selection and evaluation to continue after deployment
  - Ensure future projects that use model are evaluated against framework at each step
- Actuarial Control Cycle: use review and feedback to improve future model performance

SOCIETY OF ACTUARIES®

@soactuaries

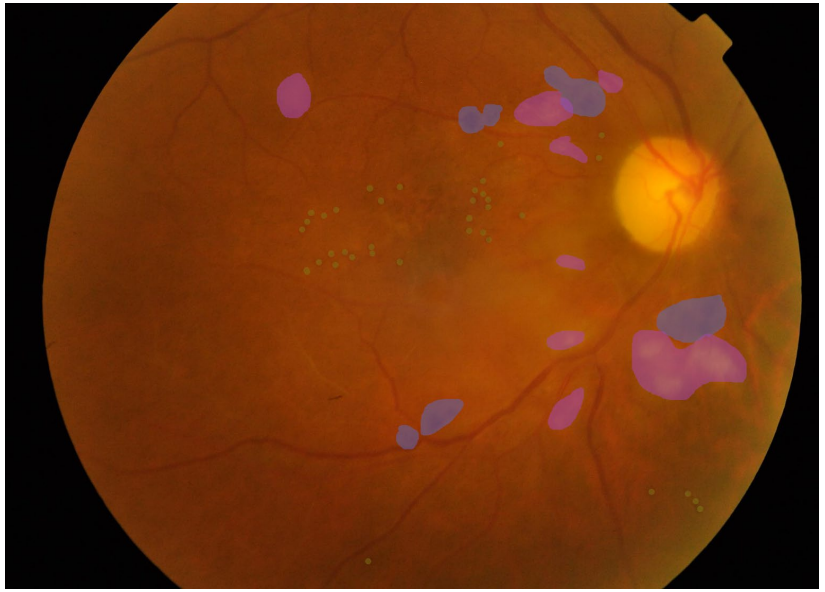
www.soa.org

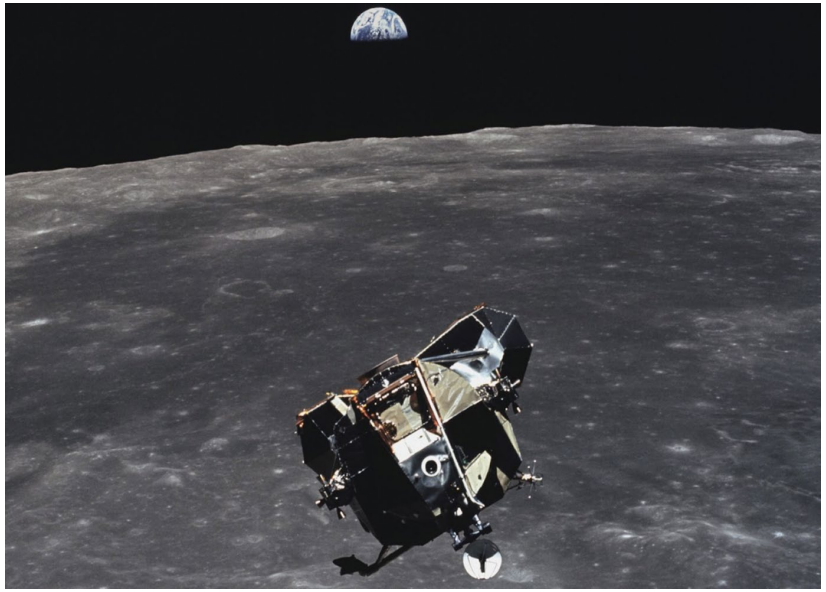
Today’s Approaches to Algorithmic Bias

Tulsee Doshi

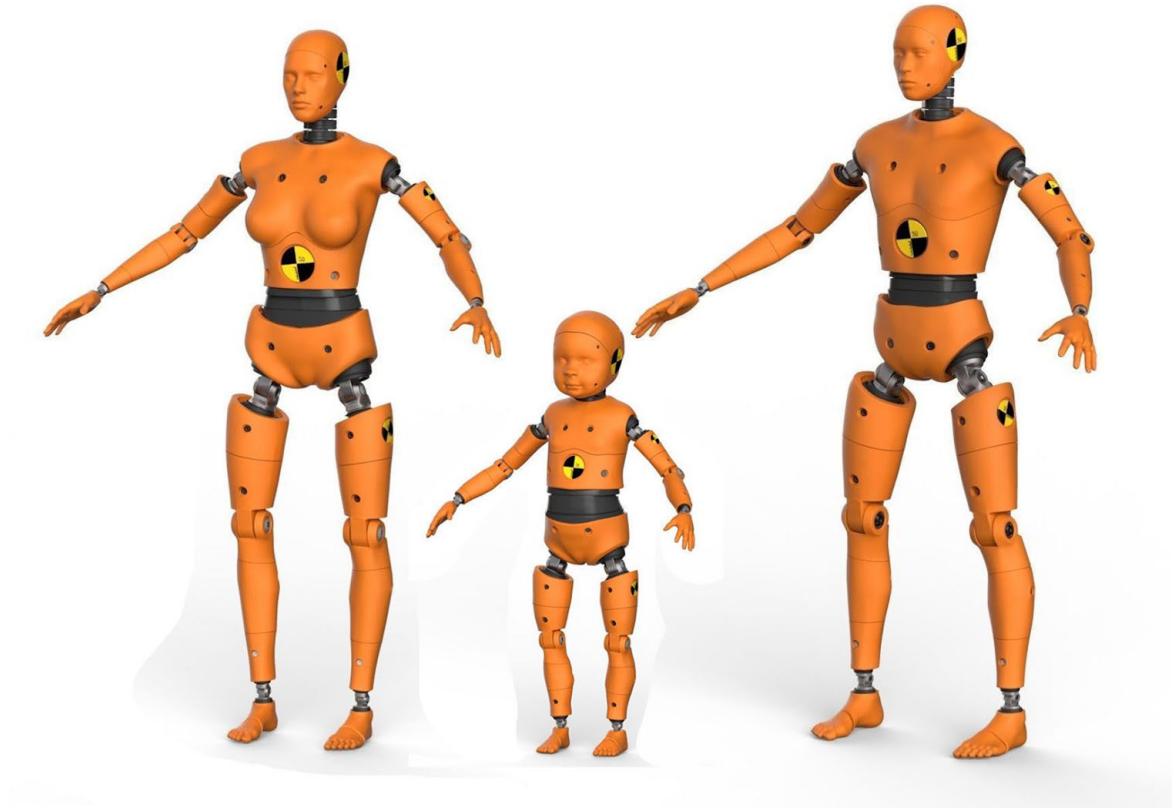
Lemonade Insurance Company

AI is being used around the world





Humans have a history of making product design decisions that are not always in line with the needs of everyone.



Female drivers were

47%

more likely to be severely injured in an auto accident

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3222446/

Source: Vulnerability of Female Drivers Involved in Motor Vehicle Crashes: An Analysis of US Population at Risk, 2015

1950s

1990s

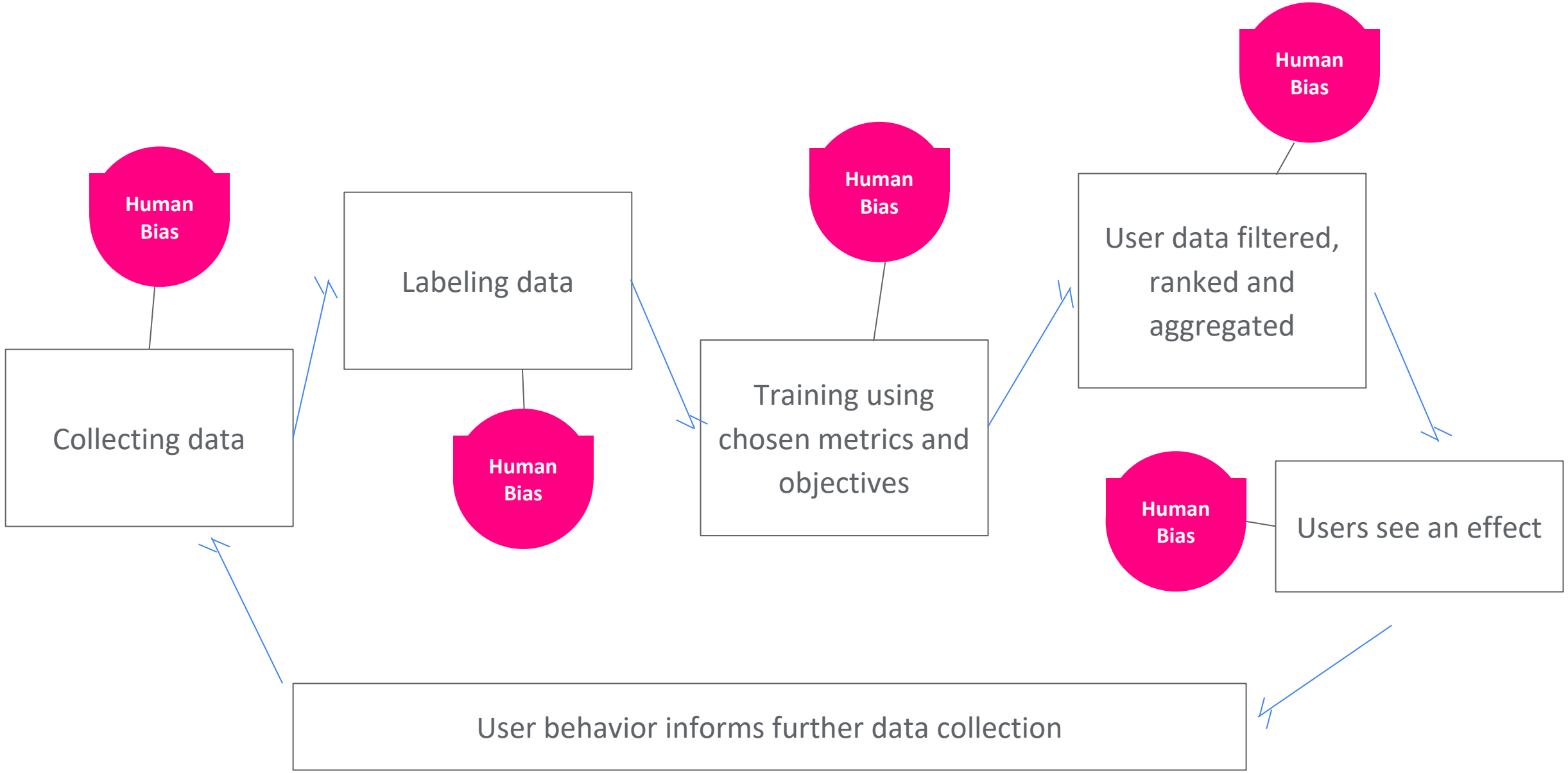
https://www.npr.org/2014/11/13/363517842/for-decades-kodak-s-shirley-cards-set-photography-s-skin-tone-standard

Humans are at the center of developing AI, and so the same biases we build into technology <u>today</u> elsewhere can exist in AI

These are not AI <u>only</u> concerns

People with the same risk should be priced the same

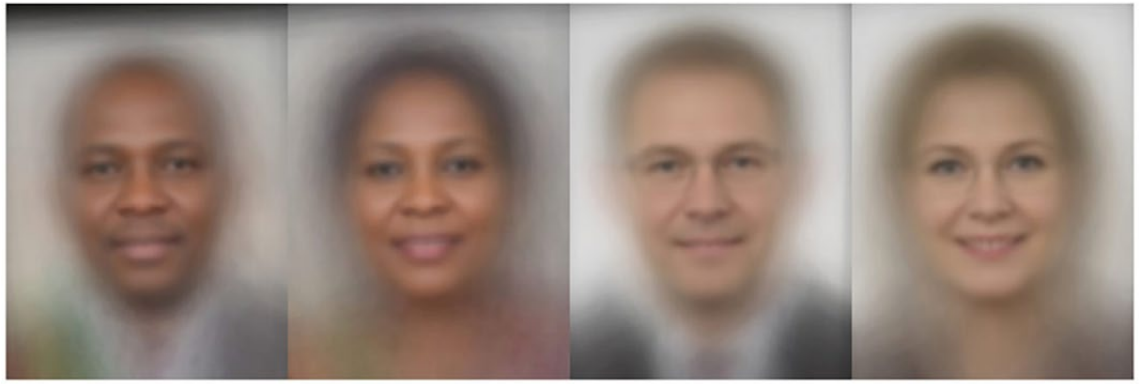
We should be pricing everyone the same and working to reduce the risk of higher risk areas / communities


Human Bias
Collecting data
Labeling data
Human Bias
Human Bias
Training using chosen metrics and objectives
Human Bias
User data filtered, ranked and aggregated
Human Bias
Users see an effect
User behavior informs further data collection

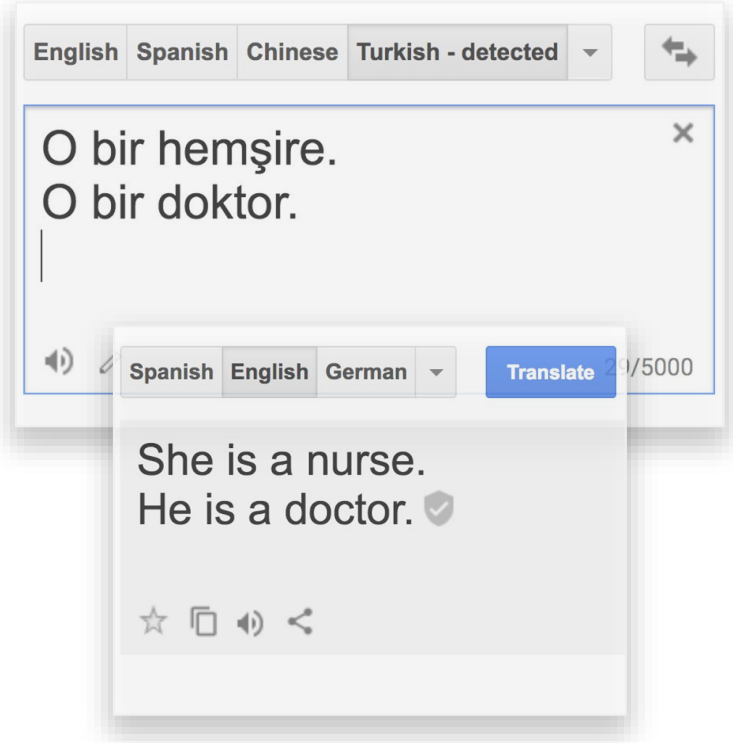
Because ML uses and users are so diverse, ML Fairness concerns can take many different forms

Different Models

Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
94.0%	79.2%	100%	98.3%	20.8%
99.3%	65.5%	99.2%	94.0%	33.8%
88.0%	65.3%	99.7%	92.9%	34.4%

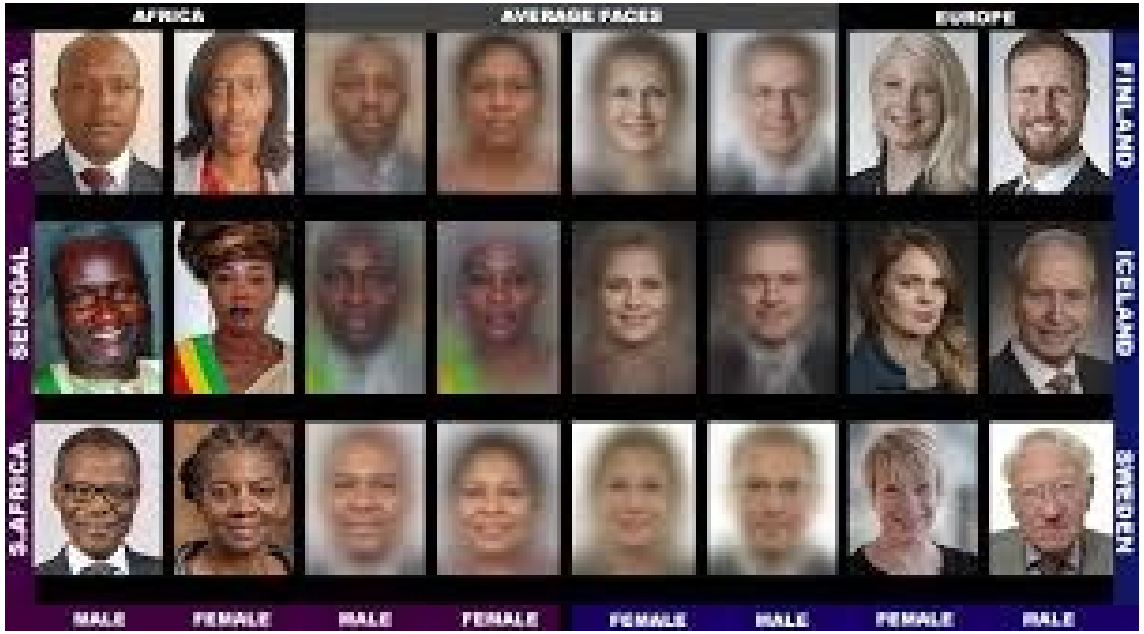


Gender Shades: Binary Gender Classification

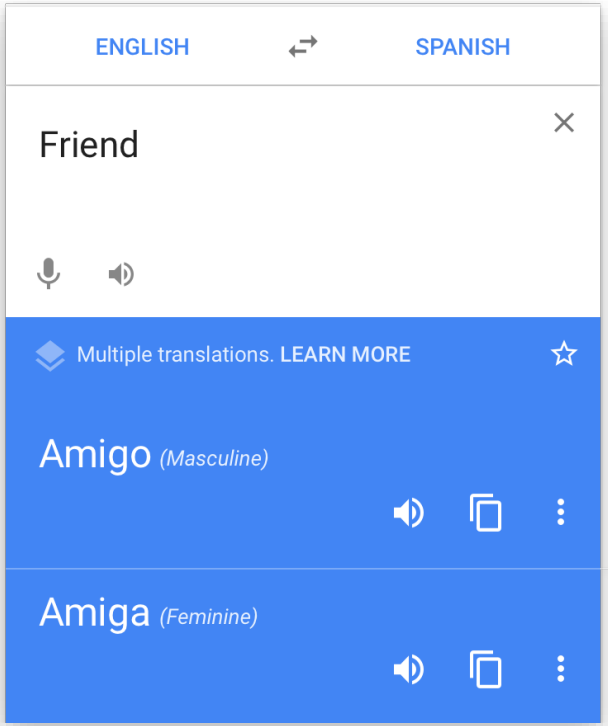


Stereotypes in Translation

And the solutions may look different as well



Data Augmentation and Collection



Product Design Approaches

How to approach fairness in model development?

Establish the problem **with diverse users in mind**

Collect data across your user base

Define fairness opportunities, and test, test, test

Design mitigation approaches and monitoring

Develop and grow an inclusive workforce, and bring in diverse perspectives

3 Stages of Evaluation

Model Features

1 Understand what features may be strong proxies, and how they affect model performance

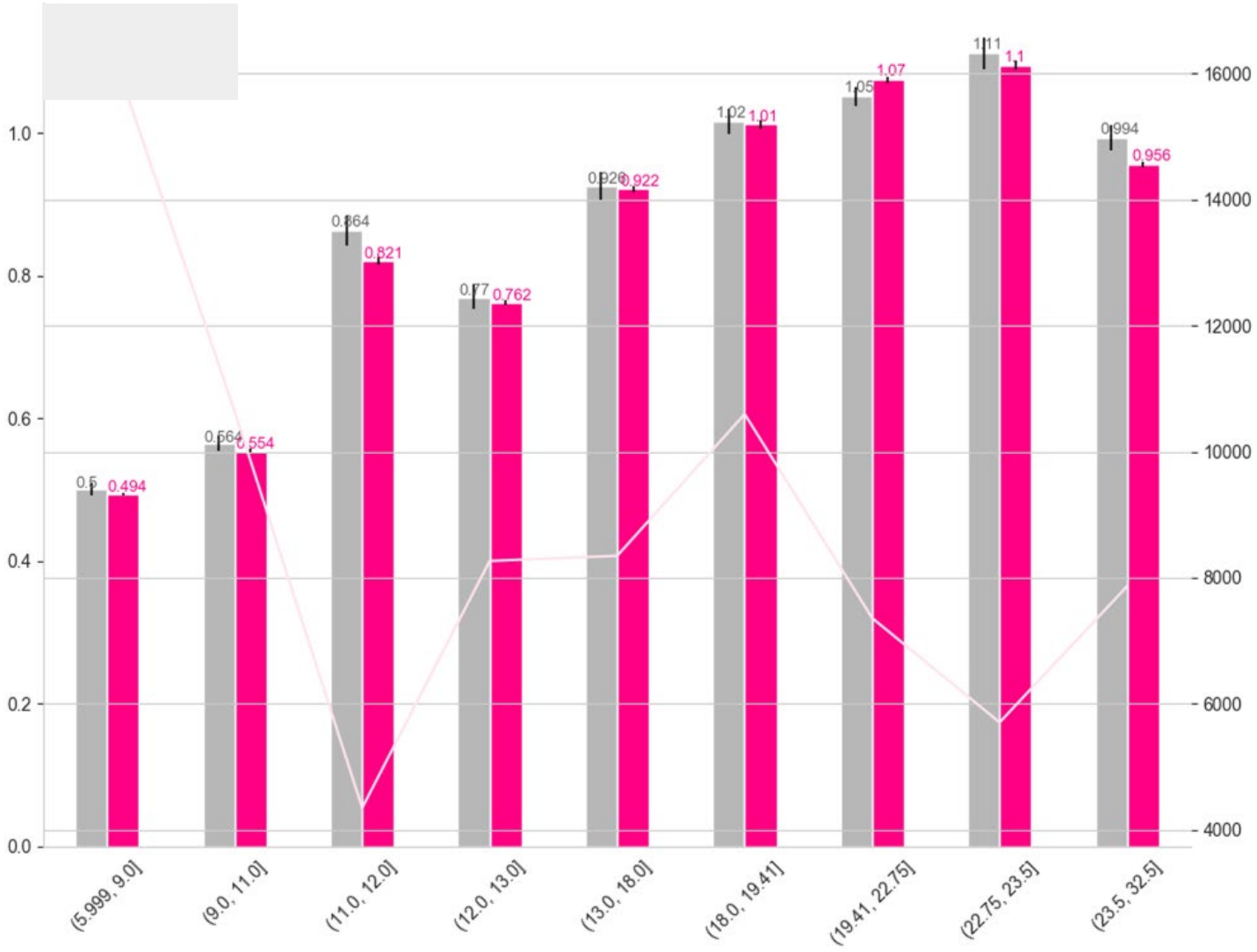
Model Outputs

2 Evaluate the performance of the model compared to the model's ground truth (eg. how does the model's predicted claim frequency differ from what is in the data?)

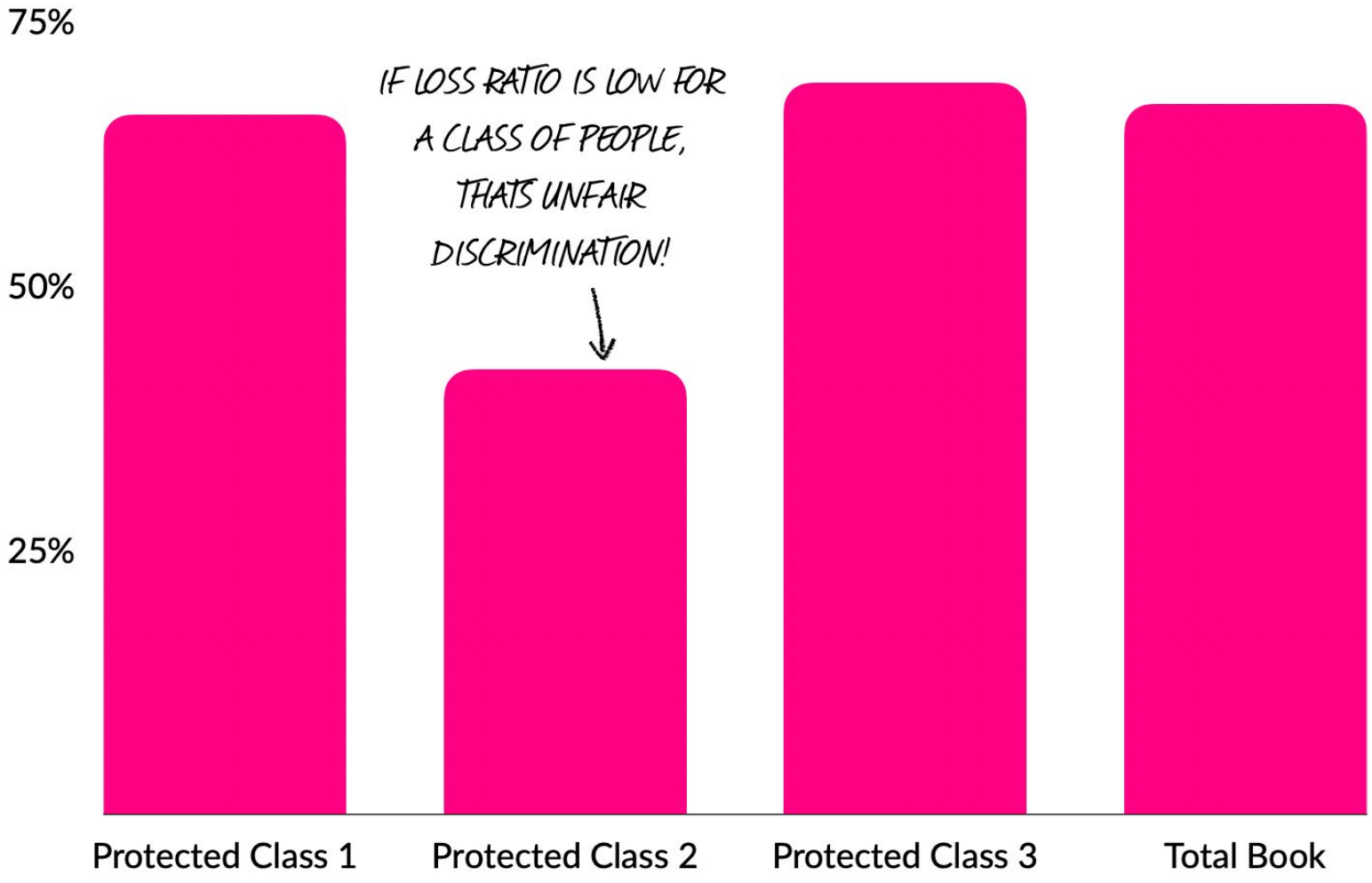
Long Term

3 Track overtime how loss ratios differ across communities to understand whether are systematic gaps across models, products, etc.

Evaluate Model Outputs



Long Term: Uniform Loss Ratio



Model Cards: Nutrition Labels for Models

Face Detection

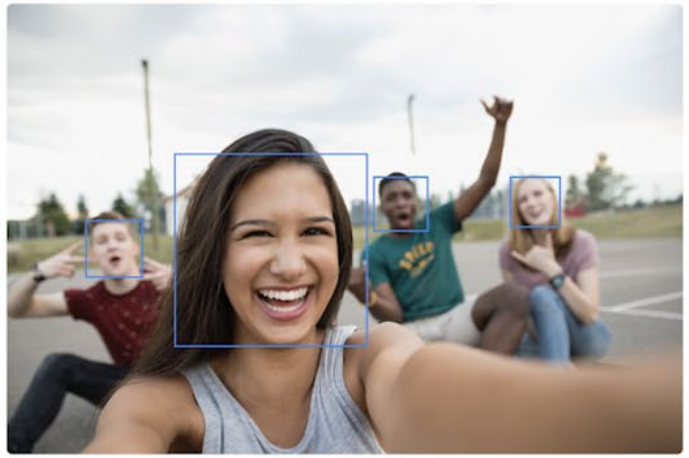
Version 0 Google Cloud Vision API

Face Detection

The model analyzed in this card detects one or more faces within an image or a video frame, and returns a box around each face along with the location of the faces' major landmarks. The model's goal is exclusively to identify the existence and location of faces in an image. It does not attempt to discover identities or demographics.

On this page, you can learn more about how well the model performs on images with different characteristics, including face demographics, and what kinds of images you should expect the model to perform well or poorly on.

MODEL DESCRIPTION

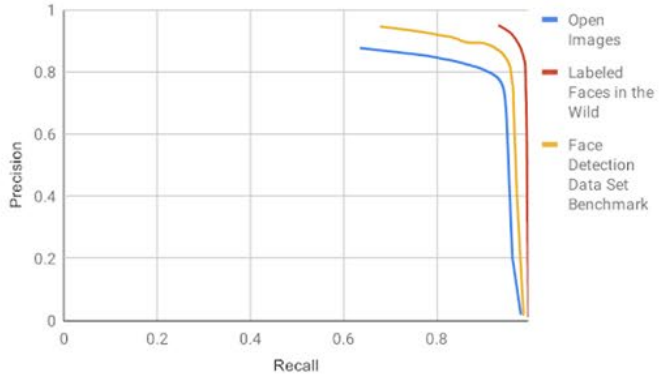


Input: Photo(s) or video(s)

Output: For each face detected in a photo or video, the model outputs:

- Bounding box coordinates
- Facial landmarks (up to 34 per face)

PEFORMANCE



Overall model performance, and performance sliced by different image and face characteristics, were assessed, including:

- Derived characteristics (face size, orientation, and occlusion)
- Face demographics (human-perceived gender presentation, age, and

Challenges

- **Protected Class Attributes are needed to be able to test:** Industry currently either individually collecting data, proxying, or leveraging other attributes (eg. zip-codes with majority black populations)
- **Hard to slice model performance with sparse data:** Attributes like losses don't happen frequently, and so especially for smaller companies, having enough data to meaningfully slice and see patterns can be hard
- **Incentives can be challenging if finding a bias makes companies liable:** Solving for algorithmic bias isn't always a quick fix. If a company evaluates for algorithmic bias and finds a challenge, does that make them immediately vulnerable to legal action while they seek to improve the problem?

NAIC

NATIONAL MEETING

SUMMER 2022

PORTLAND

Daniel Schwarcz

Fredrikson & Byron Professor of Law
University of Minnesota Law School

Schwarcz@umn.edu

Overview

- (1) Understanding the Risk of Biased AI in General
- (2) Understanding the Risk of Biased AI in Insurance
- (3) Limiting the Risk of Biased AI Requires Initial (not final) step of Testing AIs for Disparate Impact
- (4) If Testing of AIs shows Disparate Impact, then Unfair Bias MAY be present depending on Explanation for Unequal Results Across Protected Groups

(1) Understanding the Risk of Biased AI in General

- Machine-learning AIs are programmed to maximize “target variable” by inductively developing algorithms based on massive amounts of historic training data.
- Machine-learning AIs can create risk of bias against statutorily protected groups even though they do not have direct data that includes membership in protected groups. This can occur when:
  - Training data is itself biased
  - Membership in protected group is directly predictive of target variable
  - Target variable is a poor proxy for actual measure of interest
- Bias of AI models against protected groups is well documented outside of insurance setting:
  - Compas algorithm disproportionately over-predicted recidivism for blacks and underpredicted it for whites
  - Amazon hiring algorithm discriminated against women
  - Optum health algorithm directed resources disproportionately to white individuals

(2) Understanding the Risk of Biased AI in Insurance

- In underwriting/rating, AIs will inevitably discriminate based on proxies for protected status when protected status is actuarially predictive of claims. (Proxy discrimination)
  - Example: Because sex is predictive of auto claims, AIs will select proxies for sex in training data (i.e. social media, occupation) even if state law prohibits such discrimination and training data does not include policyholder sex.
  - Anya Prince & Daniel Schwarcz & , *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257 (2020).
- In fraud detection, any bias contained in the training data regarding when policyholder fraud is suspected or can be proven will be reproduced by the AI.
  - Example: If an insurers' fraud team has historically subconsciously more closely scrutinized claims from predominantly African American region, then AI will do the same.
- In marketing, AIs that are trained based on target variable of profit expectations across all lines of business will drive marketing to relatively wealthy areas where there is more cross-selling potential and total property value to insure.

(3) Limiting the Risk of Biased AI Requires Initial (not final) step of Testing AIs for Disparate Impact

- In all well-known cases of model bias, problem was discovered after auditing the relevant models for potential bias by examining model outcomes with respect to protected groups.
- Subjecting AIs to Risk-Management tools other than testing for disparate impact can only minimally limit risk of bias because of surprising ways bias can manifest.
  - Designers can try to carefully select target variables and limit bias in training data, but there are huge limitations to these efforts.
- Insurers currently do not typically have systematic data about policyholders' membership in protected classes.
  - Various techniques are available for proxying for membership in protected class, such as BISG
  - Groups like ORCAA already provide capacity to test algorithms in this way.
- Ultimately, insurers should be explicitly permitted and even required to collect information about policyholders' membership in protected groups.
  - Already required by HMDA for decades!

(4) If Testing of AIs shows Disparate Impact, then Unfair Bias MAY be present depending on Explanation for Unequal Results Across Protected Groups

- Standing alone, disparate impact may not represent a legal or regulatory problem, depending on the state and line of coverage.
- However, existence of disparate impact requires asking several hard questions (depending on the context) to determine presence of unfair bias:
  - In rating/underwriting context, crucial to ask:
    - Does disparate impact exist just for premiums, or also for loss ratios?
    - Does disparate impact reflect differences across protected classes that are unrelated to individuals' membership in a protected class?
  - In fraud context, key to ask:
    - Does training data include unfounded biases?
    - Even if there are differences across protected groups, is rate of false-positives (falsely-flagged fraud) higher among protected groups than unprotected groups due to AI?
  - In marketing context, must ask:
    - Do individual firms' marketing techniques result in segmented markets, wherein protected groups are served by smaller subset of insurers that have relatively favorable loss ratios?

NAIC

NATIONAL MEETING

SUMMER 2022

PORTLAND

Questions