

Yale:

S&DS 265 / 565 Introductory Machine Learning

Societal Issues for Machine Learning

Tuesday, December 7

-0.75142998, 0.24975 , -0.094948 , -0.36341 , 0.24869999, -0.22667 , 0.32289001,

-0.13954 , 0.68976003, 0.21586999, 0.13715 , -1.00919998, 0.028827 , 0.11011 , -0.

0.14463 , -0.18844 , -0.75536001, -0.28704 , 0.019113 , 0.30349001, -0.12111 , -0.

, 0.72409999, 0.50796002, -0.37845999, -0.13008 , -0.13808 , 0.098928 , 0.1621599

Yale

Housekeeping

- Assignment 7 out (neural nets and RL)
- New due date: Tuesday December 14
- Assignment 6 will be graded by Thursday, Dec. 9
- Quiz 4: Today (neural nets and RL); usual protocol
- Final exam, Dec 21 at 7pm (cumulative, 3 hours, cheat sheet, practice exam end of this week)

Outline

- Recall: ML vs. AI
- Examples of bias
- Examples from recent news
- Panel discussion

Today: Home assistants



Pricing and recommending homes

THE WALL STREET JOURNAL.

Subscribe Now | Sign In

\$1 for 2 months

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life Real Estate 



CIO JOURNAL



Zillow Develops Neural Network to 'See' Like a House Hunter

Granite or stainless steel countertops? Zillow's visual recognition effort can recognize the difference

By **SARA CASTELLANOS**

Nov 11, 2016 3:29 pm ET

Data scientists at Zillow Group are developing complex computer programs that detect specific attributes in photographs of homes, which could aid in estimating their value. Advances in deep learning, big data and cloud computing have converged to allow the online real estate database firm and others to develop technology that mimics how the human brain [...]

Recommended Videos

1. Film Clip: Pirates of the Caribbean: Dead Men Tell No Tales'



2. What to do in your 40s to retire a millionaire



<https://blogs.wsj.com/cio/2016/11/11/zillow-develops-neural-network-to-see-like-a-home-buyer/>

Email suggestions

The screenshot shows an email inbox interface. A dark gray callout box is overlaid on the screen, containing the text "Taco Tuesday". This box is positioned above an email from Jacqueline Bruzek. The email subject is "Taco Tuesday". The body of the email starts with "Hey Jacqueline," followed by the partially typed message "Haven't seen you in a while and I hope you're doing well.". In the top right corner of the email preview, there is a red square icon with the letters "ET" in white.

Taco Tuesday

Jacqueline Bruzek

Taco Tuesday

Hey Jacqueline,

Haven't seen you in a while and I hope you're doing well.

<https://www.youtube.com/watch?v=nZ-C8I-8BZw&t=0m16s>

AI vs. ML

Machine learning focuses on making predictions and inferences from data.

AI combines machine learning components into a larger system that includes a decision making component.

An AI system exhibits a behavior, resulting from the collective decisions that are made.

Machine learning frameworks

- Supervised, unsupervised, semi-supervised
- Reinforcement learning
- Generative vs. discriminative models
- Representation learning

Example of representation learning: Word embeddings

- Each word in vocab is mapped to 100 or 500 dimensional vector
- Based solely on co-occurrence statistics in corpus of text

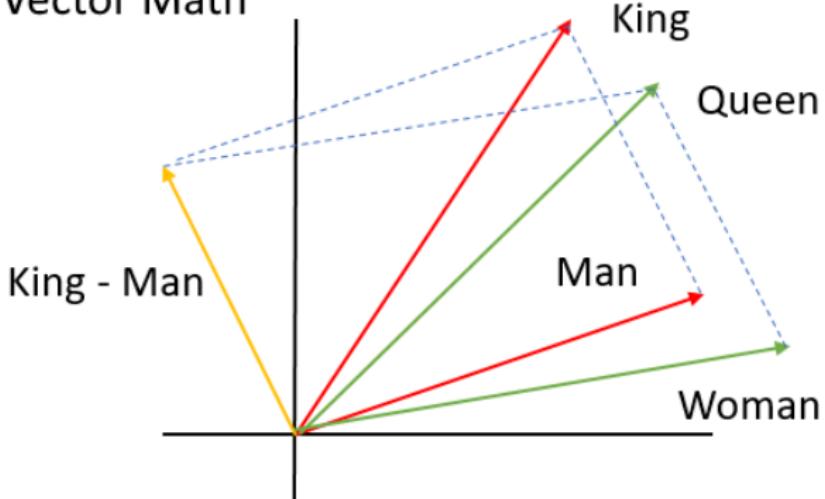
Example of representation learning: Word embeddings

Yale:

```
[ 0.78310001, 0.51717001, -0.38207 , -0.23722 , -0.31615999, 0.30805001, 0.76389998, 0.064106 , -0.74913001,  
 0.60585999, -0.23871 , -0.16876 , -0.25634 , 1.07270002, -0.29967999, 0.020095 , 0.54500997, -0.17847 , -0.26675999,  
 -0.11798 , -0.48692 , 0.22712 , 0.017473 , -0.4747 , 0.44861001, -0.084281 , -0.30412999, -1.13510001, -0.14869 , -0.11182 ,  
 -0.32530001, 1.0029 , -0.35742 , 0.35148999, -1.10679996, -0.064142 , -0.72284001, 0.14114 , -0.41247001, -0.16184001,  
 -0.54576999, -0.12958001, -0.88356 , -0.089722 , 0.10555 , -0.12288 , 0.92851001, 0.50032002, 0.1349 , 0.21457 ,  
 0.35073999, -0.73132998, 0.39633 , -0.43239999, -0.38815999, -1.34669995, 0.37463999, -0.79386002, 0.11185 , 0.18007 ,  
 -0.75142998, 0.24975 , -0.094948 , -0.36341 , 0.24869999, -0.22667 , 0.32289001, 1.29489994, 0.42658001, 1.29120004,  
 -0.13954 , 0.68976003, 0.21586999, 0.13715 , -1.00919998, 0.028827 , 0.11011 , -0.1912 , -0.073198 , -0.52449 , 0.49199 ,  
 0.14463 , -0.18844 , -0.75536001, -0.28704 , 0.019113 , 0.30349001, -0.74425 , -0.072221 , -0.40647 , 0.26899001, -0.28318  
, 0.72409999, 0.50796002, -0.37845999, -0.13008 , -0.13808 , 0.098928 , 0.16215999, 0.16293 ]
```

Word geometry

Vector Math



Embeddings encode societal bias

$$\phi(\text{scientist}) - \phi(\text{woman}) + \phi(\text{man}):$$

geologist
engineer
astronomer
mathematician
science

$$\phi(\text{scientist}) - \phi(\text{man}) + \phi(\text{woman}):$$

anthropologist
sociologist
psychologist
geneticist
biochemist

Embeddings encode societal bias

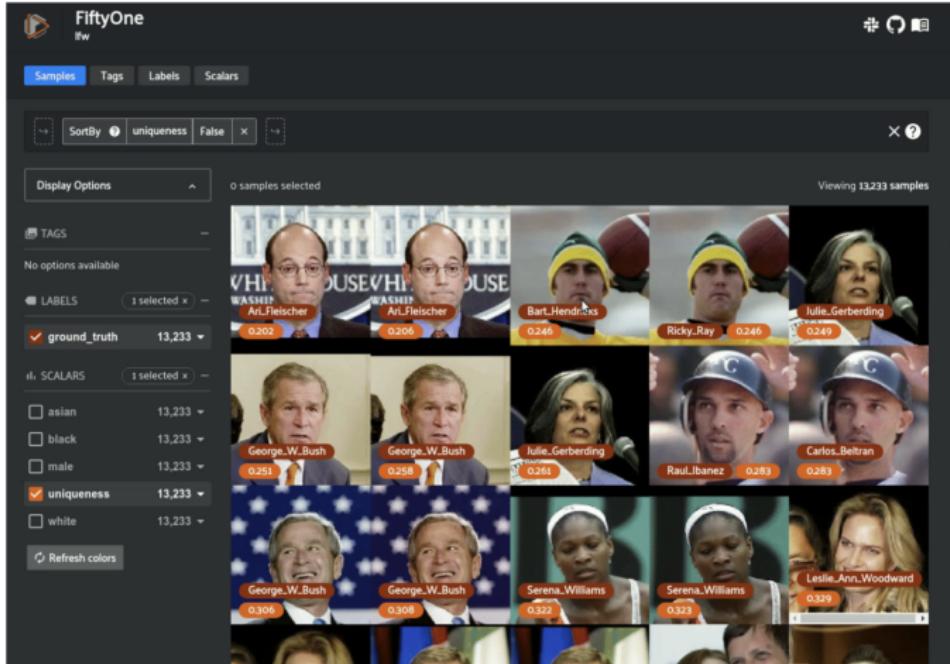
$$\phi(\text{smart}) - \phi(\text{girl}) + \phi(\text{boy}):$$

wise
better
guy
kind
good
kid

$$\phi(\text{smart}) - \phi(\text{boy}) + \phi(\text{girl}):$$

sexy
pretty
incredibly
cute
exciting
funny

Bias in LFW dataset



Sorting by the least unique images to find duplicates and incorrect labels

Hacking AI systems

[SUBSCRIBE](#)[SIGN IN](#)

TESLA AUTOPILOT —

Researchers trick Tesla Autopilot into steering into oncoming traffic

Stickers that are invisible to drivers and fool autopilot.

DAN GOODIN - 4/1/2019, 8:50 PM

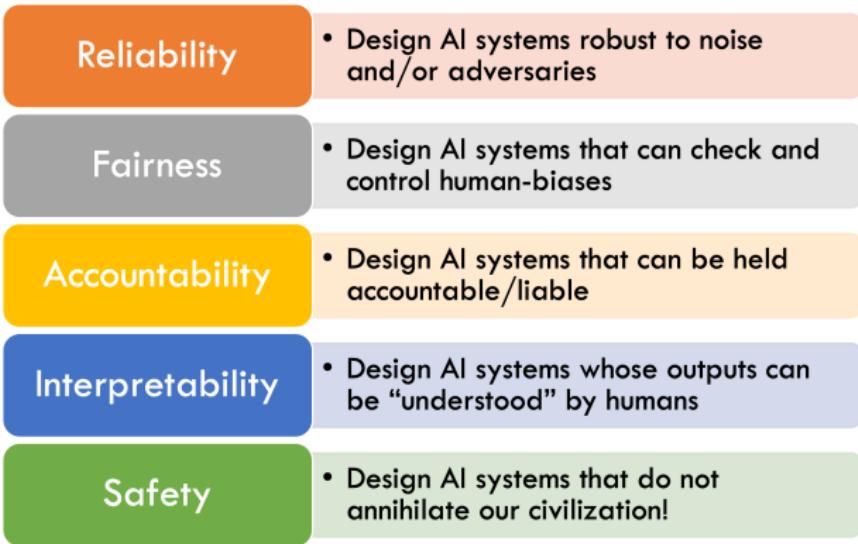
Keen Security Lab



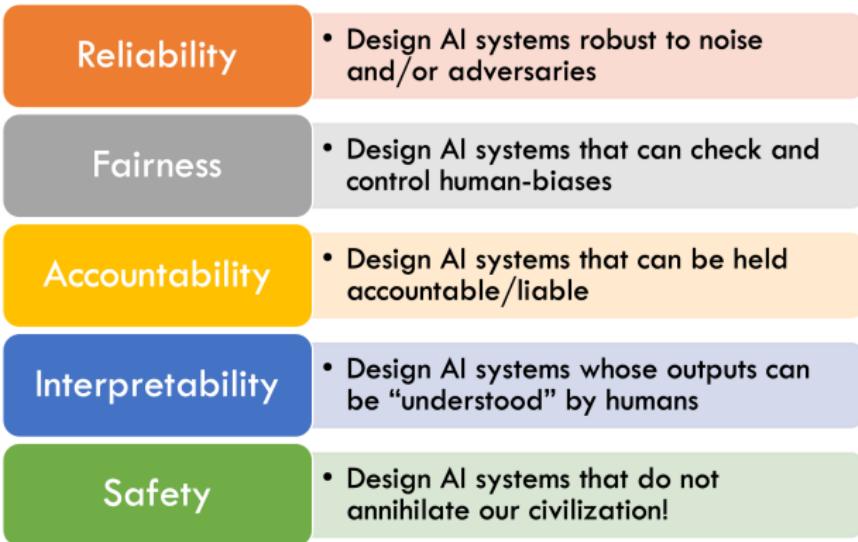
Machine learning at a large Internet company

- Typical project lifetime: 6 months to 1 year
- Ads projects involve thousands of software engineers
- Often adding new “feature” to existing black box model
- No single person understands entire model
- Not interpretable
- Security issues with data

Important directions



Important directions



- Requires a principled, interdisciplinary approach

Descriptions in press

The Scientist and the A.I.-Assisted, Remote-Control Killing Machine

Israeli agents had wanted to kill Iran's top nuclear scientist for years. Then they came up with a way to do it with no operatives present.



Descriptions in press

Jerusalem Post > Middle East > Iran News

Iran denies NYT Mossad assassination report

The New York Times published a report detailing the assassination of Iran's leading nuclear scientist by a Mossad-operated AI machine gun.

By JERUSALEM POST STAFF Published: SEPTEMBER 19, 2021 20:36



Descriptions in press

NONFICTION

A Robot Wrote This Book Review



Elliot Ulm

Descriptions in press

By **Kevin Roose**

Nov. 21, 2021

THE AGE OF AI

And Our Human Future

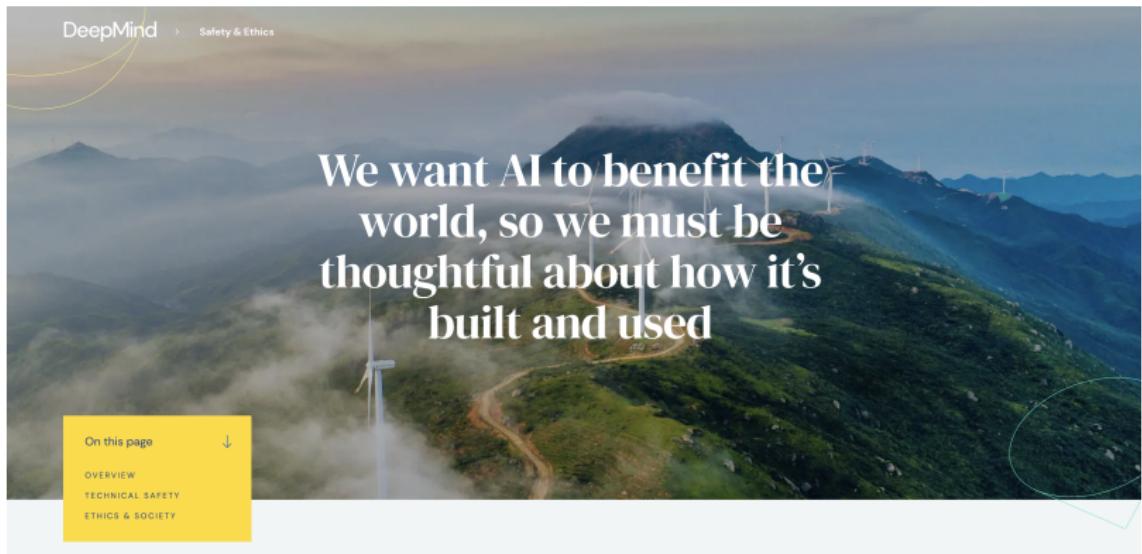
By Henry A. Kissinger, Eric Schmidt and Daniel Huttenlocher

One of the great promises of technology is that it can do the work that humans find too boring or arduous.

In the 19th and 20th centuries, factory machines relieved us of repetitive manual labor and backbreaking farm work. In this century, artificial intelligence has taken care of a few more tasks — curating Spotify playlists, selecting the next YouTube video, vacuuming the floor and so on — but many more mind-numbing activities remain ripe for the picking. The experts promise us that someday, all of our least favorite chores — including complex cognitive ones, like interviewing job candidates or managing global supply chains — will be outsourced to machines.

But that day has not yet arrived. Or has it?

Corporate initiatives



DeepMind Safety & Ethics

We want AI to benefit the world, so we must be thoughtful about how it's built and used

On this page ↓

- OVERVIEW
- TECHNICAL SAFETY
- ETHICS & SOCIETY

Corporate initiatives

TOM SIMONITE

BUSINESS 12.02.2021 08:00 AM

Ex-Googler Timnit Gebru Starts Her Own AI Research Center

The researcher, who says Google fired her a year ago, wants to ask questions about responsible use of artificial intelligence.



ILLUSTRATION WIRED STAFF; GETTY IMAGES

Further listening



PLAY ►

Is A.I. the Problem? Or Are We?

The Ezra Klein Show

Society & Culture

[Listen on Apple Podcasts ↗](#)



If you talk to many of the people working on the cutting edge of artificial intelligence research, you'll hear that we are on the cusp of a technology that will be far more transformative than simply computers and the internet, one that could bring about a new industrial revolution and usher in a utopia — or perhaps pose the greatest threat in our species's history.

Others, of course, will tell you those folks are nuts.

Panel Discussion

Team A



Boris Cyusa



Eliza Oak



Lucas Zheng

Panel Discussion

Team A

Dystopian view: AI and ML are going to result in major societal inequities. The harm may outweigh the benefits unless “we” are very careful, including corporations and government entities.

Panel Discussion

Team B



Aishwarya Iyer

Jorge Herrera

Medha Majety

Panel Discussion

Team B

Utopian view: AI and ML are going to result in major societal benefits. The technology is unprecedented, and people will adapt to harness it in ways that will improve the human condition.