Today: Outline

- **Free-choice Lectures**
- **Ethics in ML and DL**
- **Project Breakout Session**

- Announcement:
  - Practice problems available
  - Thu is a free-choice lecture
- Reminders:
  - Exam Jun 22 in class
    (and ~12 hrs before for remote only students)
  - *Team Registration Deadline due today*
  - *Regrade requests for PS1 due today*
  - *Be sure your scores are updated on webpage*
    http://cs-people.bu.edu/sbargal/cs523/grades.html

Deep Learning

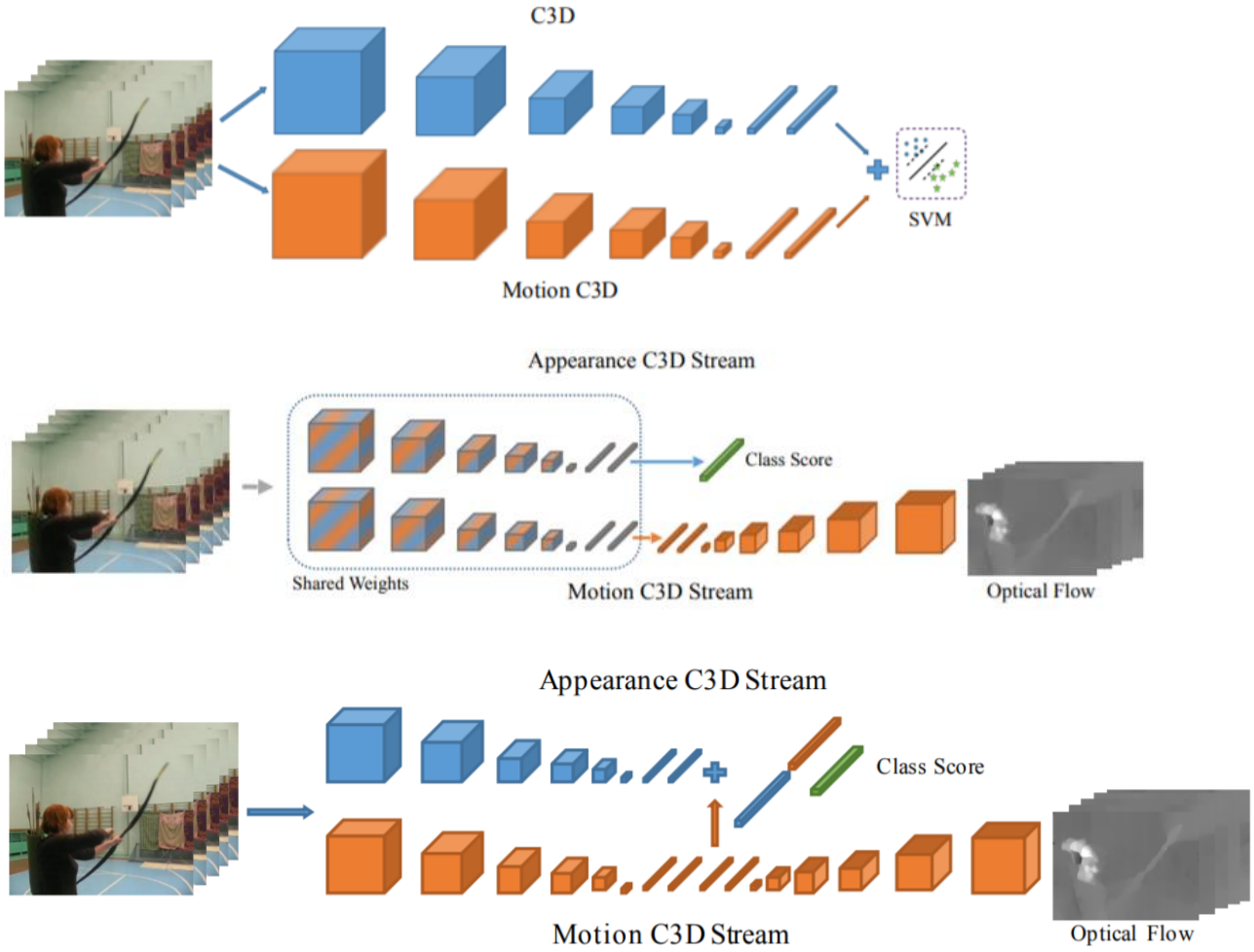
CS 523 Summer1 2021

You can use your random ID to access your row of grades for the course.

Grades**

Rand ID	PS1 (/100)	PS2 (/100)	PS3 (/100)	Par (/5)	Bonus (/2)	Pre-lec1 (/1)	Pre-lec2 (/1)	Pre-lec3 (/1)	Exam (/100)	Project (/100)
42	76	-	-	-	-	1	1	1	-	-
29	72.5	-	-	-	-	1	1	0	-	-
41	81	-	-	-	-	1	1	1	-	-
14	72.5	-	-	-	-	0.75	0	0	-	-
47	74.5	-	-	-	-	1	1	1	-	-
18	0	-	-	-	-	0	0	0	-	-
37	73.5	-	-	-	-	1	1	1	-	-
46	78	-	-	-	-	1	1	1	-	-
23	71.5	-	-	-	-	1	0.75	1	-	-
48	73.5	-	-	-	-	1	1	1	-	-
44	79	-	-	-	-	1	1	0.75	-	-
45	85	-	-	-	-	1	1	1	-	-

Multiple Modalities & Auxiliary Tasks



Pre-lecture Material

Grad-CAM

Which of the following is true?
[Multiple choices allowed]

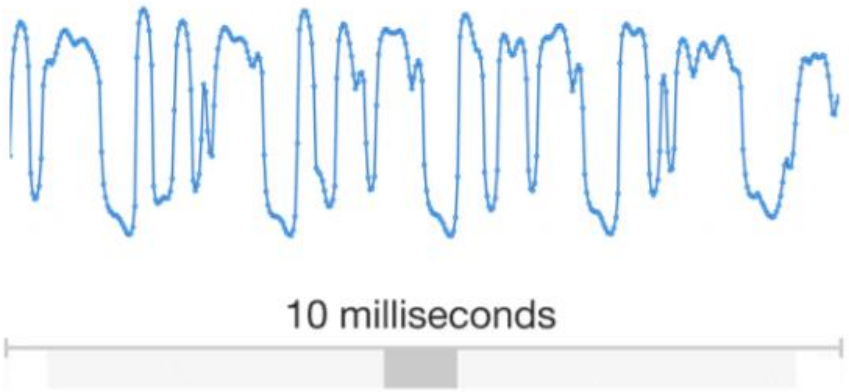
#pin

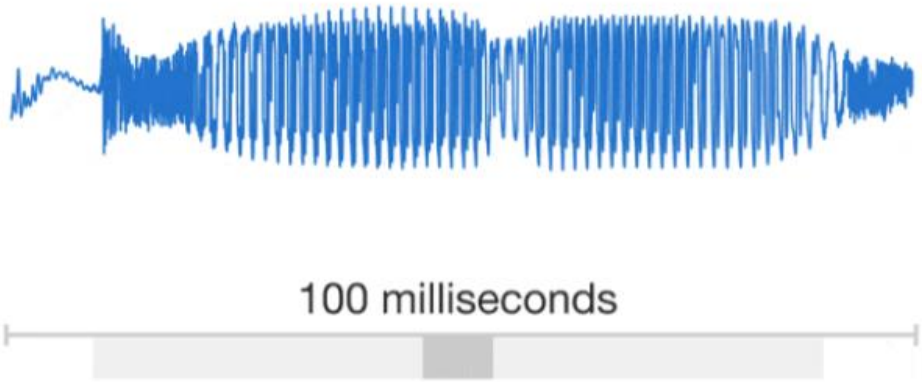
- [ ] Grad-CAM is a an interpretability method for deep learning models
- [ ] Grad-CAM provides transparency into why a model makes a specific prediction
- [ ] All applications presented in the paper are for visual tasks (takss that use images and/or videos)
- [ ] One can compute a Grad-CAM based saliency map without the knowledge of network parameters
- [ ] Grad-CAM can be used as a tool to diagnose model bias

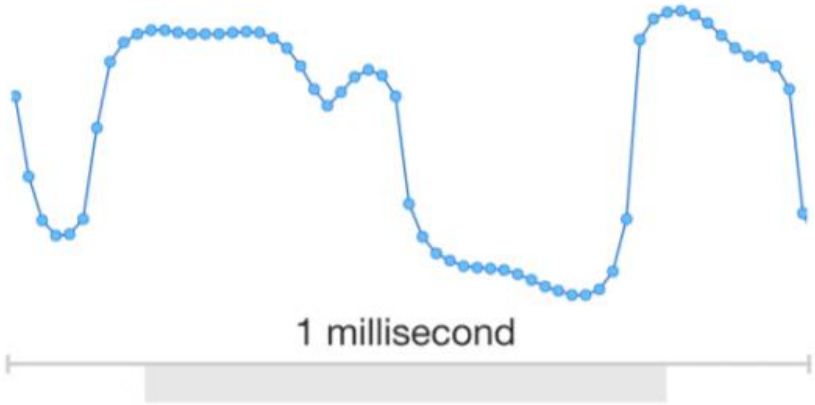
Audio

Free-choice Lecture 1

Raw Audio

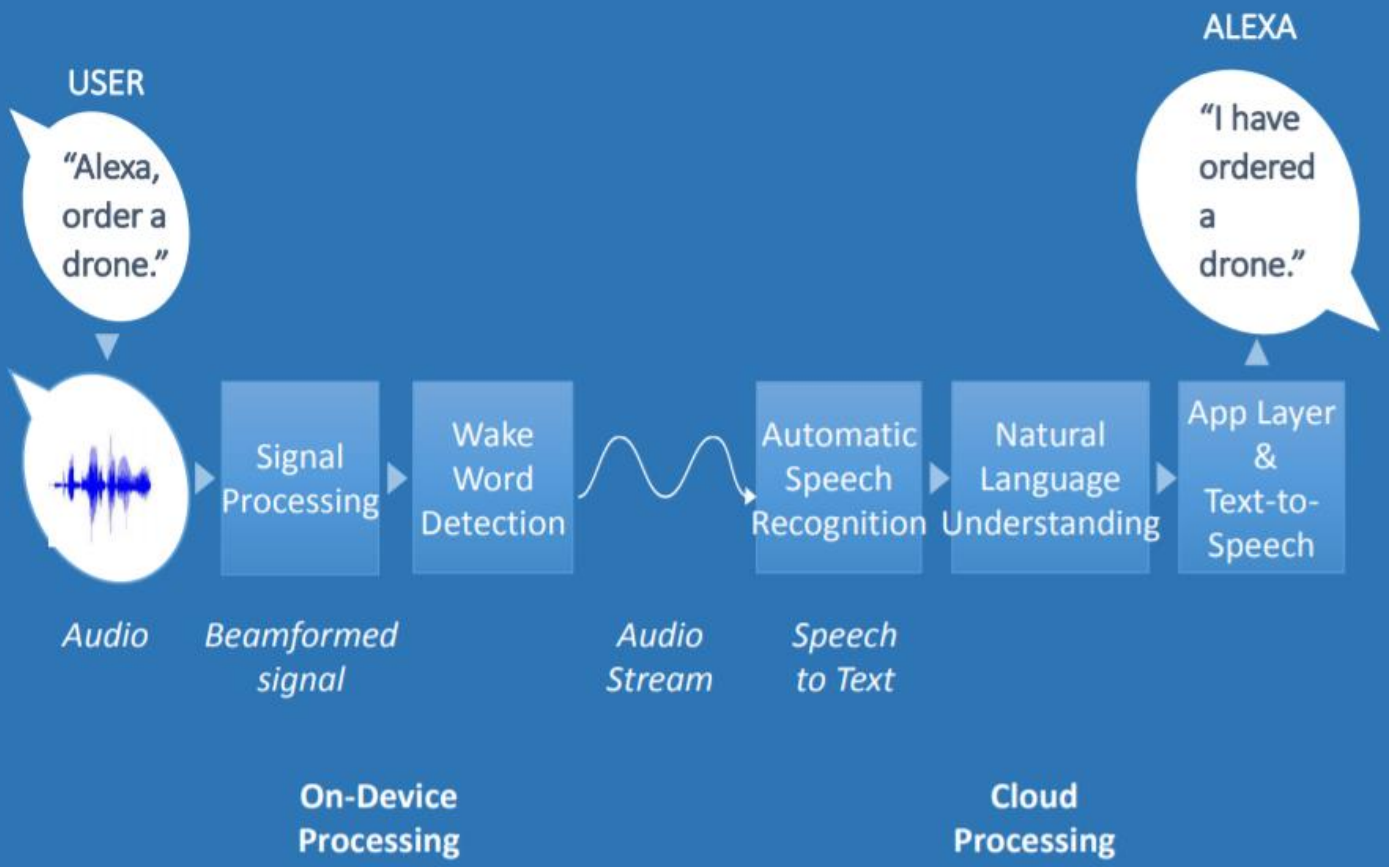






[©Brian Kulis, Ashok Cutkosky]

The Alexa Pipeline



Vision + Language

Free-choice Lecture 2

What are vision-language problems?

Input → output might be...

A. Image → text
B. Text → image
C. Image + text → text

Examples...

A. Image Captioning
B. Sentence to Image Retrieval
C. Visual Question Answering

[©Bryan Plummer, Andrea Burns]

Image Captioning

A woman in a white shirt and denim overalls is walking six dogs in the park.

Generating natural language sentences given the input image

[©Bryan Plummer, Andrea Burns]

Sentence to Image Retrieval

A woman in a white shirt and denim overalls is walking six dogs in the park.

Retrieving an image given the input natural language sentence

[©Bryan Plummer, Andrea Burns]

Sentence to Image Retrieval

A woman in a white shirt and denim overalls is walking six dogs in the park.

Retrieving an image given the input natural language sentence

[©Bryan Plummer, Andrea Burns]

Visual Question Answering

\+ How many dogs are being walked?

→ Six

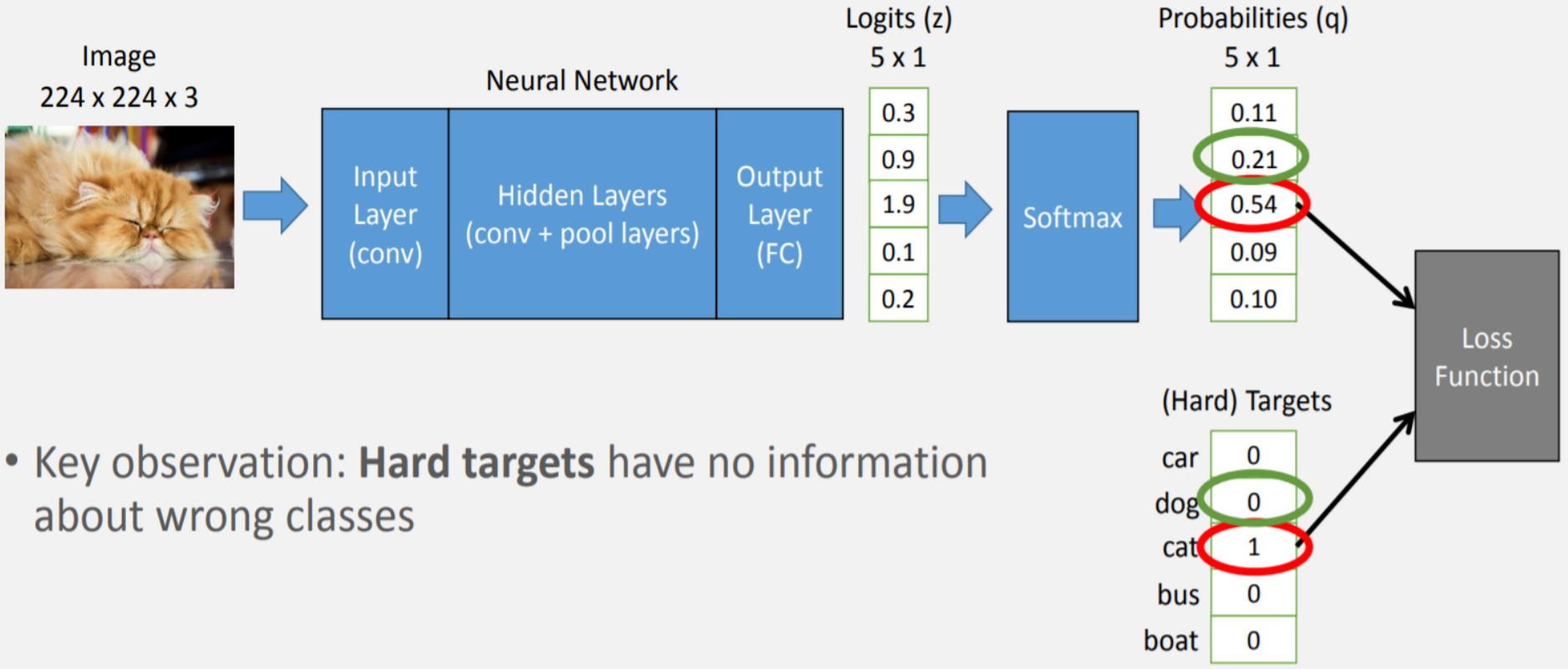
Generating an answer to a visual question, often framed as a classification problem

[©Bryan Plummer, Andrea Burns]

Knowledge Distillation

Free-choice Lecture 3

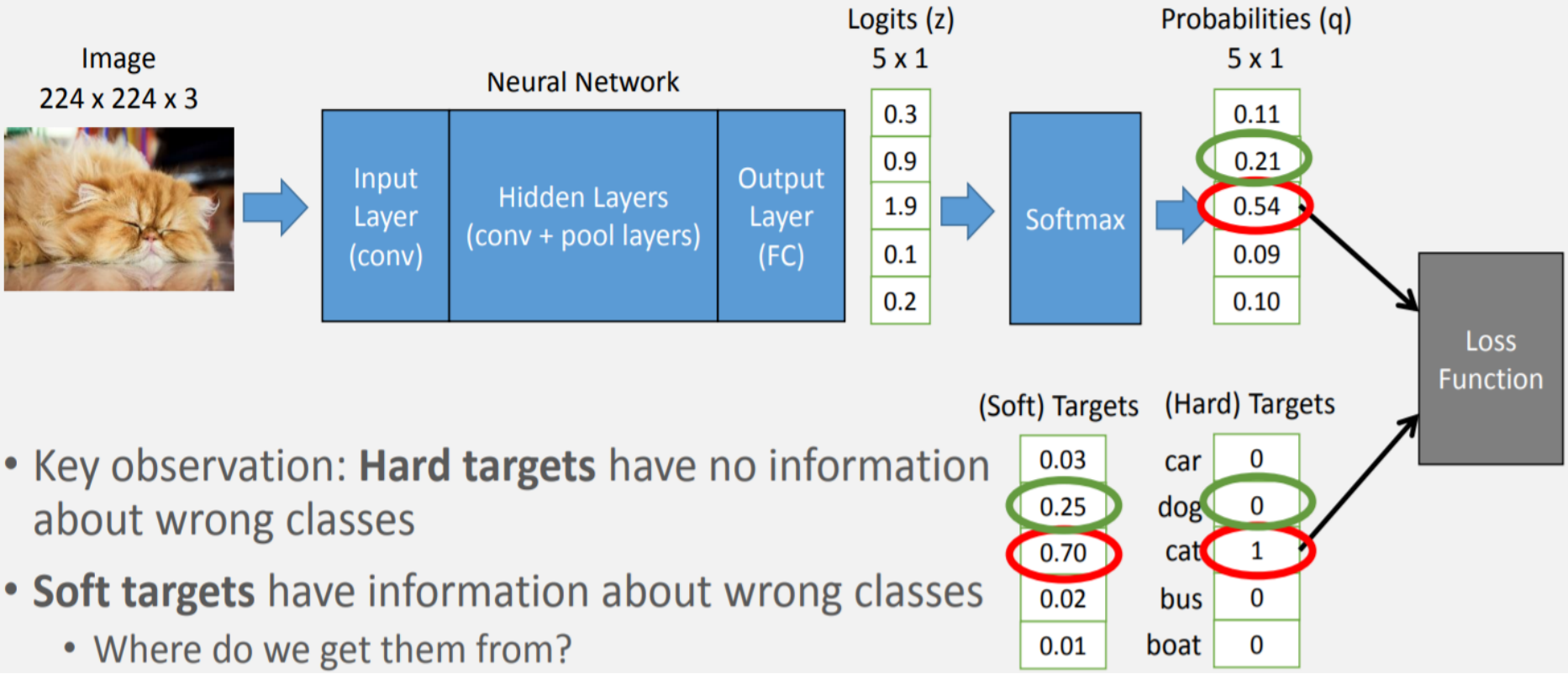
Training a neural network



- Key observation: **Hard targets** have no information about wrong classes

[©Mohamed Abdelfatah]

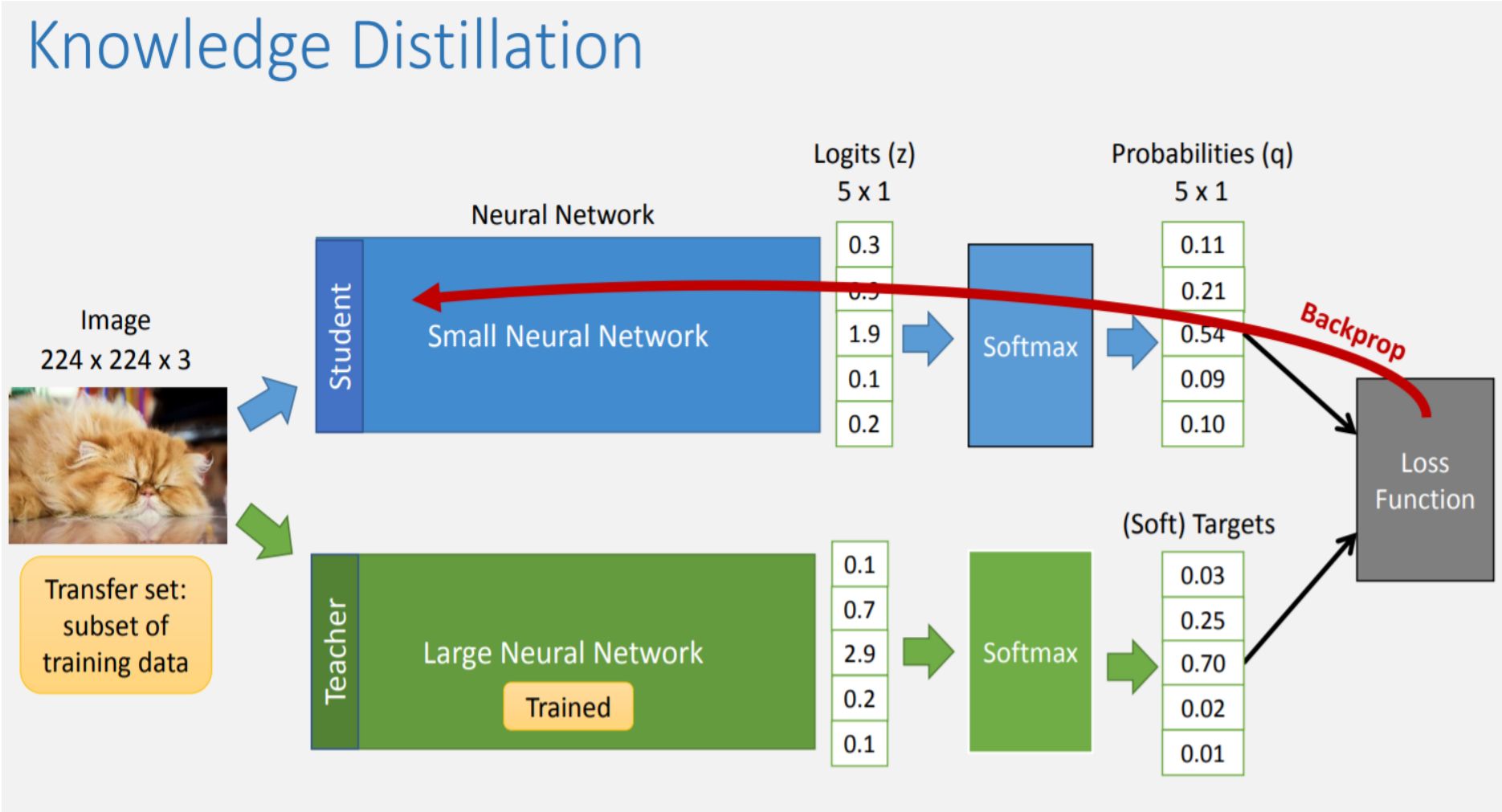
Training a neural network



- Key observation: **Hard targets** have no information about wrong classes
- **Soft targets** have information about wrong classes
  - Where do we get them from?

[©Mohamed Abdelfatah]

Where do we get soft targets from?



[©Mohamed Abdelfatah]

Ethics in Machine Learning

Kate Saenko

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.

Many of these problems are not new!

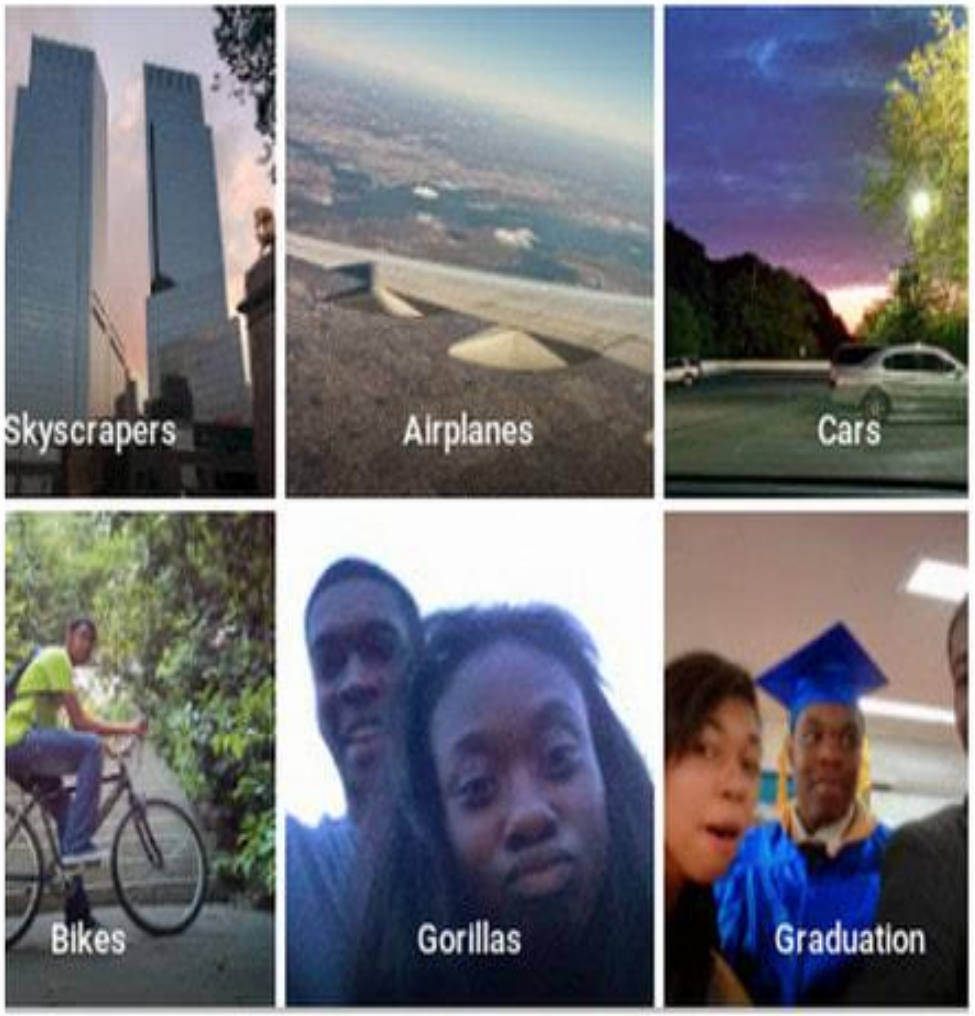
Fears about job automation, lack of privacy and inequality arise with each new innovation

- 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

Bias can lead to offensive or unfair results…



When Hiring for AI …

- Who codes matters
  Are we creating full-spectrum teams with diverse individuals who can check each other's blind spots?

- How we code matters

  Are we factoring in fairness as we're developing systems?

- Why we code matters
  We now have the opportunity to unlock even greater equality

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

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 does *not* 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

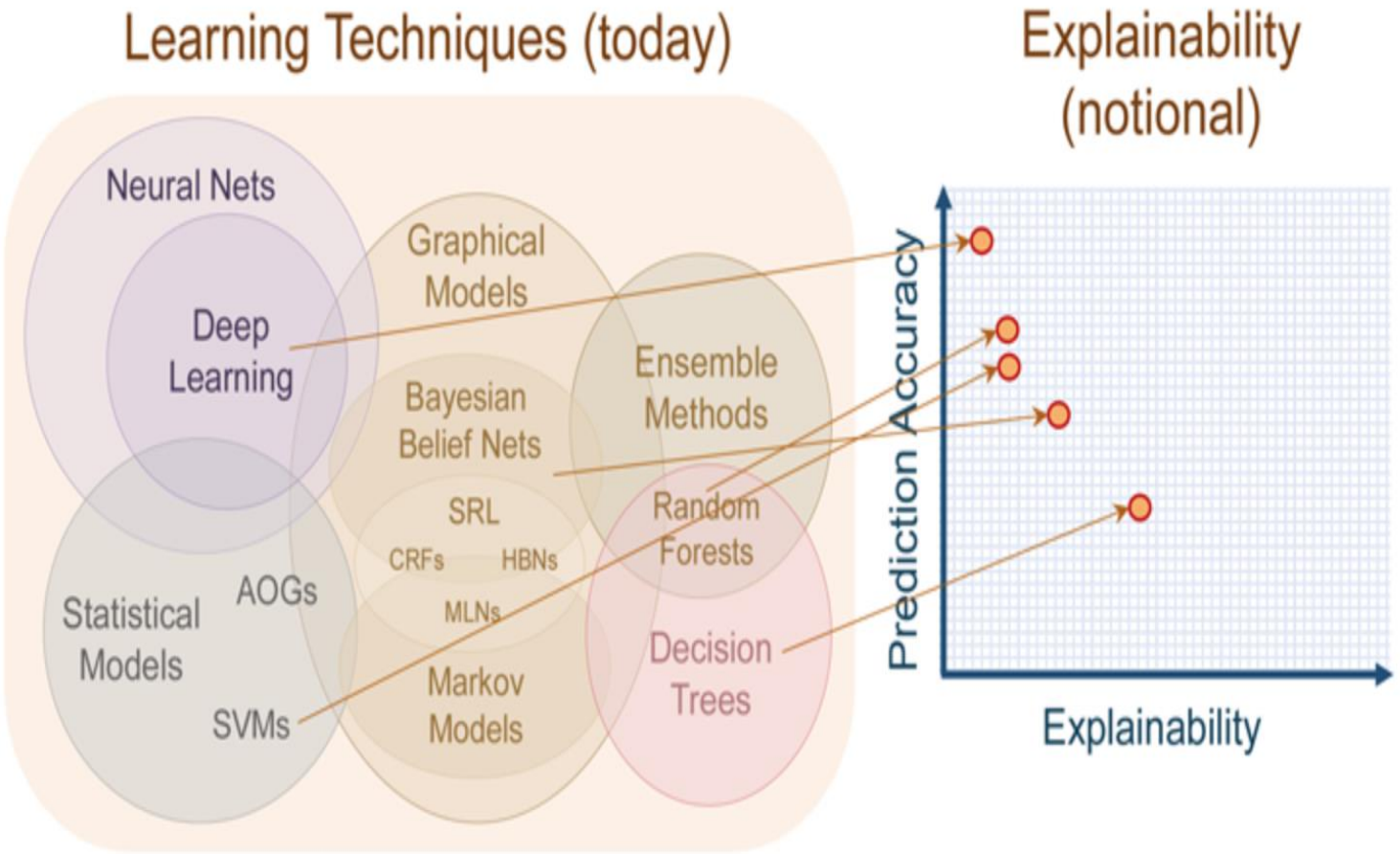
Why fairness is hard

- 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



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

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%

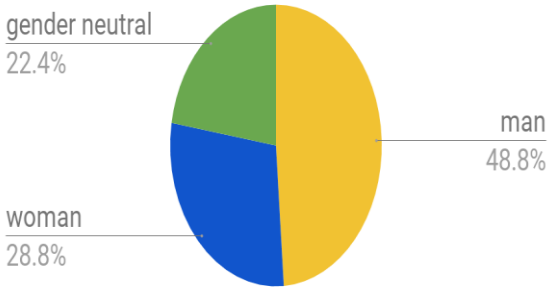
Explanation

True class: “sheep”

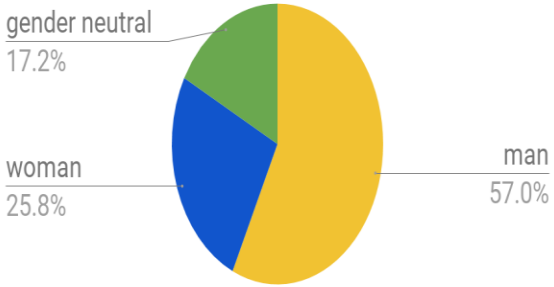
Gender bias in captioning models (Hendricks et al. 2018)

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

Ground truth captions

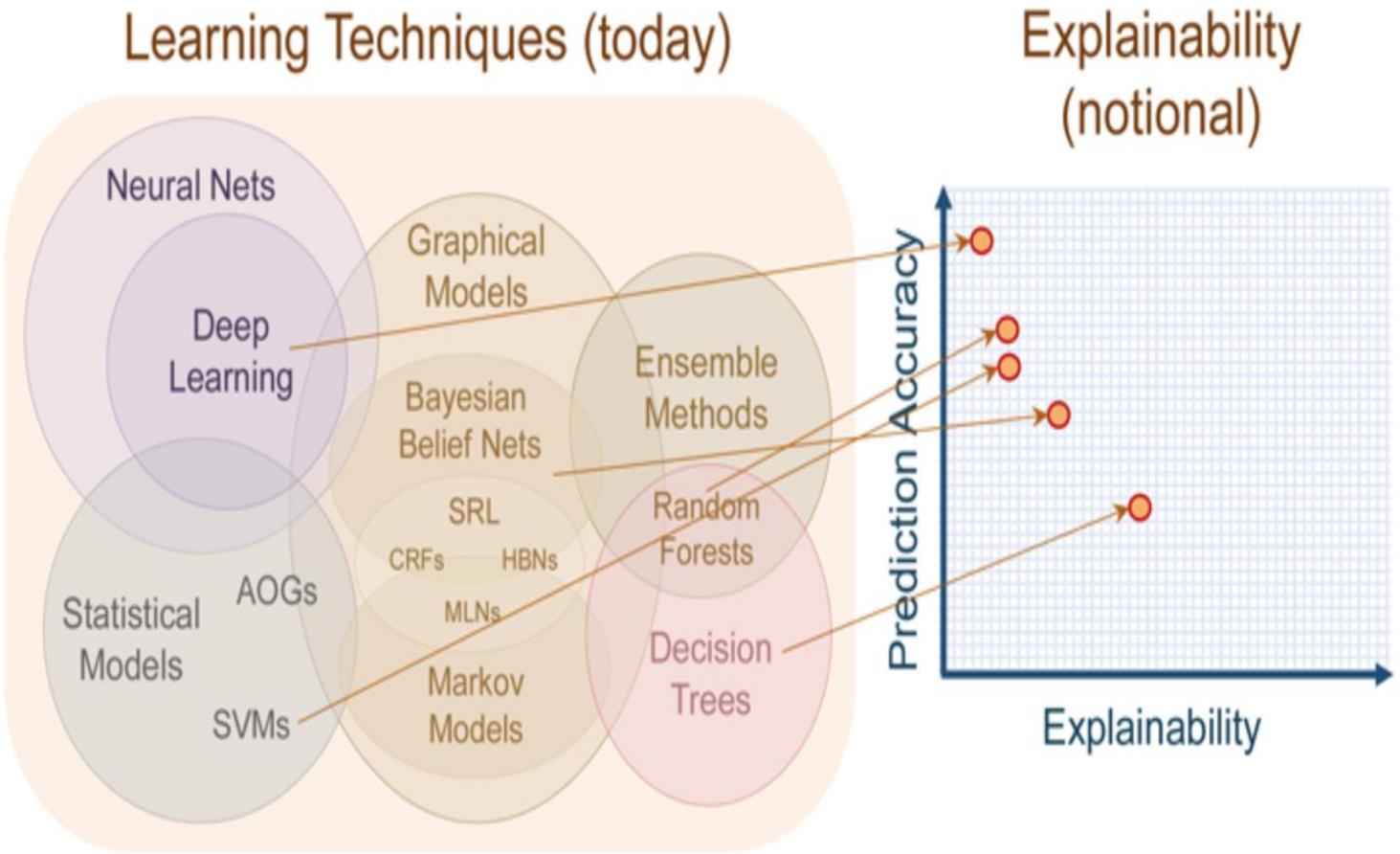


Generated captions

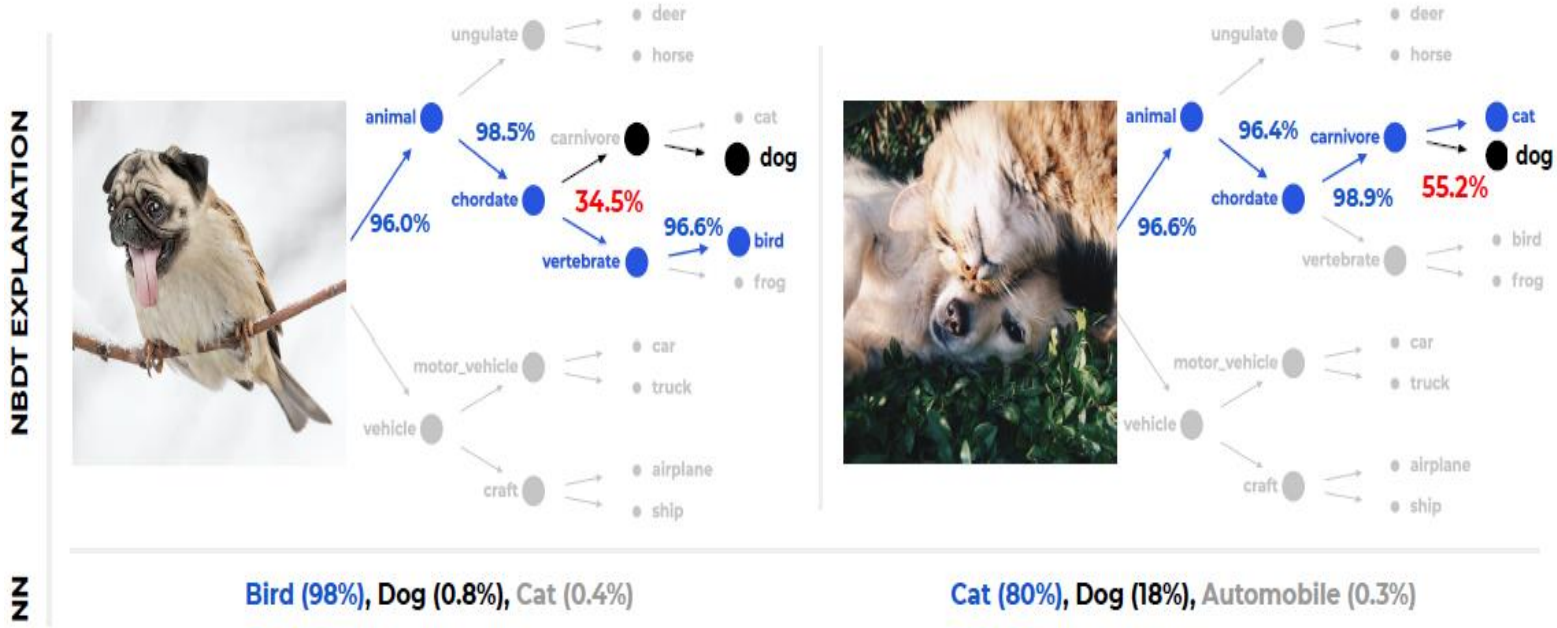


Hendricks et al. "Women Also Snowboard: Overcoming Bias in Captioning Models." ECCV 2018
Zhao et al. "Men also like shopping: Reducing gender bias amplification using corpus-level constraints." EMNLP 2017

Accuracy *vs.* explainability



NBDT: Neural-Backed Decision Trees

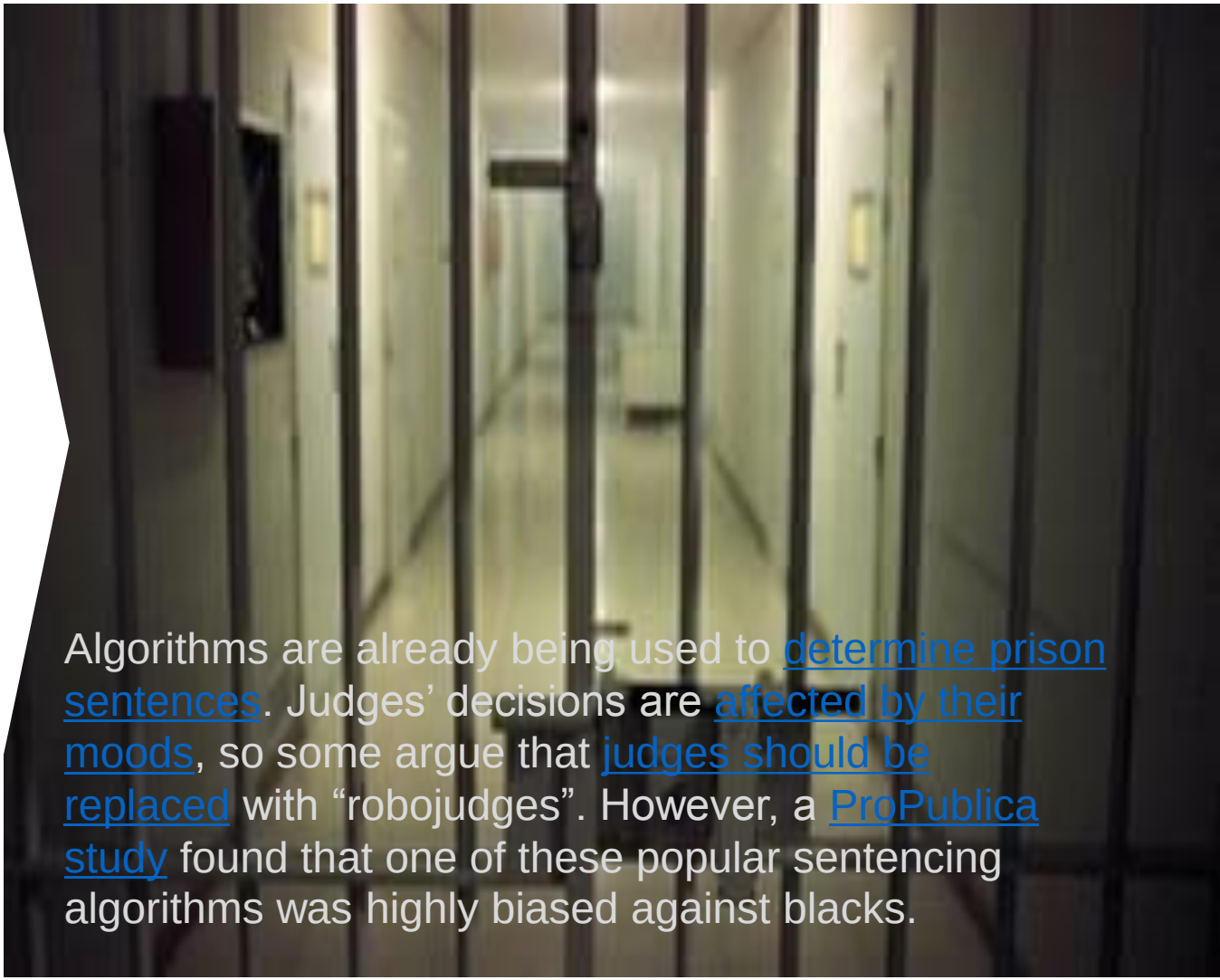


A. Wan, L. Dunlap, D. Ho, J. Yin, S. Lee, H. Jin, S. Petryk, S. A. Bargal, J. E. Gonzalez. NBDT: Neural-Backed Decision Tree. *International Conference on Learning Representations (ICLR) ,* 2021.

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?



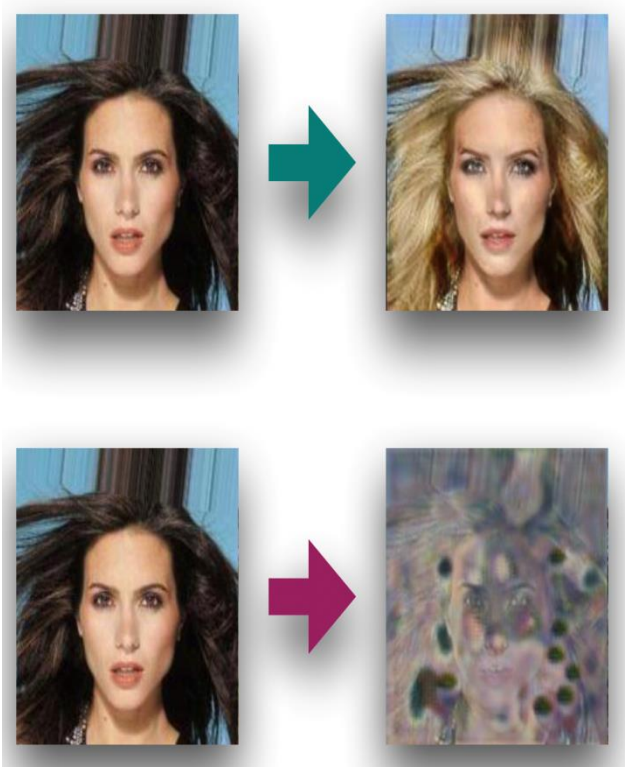
Algorithms are already being used to determine prison sentences. Judges' decisions are affected by their moods, so some argue that judges should be replaced with "robojudges". However, a ProPublica study found that one of these popular sentencing algorithms was highly biased against blacks.

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

Disrupting Deepfakes



N. Ruiz, S. A. Bargal, S. Sclaroff. Disrupting DeepFakes: Adversarial Attacks Against Conditional Image Translation Networks and Facial Manipulation Systems. Workshop on Adversarial Machine Learning in Computer Vision at IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

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 Centre for New American security said in a report that the Chinese company Ziyan is negotiating to sell Blowfish A2, a killer robot capable of 60 millimeter mortar shells or a 35-40 millimeter grenade launcher, to the governments of Pakistan and Saudi Arabia

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

Ethical Issues in Machine Learning

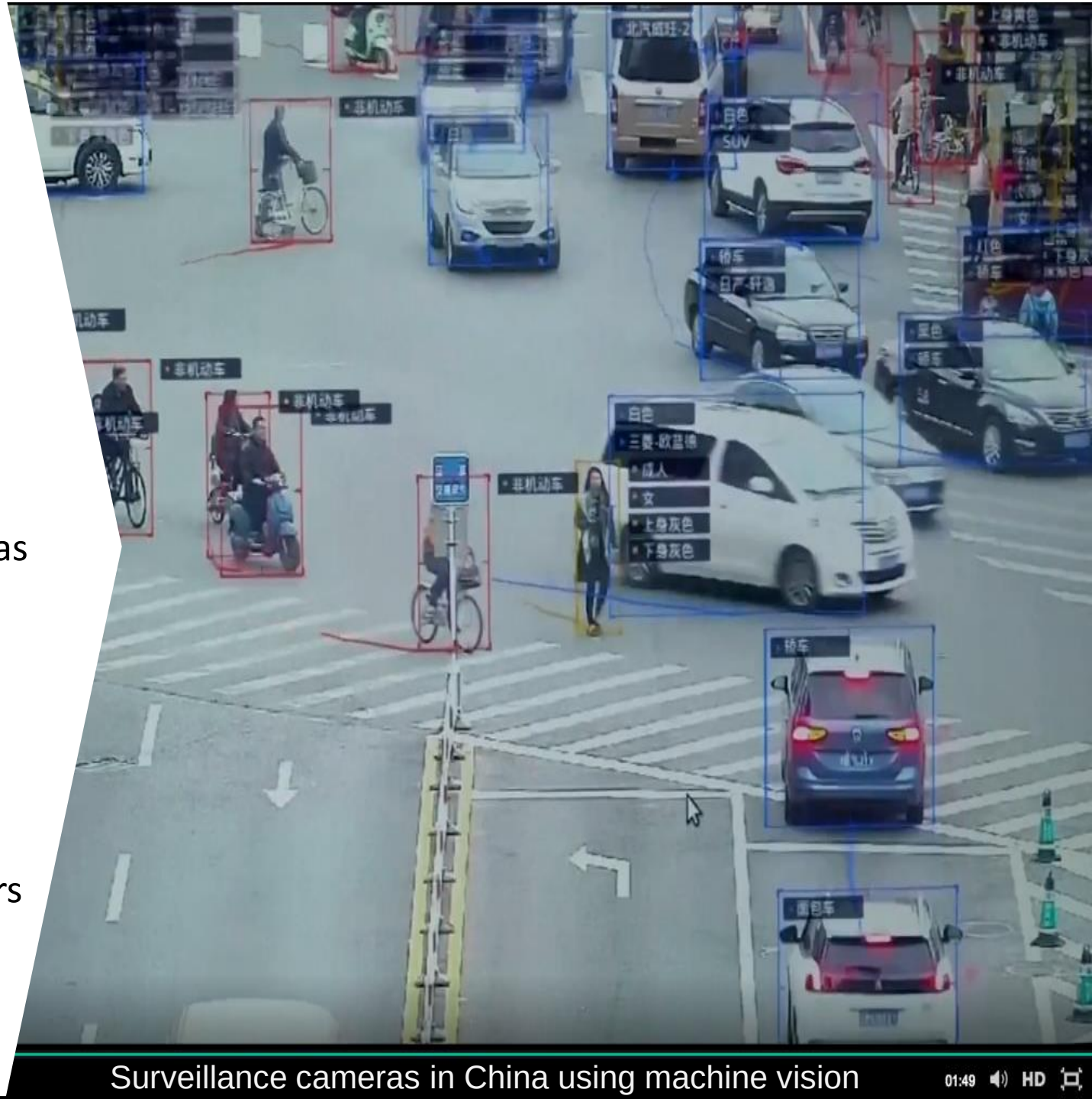
- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance

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

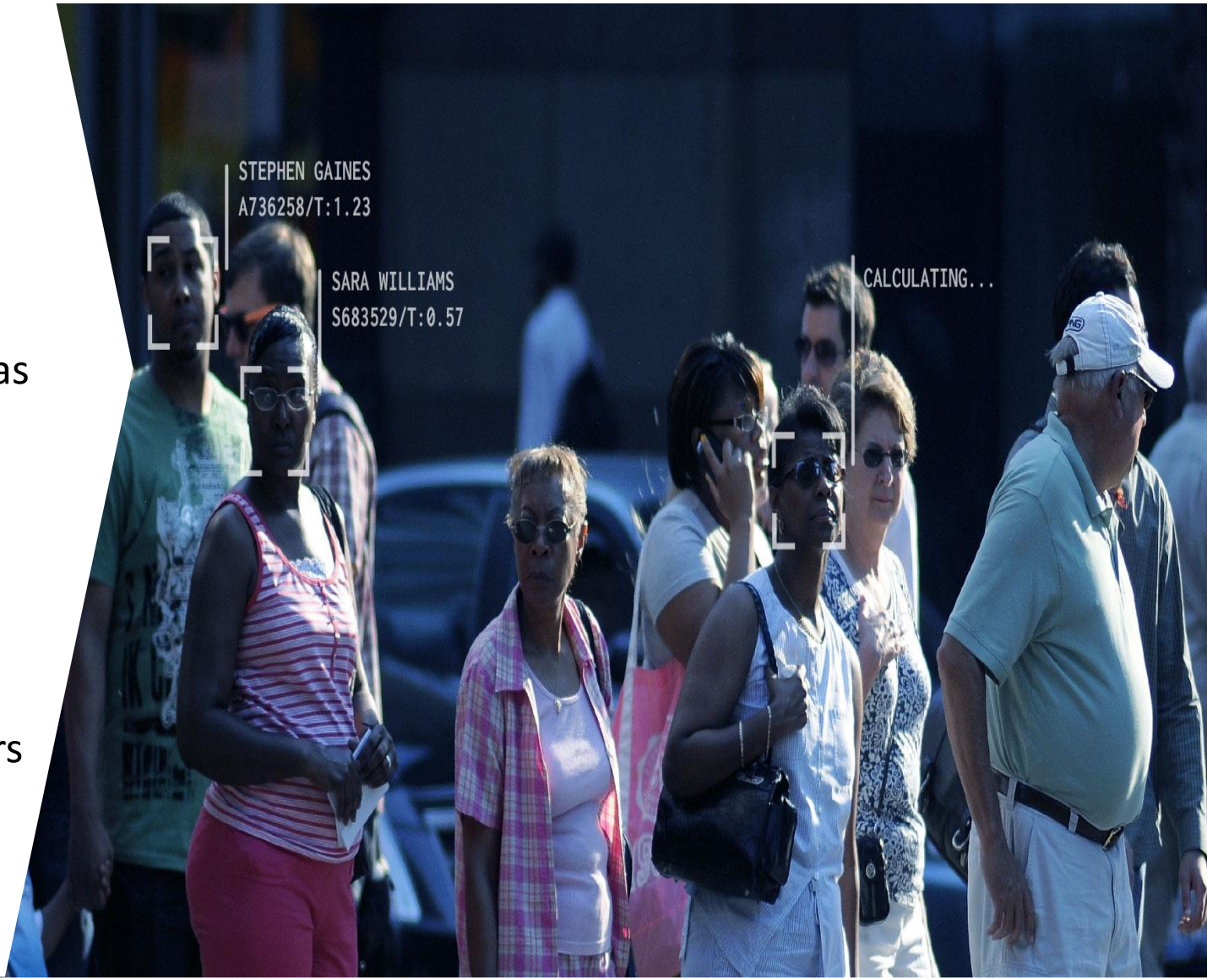
- Job Loss
- Algorithmic Bias
- Transparency
- AI Supremacy
- Fake news and videos
- Autonomous weapons
- Self-driving cars
- Privacy and surveillance



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/



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.fatml.org/