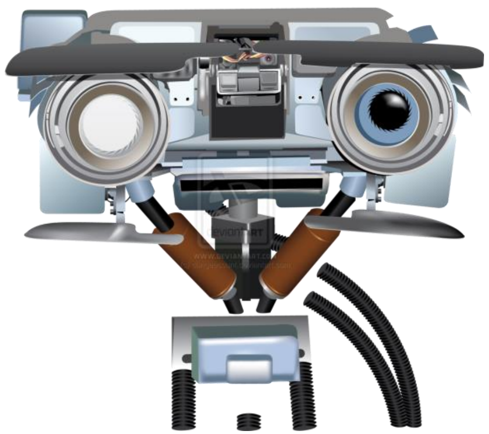


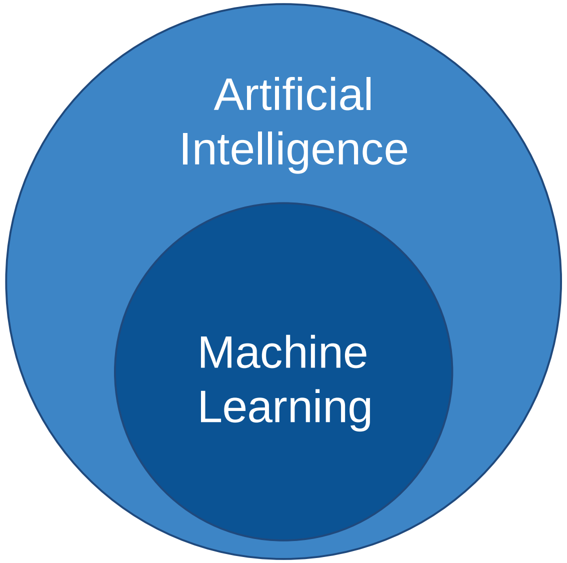
Artificial Intelligence

CS 440

Fall’20

Today: Outline

- Audio Check with Students in Class
- Piazza Tour
  - Q&A
  - Resources
    - Lecture 1
    - Linear Algebra Review
- Pre-lecture Material
- Syllabus
- Ethics in AI



Recap

Welcome to CS542!

- In this course we will:
  - Have fun!
  - Learn about artificial intelligence concepts, techniques, and algorithms.
  - Real world applications and state-of-the-art.
  - Outline:
    - Course Logistics
    - Today we will start with part of the recipe of every AI algorithm
    - **TODO:** ***Please fill out the timing zone info poll I will post on Piazza***

Recap

Course Staff

Sarah Adel Bargal

sbargal@bu.edu

Instructor

Kaihong Wang

kaiwkh@bu.edu

Teaching Fellow

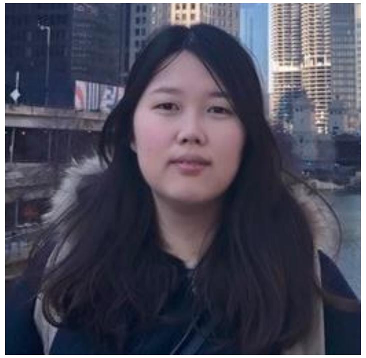
Recap

Course Staff Cont'd

Graders:

Fred Fung

Siddarth Mysore



Ximeng Sun

Mona Jalal

Class Moderator (for lab sessions): Michelle Lee

Recap

Please let us know ASAP if:

- You are not yet added to the Piazza course.
  - Please use Piazza for all course related communication
  - Please remember that: with great power comes great responsibility
- You cannot make it to *any* of the posted office hours.
- There is a lecture you cannot attend -> for Exam scheduling.

Recap

Course Pre-requisites

- CASCS112

  Introduction to Computer Science 2

- CASCS132

  Geometric Algorithms

- Calculus exposure: partial derivatives

Course Grading

Recap

- Grading Criteria

  5% Piazza, Lecture, and Lab Participation
  5% Pre-lecture materials
  30% Assignments
  10% Class challenge
  25% Exam 1
  25% Exam 2

- Late assignments will be levied a late penalty of 0.5% per hour (up to 72 hours). After 72 hours, no credit will be given.

- All course participants must adhere to the BU Academic Conduct Code: http://www.bu.edu/academics/resources/academic-conduct-code/
  All instances of academic misconduct must be reported to the College Academic Conduct Committee.

Recap

Examinations

- **Administering the exam:**
  - During lecture time
  - Open: Video camera + Microphone
  - Your hands must be in the camera’s field of view
  - Open exam pdf posted on Piazza
  - Take photos of your solutions on paper
  - Submit a pdf of the photos on a google form, just like you submit assignments
  - Confirm we received your submission before you leave (through private chat)
- **What do you need?**
  - Internet + Pen/pencil + Empty sheets of paper (~10)
  - New question, new page
  - Cell phone to take photos of your solutions at the end for submission

If you have problems supplying a stable internet connection please ask the dean of students for help NOW

Recap

Rotations for In-Person Attendance

- Lecture Rotations:
  - Attend on **Tuesday ONLY** if your last name starts with **A-K**
  - Attend on **Thursday ONLY** if your last name starts with **L-Z**

- Rotations might change as the semester progresses. Piazza announcement will be made if such changes are necessary.
  - If unexpected number of students appear, we will have to make more rotations.
  - If COVID cases increase on campus, we will have to make more rotations

- IMPORTANT: If at any time you change your status from Remote -> In-Person you need to email me *before* attending class.

Recap

Safety

- **COVID Compliance –** please remember to do your daily symptom surveys, schedule your COVID screening tests as required, always wear your mask, maintain 6+ feet of social distancing, and follow all signage especially maximum occupancies.
- On-campus and Off-campus Safety
- All Instructor and TF office hours will be held remotely.
- No face-to-face questions after class.
- Faculty will report to Associate Dean for Student Academic Life, Steven Jarvi, students who are not compliant in the classroom (e.g. a student is not properly wearing a mask).

Update from the Medical Advisory Group

BOSTON UNIVERSITY

Recap

Facial Coverings

Facial coverings and masks are an effective tool to prevent the potential spread of illnesses like COVID-19. In order to be effective, facial coverings and masks should have a snug fit and ideally consist of two or three layers of material (e.g. two-ply or three-ply). While fitted N-95 masks provide the best level of protection, these are in short supply and generally should be reserved for health care workers. Similar levels of protection can be achieved with surgical masks, three-ply masks, or two-ply masks consisting of a layer of polypropylene and cotton or two layers of cotton, and two layers of polypropylene.

Gaiters, bandanas, neck fleeces, scarves and masks with exhalation valves should not be used given potentially increased risk of aerosol transmission associated with these types of facial coverings.

Make sure you wash your hands before donning a mask, place it over your nose and mouth, make sure it fits snugly so air is passing through the mask and not around the sides of the mask, and make sure you can breathe comfortably.

Judy Platt MD, Director of Student Health Services
Ann Zaia PhD, CNP, SM, MSN, MHA Director of the Occupational Health Center
Carrie Landa PhD, Associate Director of Student Health Services
Davidson Hamer MD, Professor of Global Health and Medicine

Recap

Switching to Remote

- This could happen for several reasons:
  - University decides to do so due to high COVID spread
  - Any failure in classroom setup making it not compliant
  - Instructor is not cleared to be on-campus
  - …

Please keep an eye out for any Piazza notifications

Ethics in Artificial Intelligence

Kate Saenko

AI Fears

- Autonomous weapons – frameworks for regulation
- Future of work – deskilling / reskilling
- Worse Inequality – bias in algorithms may worsen inequality; ecological concerns in energy, storage and cooling required for ML; economic inequality
- Divided societies with algorithmic bubbles – automated recommendations, news feeds; deepfakes and election meddling

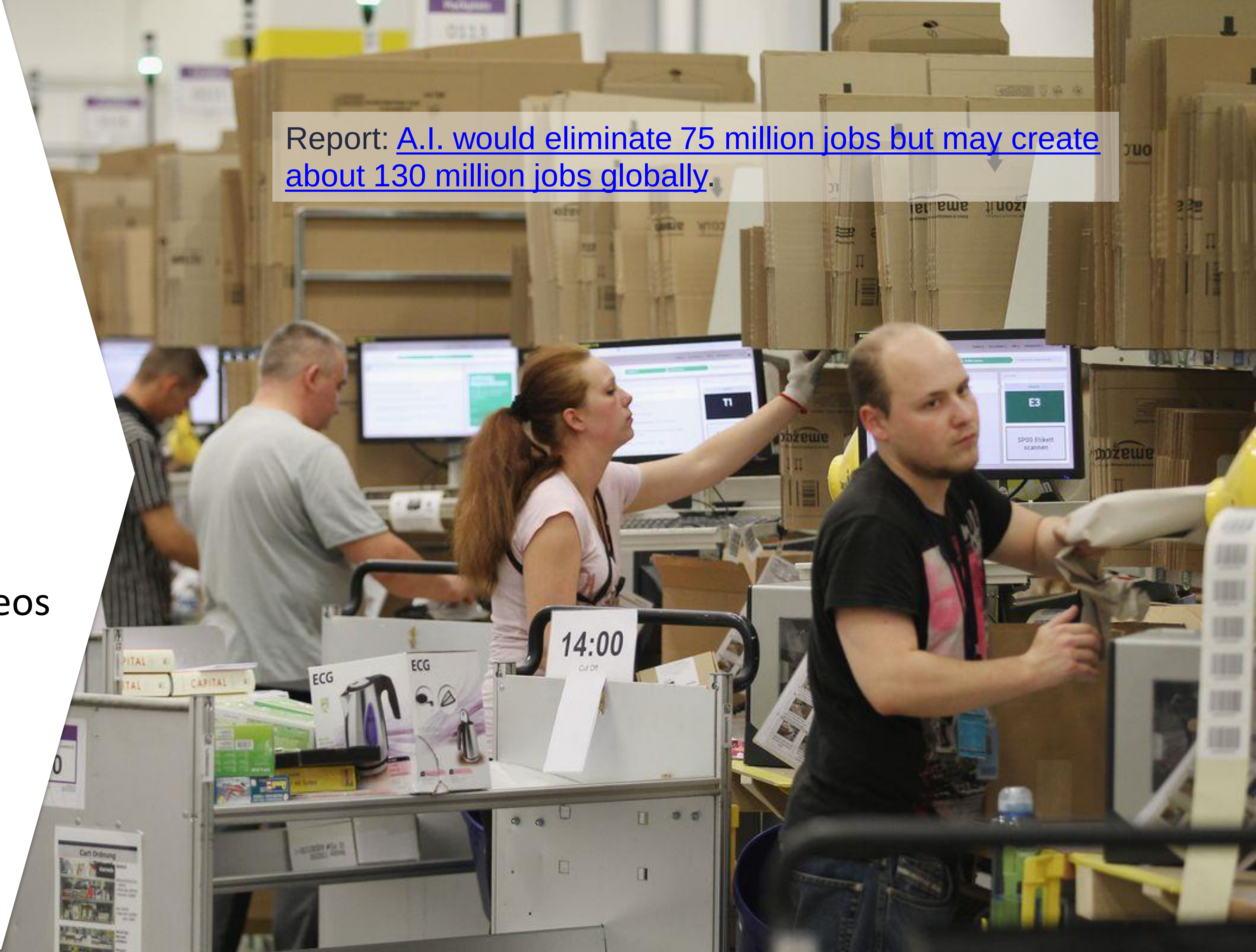
Many of these problems are not new!

Fears about job automation, lack of privacy and inequality arise with each new innovation

- Printing press
- Weapons
- Internet

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance



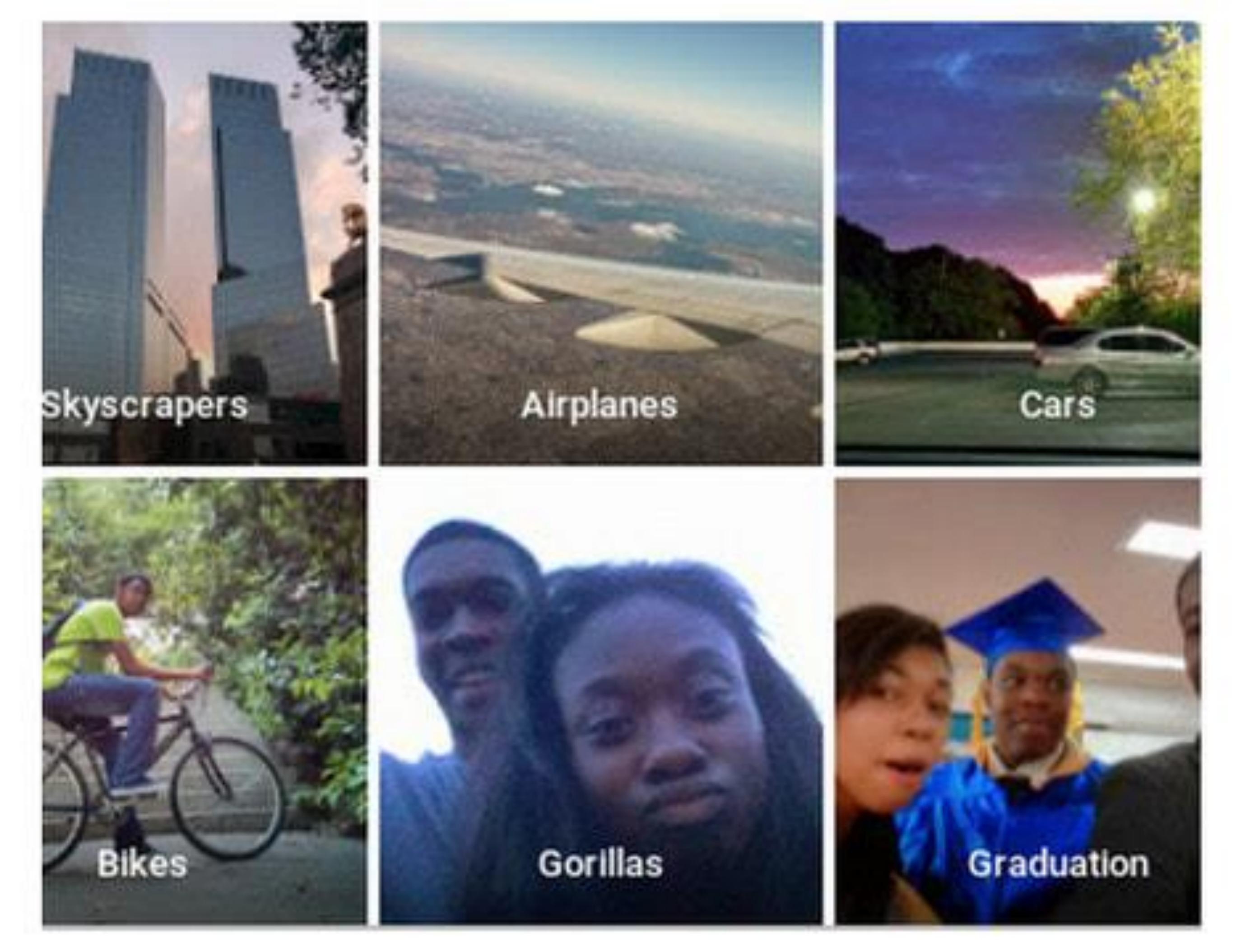
Report: A.I. would eliminate 75 million jobs but may create about 130 million jobs globally.

Stable Roles	New Roles	Redundant Roles
Managing Directors and Chief Executives	Data Analysts and Scientists*	Data Entry Clerks
General and Operations Managers*	AI and Machine Learning Specialists	Accounting, Bookkeeping and Payroll Clerks
Software and Applications Developers and Analysts*	General and Operations Managers*	Administrative and Executive Secretaries
Data Analysts and Scientists*	Big Data Specialists	Assembly and Factory Workers
Sales and Marketing Professionals*	Digital Transformation Specialists	Client Information and Customer Service Workers*
Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products	Sales and Marketing Professionals*	Business Services and Administration Managers
Human Resources Specialists	New Technology Specialists	Accountants and Auditors
Financial and Investment Advisers	Organizational Development Specialists*	Material-Recording and Stock-Keeping Clerks
Database and Network Professionals	Software and Applications Developers and Analysts*	General and Operations Managers*
Supply Chain and Logistics Specialists	Information Technology Services	Postal Service Clerks
Risk Management Specialists	Process Automation Specialists	Financial Analysts
Information Security Analysts*	Innovation Professionals	Cashiers and Ticket Clerks
Management and Organization Analysts	Information Security Analysts*	Mechanics and Machinery Repairers
Electrotechnology Engineers	Ecommerce and Social Media Specialists	Telemarketers
Organizational Development Specialists*	User Experience and Human-Machine Interaction Designers	Electronics and Telecommunications Installers and Repairers
Chemical Processing Plant Operators	Training and Development Specialists	Bank Tellers and Related Clerks
University and Higher Education Teachers	Robotics Specialists and Engineers	Car, Van and Motorcycle Drivers
Compliance Officers	People and Culture Specialists	Sales and Purchasing Agents and Brokers
Energy and Petroleum Engineers	Client Information and Customer Service Workers*	Door-To-Door Sales Workers, News and Street Vendors, and Related Workers
Robotics Specialists and Engineers	Service and Solutions Designers	Statistical, Finance and Insurance Clerks
		Lawyers

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

Bias can lead to offensive or unfair results…



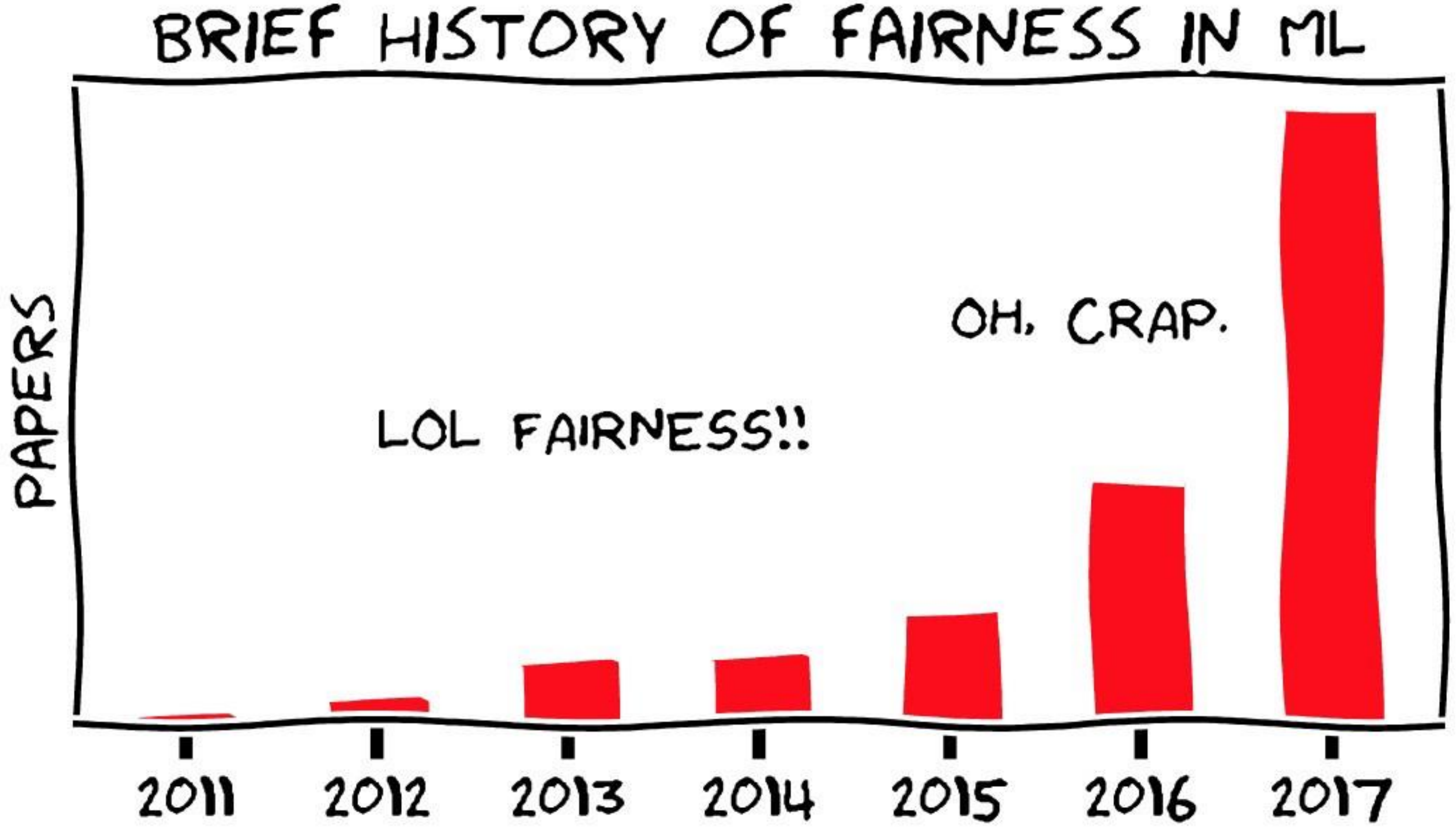
Gender Shades (Buolamwini & Gebru, 2018)

- Evaluated commercial gender classifiers from Microsoft, FACE++, IBM
- Found large disparity in error between population subgroups based on gender, skin color

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." Conference on Fairness, Accountability and Transparency. 2018.

Fairness in Machine Learning

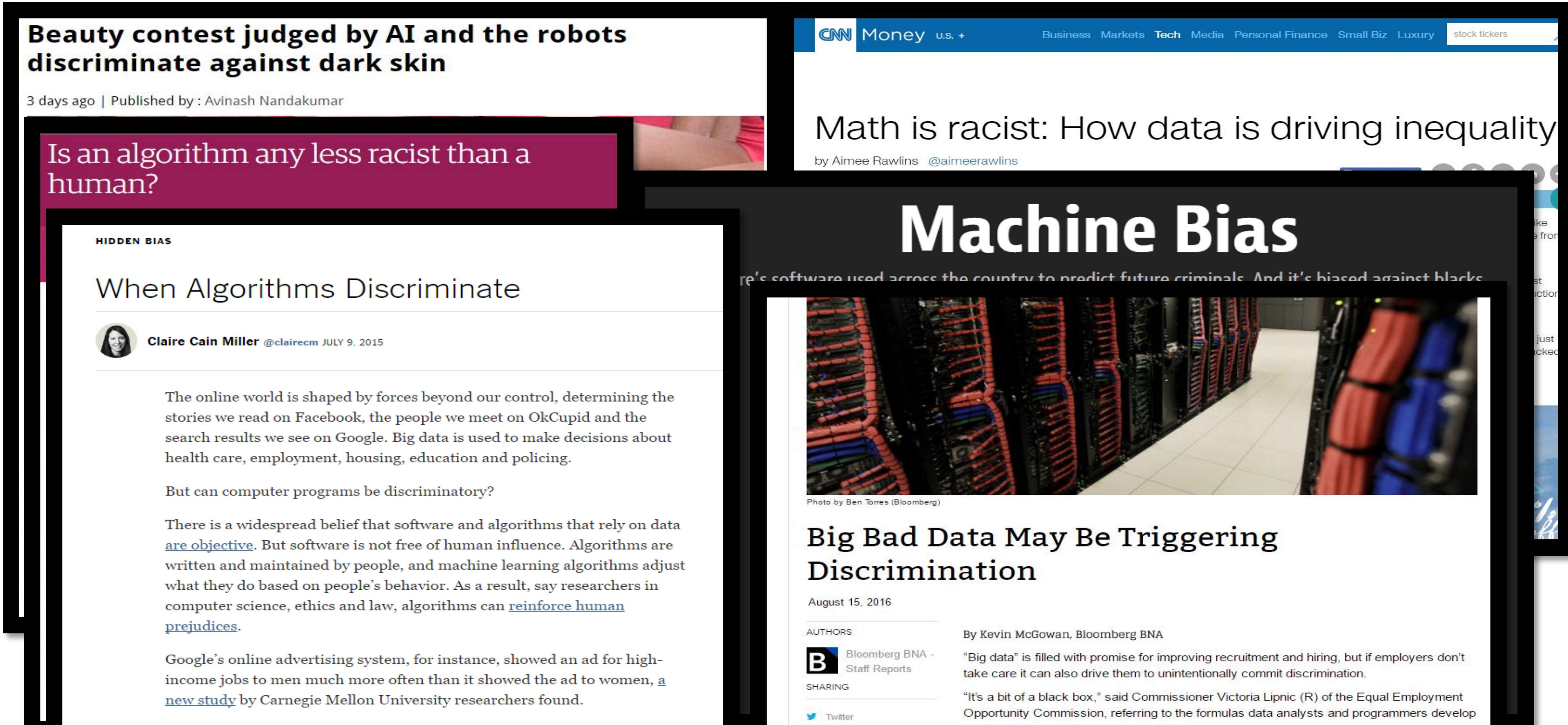


Moritz Hardt

Aaron Roth (University of Pennsylvania)

Initially: AI is better than humans!

Wait, maybe not such a good idea…



Beauty contest judged by AI and the robots discriminate against dark skin

3 days ago | Published by : Avinash Nandakumar

Is an algorithm any less racist than a human?

HIDDEN BIAS

When Algorithms Discriminate

Claire Cain Miller @clairecm JULY 9, 2015

The online world is shaped by forces beyond our control, determining the stories we read on Facebook, the people we meet on OkCupid and the search results we see on Google. Big data is used to make decisions about health care, employment, housing, education and policing.

But can computer programs be discriminatory?

There is a widespread belief that software and algorithms that rely on data are objective. But software is not free of human influence. Algorithms are written and maintained by people, and machine learning algorithms adjust what they do based on people's behavior. As a result, say researchers in computer science, ethics and law, algorithms can reinforce human prejudices.

Google's online advertising system, for instance, showed an ad for high-income jobs to men much more often than it showed the ad to women, a new study by Carnegie Mellon University researchers found.

CNN Money U.S. + Business Markets Tech Media Personal Finance Small Biz Luxury stock tickers

Math is racist: How data is driving inequality

by Aimee Rawlins @aimeerawlins

Machine Bias

…re's software used across the country to predict future criminals. And it's biased against blacks.

Photo by Ben Torres (Bloomberg)

Big Bad Data May Be Triggering Discrimination

August 15, 2016

AUTHORS

Bloomberg BNA - Staff Reports

SHARING

Twitter

By Kevin McGowan, Bloomberg BNA

"Big data" is filled with promise for improving recruitment and hiring, but if employers don't take care it can also drive them to unintentionally commit discrimination.

"It's a bit of a black box," said Commissioner Victoria Lipnic (R) of the Equal Employment Opportunity Commission, referring to the formulas data analysts and programmers develop

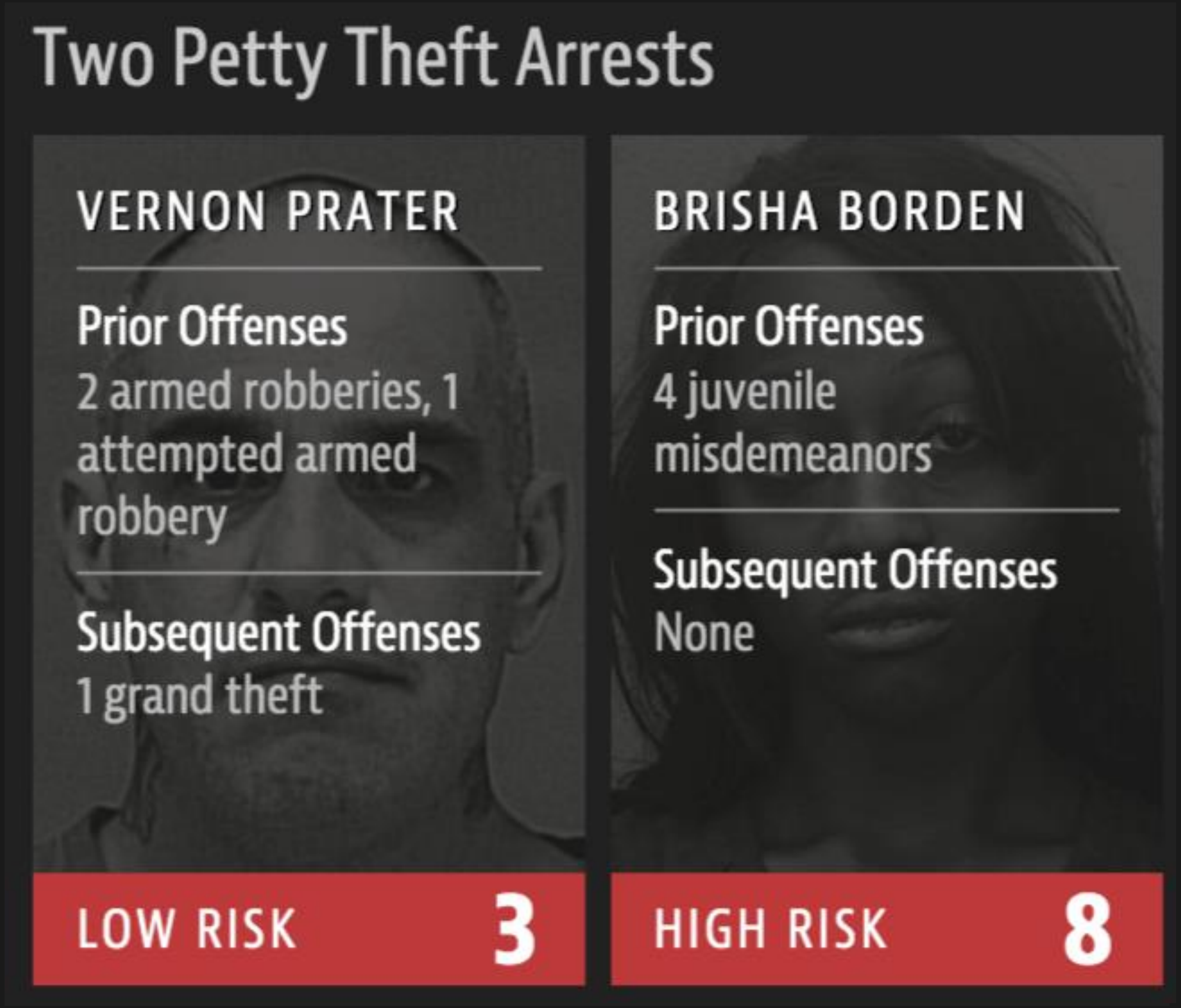
Example of ML (un)fairness: COMPAS

- Criminal justice: recidivism algorithms (COMPAS)
- Predicting if a defendant should receive bail
- Unbalanced false positive rates: more likely to wrongly deny a black person bail

ProPublica Analysis of COMPAS Algorithm

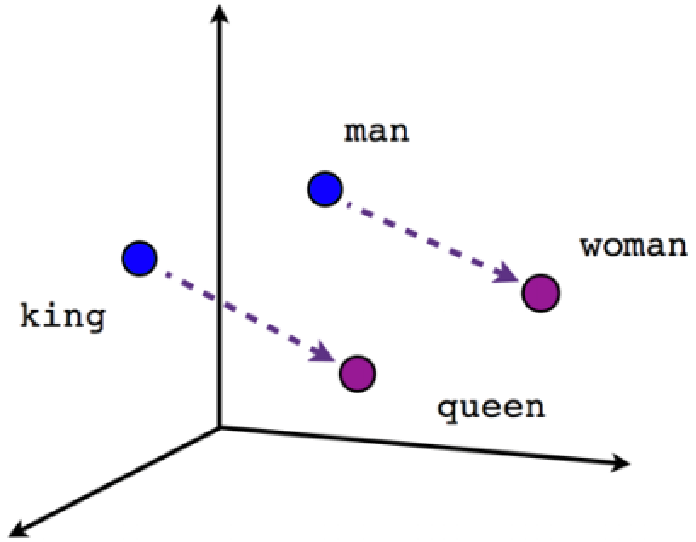
	White	Black
Wrongly Labeled High-Risk	23.5%	**44.9%**
Wrongly Labeled Low-Risk	**47.7%**	28.0%

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing


Two Petty Theft Arrests
VERNON PRATER
Prior Offenses
2 armed robberies, 1 attempted armed robbery
Subsequent Offenses
1 grand theft
LOW RISK
3
BRISHA BORDEN
Prior Offenses
4 juvenile misdemeanors
Subsequent Offenses
None
HIGH RISK
8

Example of ML (un)fairness: word embedding

- Bias found in word embeddings (Bolukbasi et al. 2016)
  - Examined word embeddings (word2vec) trained on Google News
  - Represent each word with high-dimensional vector
  - Vector arithmetic: found analogies like
  - Paris - France = London – England
  - man - woman = programmer – homemaker = surgeon - nurse
- The good news: word embeddings learn so well!
- The bad news: sometimes too well
- Our chatbots should be less biased than we are



Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings, NIPS 2016

(Un)Fairness Where?

- Data (input)
  - e.g. more arrests where there are more police
  - try to “correct” bias
- Models (output)
  - e.g. discriminatory treatment of subpopulations
  - equality of false positive/negative rates; calibration
- Algorithms (process)
  - e.g. don’t learn outcomes of denied mortgages
  - design (sequential) algorithms that are fair

Data Bias Examples

Reporting Bias example:
A sentiment-analysis model is trained to predict whether book reviews are positive or negative based on a corpus of user submissions to a popular website. The majority of reviews in the training data set reflect extreme opinions (reviewers who either loved or hated a book), because people were less likely to submit a review of a book if they did not respond to it strongly. As a result, the model is less able to correctly predict sentiment of reviews that use more subtle language to describe a book.

Selection Bias example:
A model is trained to predict future sales of a new product based on phone surveys conducted with a sample of consumers who bought the product. Consumers who instead opted to buy a competing product were not surveyed, and as a result, this group of people was not represented in the training data.

Why fairness is hard

- Suppose we are a bank trying to fairly decide who should get a loan i.e. Who is most likely to pay us back?

- Suppose we have two groups, A and B (the sensitive attribute) This is where discrimination could occur

- The simplest approach is to remove the sensitive attribute from the data, so that our classifier doesn’t know the sensitive attribute

Why fairness is hard

Table 2: To Loan or Not to Loan?

Age	Gender	Postal Code	Req Amt	A or B?	Pay
46	F	M5E	\$300	A	1
24	M	M4C	\$1000	B	1
33	M	M3H	\$250	A	1
34	F	M9C	\$2000	A	0
71	F	M3B	\$200	A	0
28	M	M5W	\$1500	B	0

Why fairness is hard

Table 3: To Loan or Not to Loan? (masked)

Age	Gender	Postal Code	Req Amt	A or B?	Pay
46	F	M5E	\$300	?	1
24	M	M4C	\$1000	?	1
33	M	M3H	\$250	?	1
34	F	M9C	\$2000	?	0
71	F	M3B	\$200	?	0
28	M	M5W	\$1500	?	0

Why fairness is hard

- However, if the sensitive attribute is correlated with the other attributes, this isn’t good enough

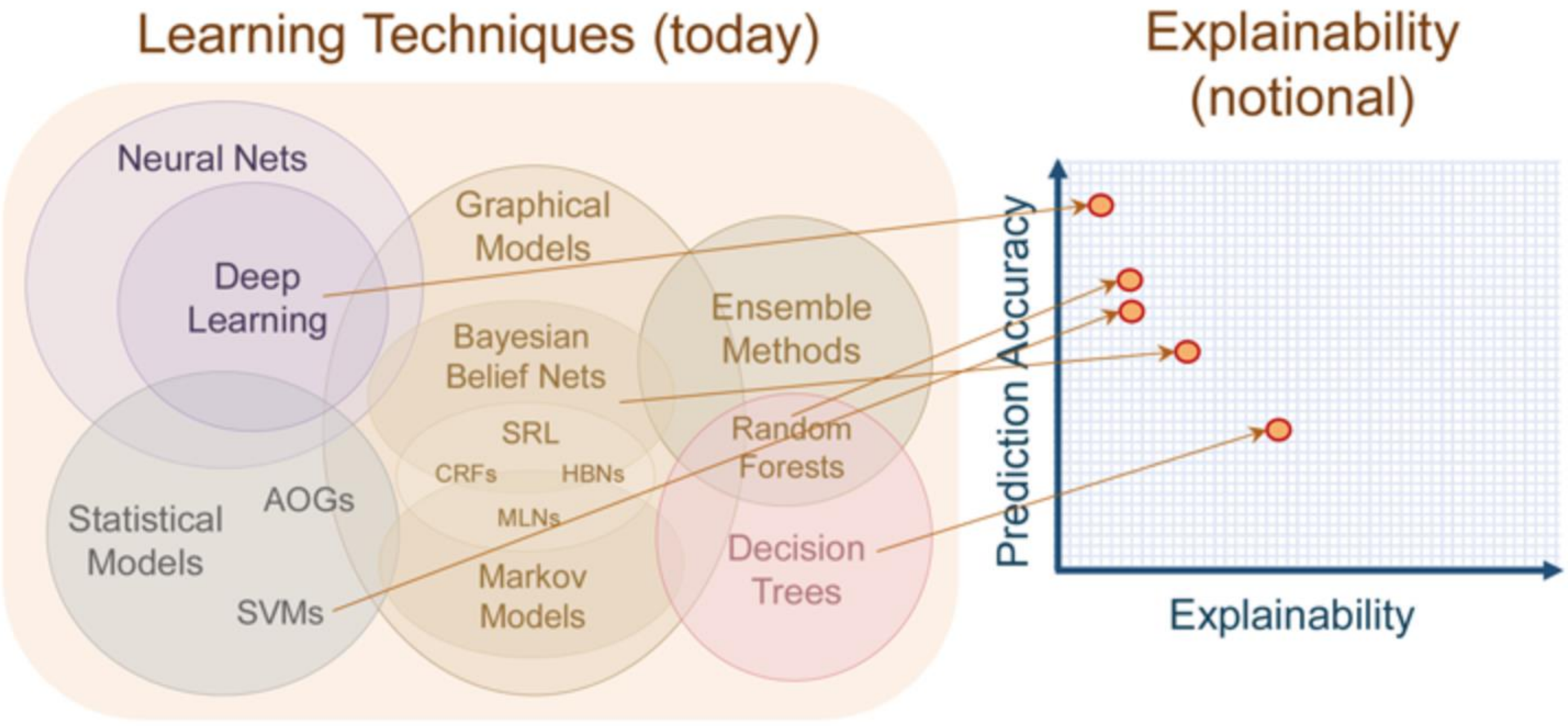
- It is easy to predict race if you have lots of other information (e.g. home address, spending patterns)

- More advanced approaches are necessary

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

Accuracy *vs.* explainability



E.g.: dataset bias leads to higher errors on ‘novel’ data… Can an explanation point to such bias?

Training

Most cows are black/brown

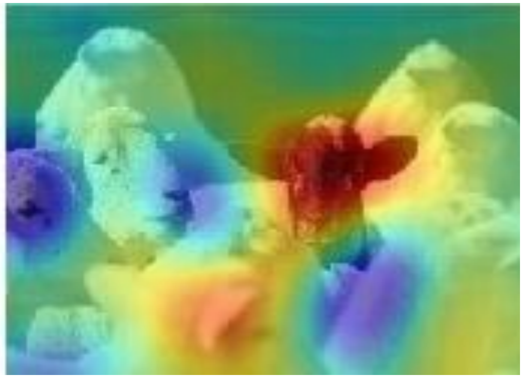
Most sheep are white

Test

Prediction: “cow” 76%

True class: “sheep”

Explanation

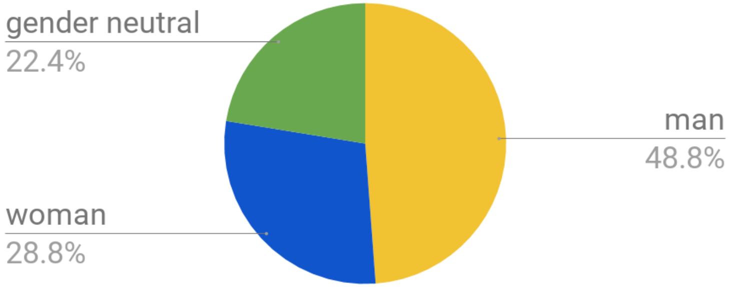


Gender bias in captioning models (Hendricks et al. 2018)

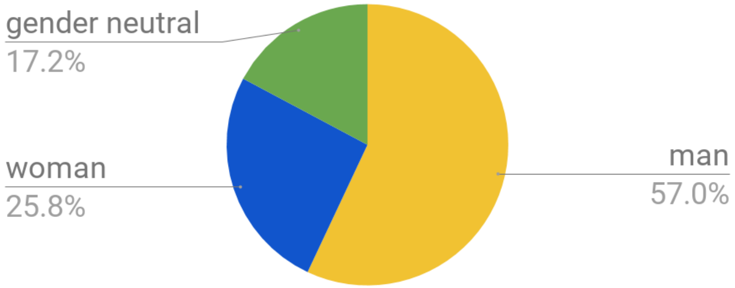
Evidence for "man"

Baseline: *A **man** sitting at a desk with a laptop computer.*

Ground truth captions



Generated captions



Hendricks et al. "Women Also Snowboard: Overcoming Bias in Captioning Models." ECCV 2018

Zhao et al. "Men also like shopping: Reducing gender bias amplification using corpus-level constraints." EMNLP 2017

Sample Misclassification

Ground Truth:
BabyCrawling

Classified as:
Pushups

- Explainablity would tell us “why”, or at least highlight pixels responsible for the prediction

Why is an algorithm predicting "Pedestrians Crossing the road" very well?

- Because of the periodic motion of the legs? If so, then we would have a problem in the following test scenario where the legs of the pedestrians are completely occluded.

Train

Test

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

If we start trusting algorithms to make decisions, who will have the final word on important decisions? Will it be humans, or algorithms?

Algorithms are already being used to [determine prison sentences]. Judges' decisions are [affected by their moods], so some argue that [judges should be replaced] with "robojudges". However, a [ProPublica study] found that one of these popular sentencing algorithms was highly biased against blacks.

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

https://www.youtube.com/watch?v=VWrhRBb-1Ig

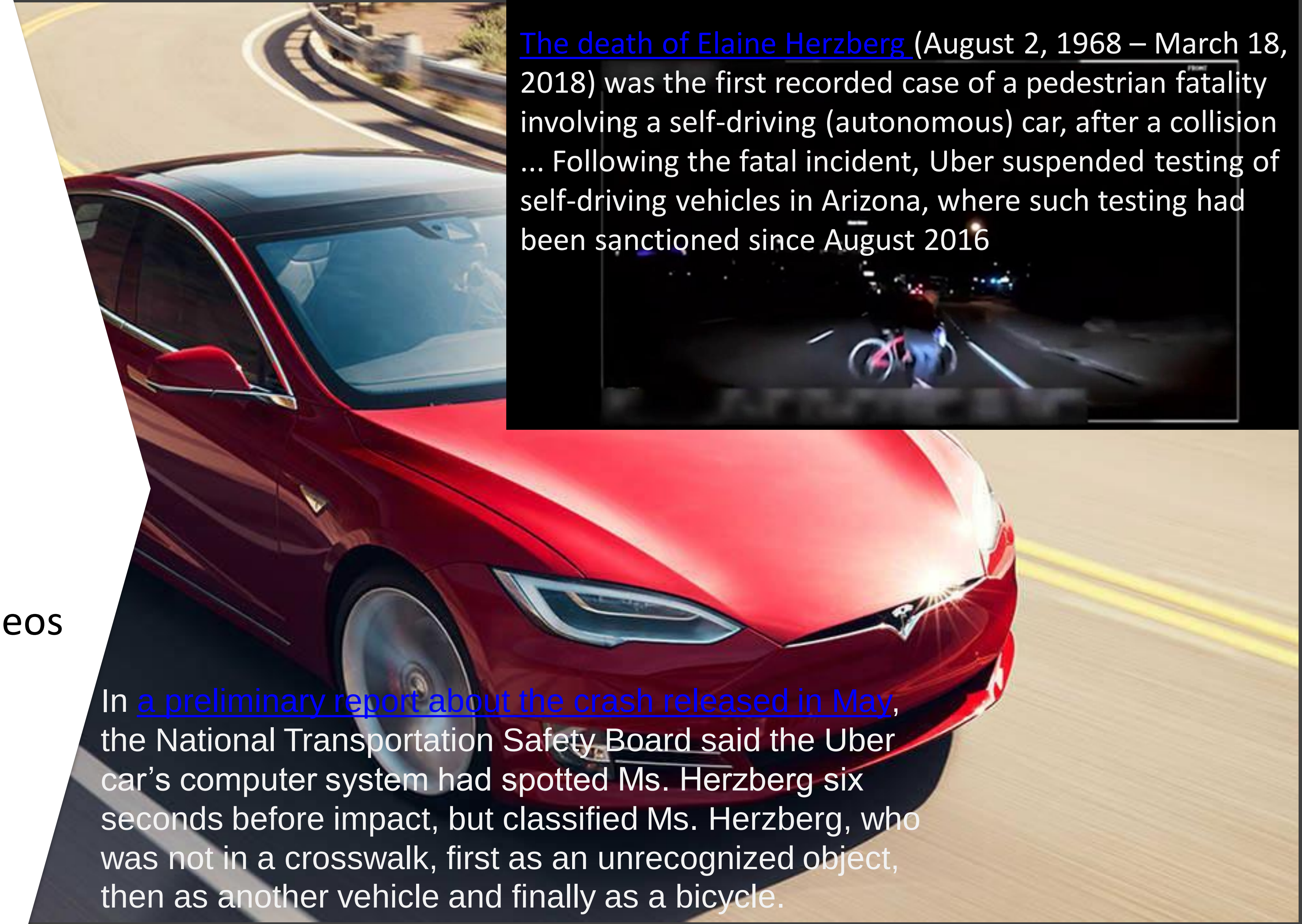
Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

https://www.albawaba.com/news/china-selling-autonomous-weaponized-drones-saudi-arabia-and-pakistan-1321951

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

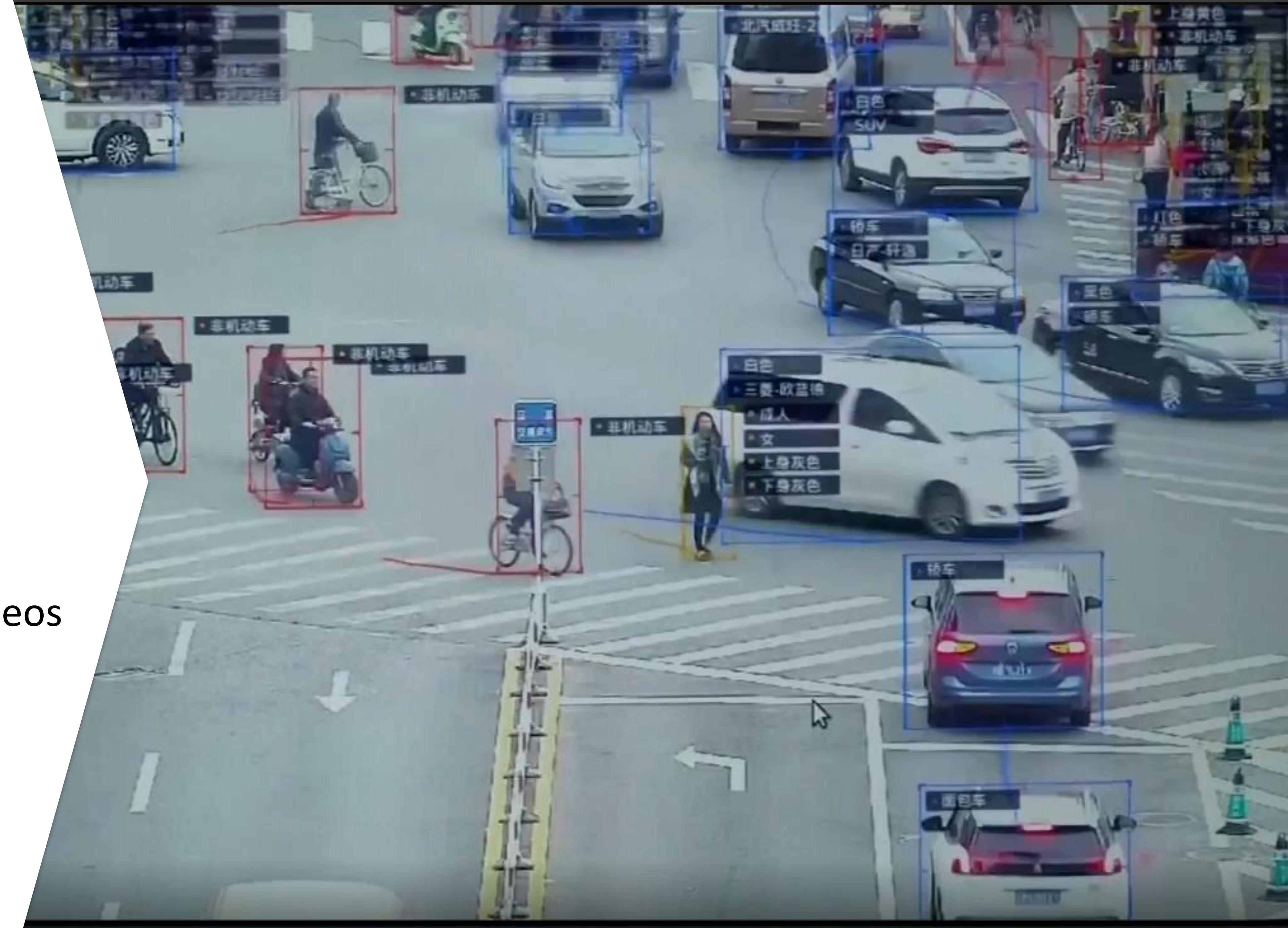


[The death of Elaine Herzberg](#) (August 2, 1968 – March 18, 2018) was the first recorded case of a pedestrian fatality involving a self-driving (autonomous) car, after a collision ... Following the fatal incident, Uber suspended testing of self-driving vehicles in Arizona, where such testing had been sanctioned since August 2016

In [a preliminary report about the crash released in May](#), the National Transportation Safety Board said the Uber car's computer system had spotted Ms. Herzberg six seconds before impact, but classified Ms. Herzberg, who was not in a crosswalk, first as an unrecognized object, then as another vehicle and finally as a bicycle.

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance



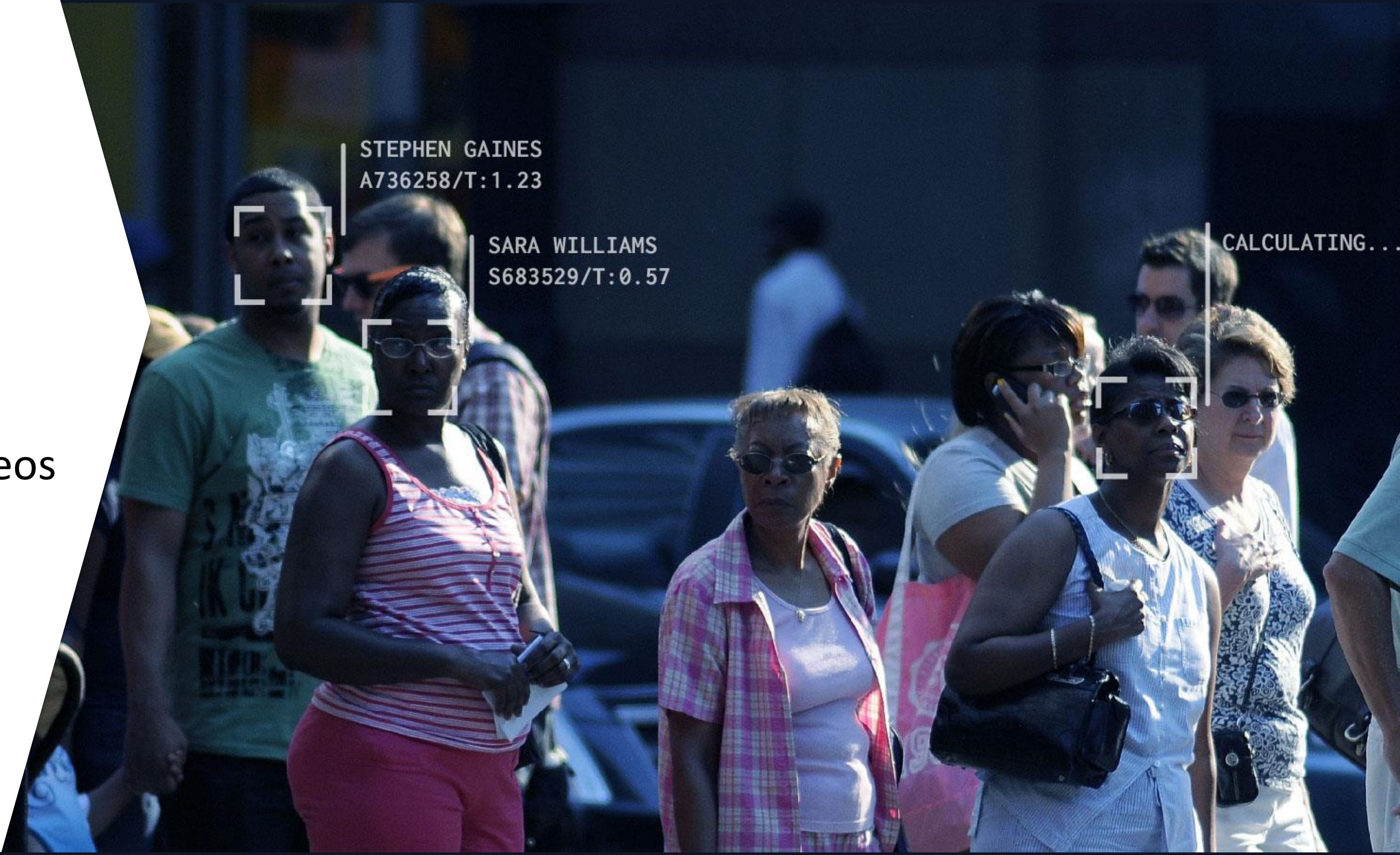
Surveillance cameras in China using machine vision

Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

One in two American adults is in a law enforcement face recognition network-- https://www.perpetuallineup.org/

Most law enforcement agencies do little to ensure their systems are accurate.



Ethical Issues in Machine Learning

- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

Further reading:

FATML Conference: https://www.fatml.org/
ACM FAT* Conference: https://fatconference.org/