Georgia
Tech

Lecture 24
Ethical Issues in using Machine Learning and AI

Mahdi Roozbahani
Georgia Tech

These slides are from **Nakul Gopalan** at Georgia Tech

A Few Notes (From Nakul's notes)

- I am trying this for the first time the topic itself is nascent in its development

- Topics might include ideas of discrimination, bias, ethnicity

- Do not assume malice, instead assume people are making arguments in good faith!!

Outline

- Success of Machine learning/ AI
- Case studies in negative effects of ML:
    - Job Losses
    - Privacy
    - Bias
    - Manipulation with ML
- Ethical considerations in other sciences
- Solutions?
    - Explainability
    - Bias reduction
    - Novel class of algorithms

Outline

- Success of Machine learning/ AI
- Case studies in negative effects of ML:
  - Job Losses
  - Privacy
  - Bias
  - Manipulation with ML
- Ethical considerations in other sciences
- Solutions?
  - Explainability
  - Bias reduction
  - Novel class of algorithms

The Wins

- Natural Language Processing

- Search

- Forecasting: Energy/ Weather/ Finance/ Covid

- Autonomous driving/ Submarines/ Drones

- Learning Controllers - Nest

- Drug discovery/ automated image analysis

- Game Playing

- Traffic prediction

ML applications in social good

- Forecasting severe weather events: US, Australia, Floods in Asia etc.
- Tracking and protecting wildlife: in Tanzania, marine wildlife
- Rural water supply
- Helping blind people sense
- Detecting plant diseases
- Other healthcare based applications
- https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good
- https://ai.google/social-good/
- Series of workshops about AI for social good: https://aiforsocialgood.github.io

Outline

- Success of Machine learning/ AI
- Case studies in negative effects of ML:
  - Job Losses
  - Privacy
  - Bias
  - Manipulation with ML
- Ethical considerations in other sciences
- Solutions?
  - Explainability
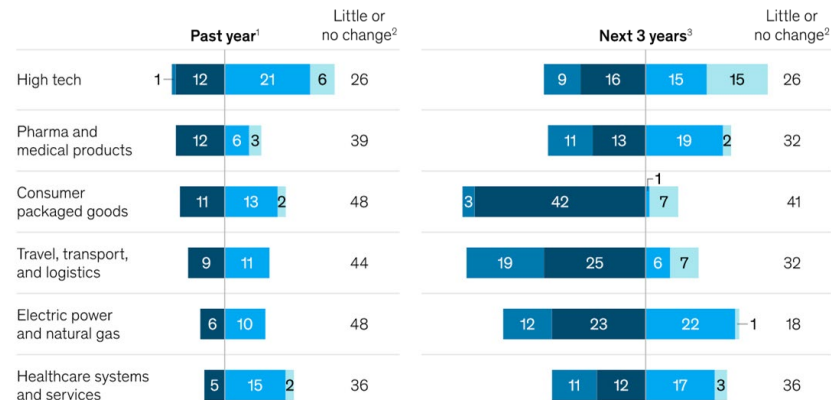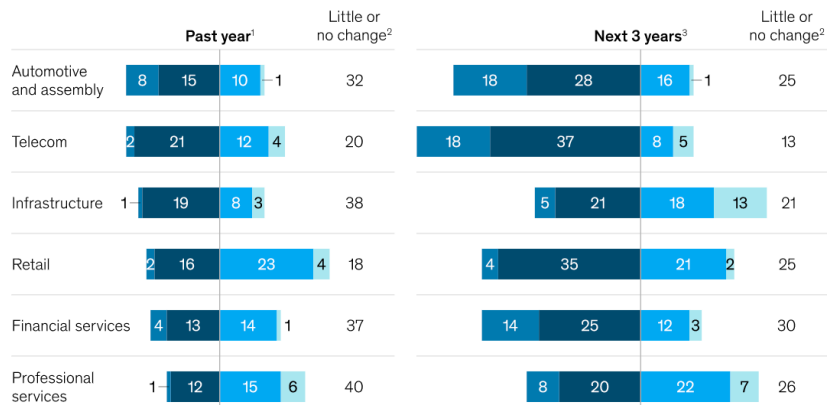 - Bias reduction
  - Novel class of algorithms

Job Losses

- Trucking
- Manufacturing
- Clerical and Secretarial jobs
- Coding???

Jobs at risk of automation in the EU 19 countries

▶ **Executives from 1,872 enterprises worldwide report the largest AI-induced workforce contraction in automotive and assembly and telecoms in the last year. Looking forward, the CPG, transport, utilities, retail and financial services are expected to follow.**

Change in workforce due to AI adoption, % of respondents

- Decrease >10%
- Decrease 3–10%
- Increase 3–10%
- Increase >10%

	Past year[1]				Little or no change[2]		Next 3 years[3]				Little or no change[2]
Automotive and assembly	8	15	10	1	32		18	28	16	1	25
Telecom	2	21	12	4	20		18	37	8	5	13
Infrastructure	1	19	8	3	38			5	21	18	13	21
Retail	2	16	23	4	18		4	35	21	2	25
Financial services	4	13	14	1	37		14	25	12	3	30
Professional services	1	12	15	6	40		8	20	22	7	26

	Past year[1]				Little or no change[2]		Next 3 years[3]				Little or no change[2]
High tech	1	12	21	6	26		9	16	15	15	26
Pharma and medical products	12	6	3		39		11	13	19	2	32
Consumer packaged goods	11	13	2		48		3	42	7	1	41
Travel, transport, and logistics	9	11			44		19	25	6	7	32
Electric power and natural gas	6	10			48		12	23	22	1	18
Healthcare systems and services	5	15	2		36		11	12	17	3	36

McKinsey & Company

stateof.ai 2020

Job Losses

The loom unemployed mill workers??? (Increasing the output per worker)

Spreadsheets caused job losses amongst accountants???

Will governments step in and help??

What do you think?

Privacy: Cookies

- Easy tracking of people with anonymized IDs
- Selling these fingerprints for advertisements
- Apps collecting data without knowing what it will be used for: Normalized by major ML companies
- General Data Protection Regulation (GDPR):
  - Laws about what can be collected
  - Default settings
  - Sharing

Privacy: Images

- Creating customer IDs from images
- Images on apps like FB
- [Tracking customers inside and near showrooms](#)
- Traffic cameras
- Surveillance in oppressive governments – cause of major concern
- Amazon/ UPS/ large mail carriers tracking employee behavior
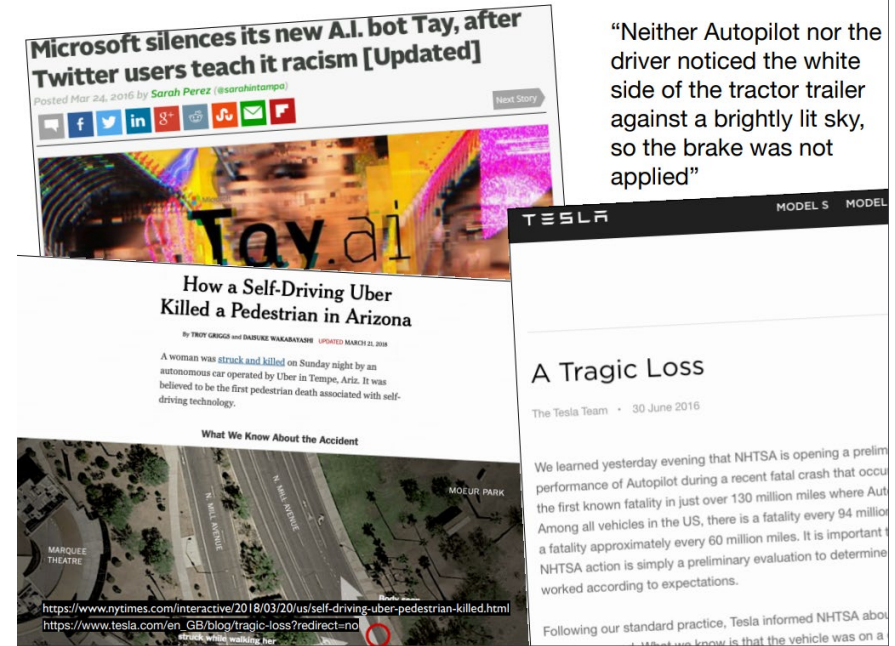
Privacy: Guiding behavior

- Advert clicks
- [Manipulating user's feed based on behavior](#) / [Wiki link](#)
- OKCupid: [We experiment on human beings](#)
- Uber/ Lyft

Privacy

What do we think????

Bias

- We use ML in criminal sentencing
- Biased data in Biased decisions out
- We us ML in deciding lending rates
- Biased data in Biased decisions out
- Bias in large language models

Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism [Updated]

Posted Mar 24, 2016 by Sarah Perez (@sarahintampa)

How a Self-Driving Uber Killed a Pedestrian in Arizona

By TROY GRIGGS and DAISUKE WAKABAYASHI   UPDATED MARCH 21, 2018

A woman was struck and killed on Sunday night by an autonomous car operated by Uber in Tempe, Ariz. It was believed to be the first pedestrian death associated with self-driving technology.

What We Know About the Accident

https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html
https://www.tesla.com/en_GB/blog/tragic-loss?redirect=no

"Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied"

TESLA   MODEL S   MODEL

A Tragic Loss

The Tesla Team • 30 June 2016

We learned yesterday evening that NHTSA is opening a prelim performance of Autopilot during a recent fatal crash that occu the first known fatality in just over 130 million miles where Aut Among all vehicles in the US, there is a fatality every 94 million a fatality approximately every 60 million miles. It is important t NHTSA action is simply a preliminary evaluation to determine worked according to expectations.

Following our standard practice, Tesla informed NHTSA abo

15
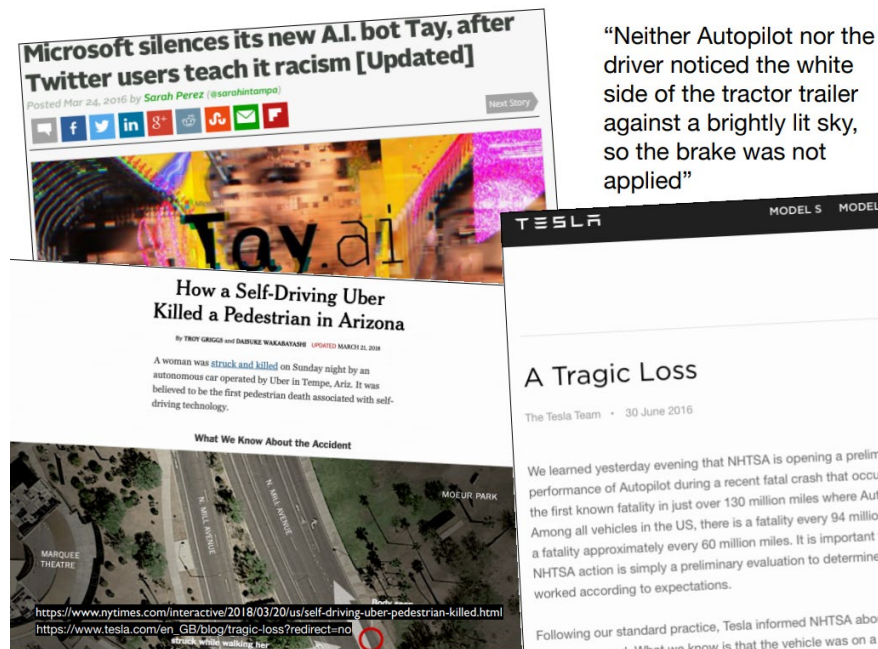
Bias

One of the larger problems in ML

Overfitting and Bias are serious problems with real life consequences

Bias

What do we think?
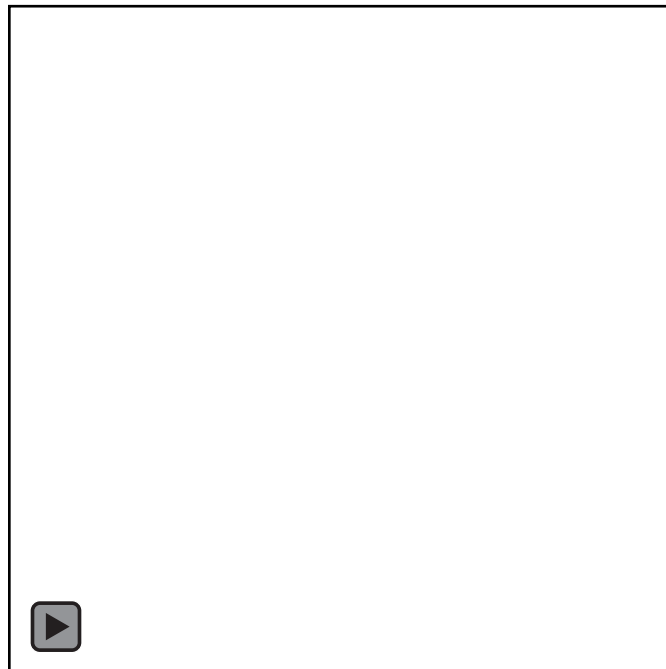
Manipulation with ML

- Text generation for News
- Picture/ Video Manipulation - Deep fakes

Microsoft silences its new A.I. bot Tay, after Twitter users teach it racism [Updated]

Posted Mar 24, 2016 by Sarah Perez (@sarahintampa)

Next Story

How a Self-Driving Uber Killed a Pedestrian in Arizona

By TROY GRIGGS and DAISUKE WAKABAYASHI   UPDATED MARCH 21, 2018

A woman was struck and killed on Sunday night by an autonomous car operated by Uber in Tempe, Ariz. It was believed to be the first pedestrian death associated with self-driving technology.

What We Know About the Accident

MOEUR PARK

MARQUEE THEATRE

https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html
https://www.tesla.com/en_GB/blog/tragic-loss?redirect=no

struck while walking her

"Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied"

TESLA   MODEL S   MODEL

A Tragic Loss

The Tesla Team   ·   30 June 2016

We learned yesterday evening that NHTSA is opening a prelim performance of Autopilot during a recent fatal crash that occu the first known fatality in just over 130 million miles where Aut Among all vehicles in the US, there is a fatality every 94 millio a fatality approximately every 60 million miles. It is important t NHTSA action is simply a preliminary evaluation to determine worked according to expectations.

Following our standard practice, Tesla informed NHTSA abou

18

Some deep fake examples:

Manipulation with ML

What do we think?

Ethical risks: A group of researchers have spent years helping to frame the ethical risks of deploying ML in certain sensitive contexts. This year those issues went mainstream.

▶ **Examples include policing, the judiciary and the military. A few trailblazing researchers include:**

- *Joy Buolamwini, Timnit Gebru*, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification (2018)
- *Clare Garvie, Alvaro Bedoya, and Jonathan Frankle.* The Perpetual Line-Up: Unregulated Police Face Recognition in America (2016)
- *Adam Harvey.* Megapixels (2017)
- *P Allo, M Taddeo, S Wachter, L Floridi.* The ethics of algorithms: Mapping the debate (2016)
- *Margaret Boden, Joanna Bryson, Alan Winfield* et al. Principles of robotics: regulating robots in the real world (2017)

STEPHEN GAINES
A736258/T:1.23

SARA WILLIAMS
S683529/T:0.57

CALCULATING...

THE PERPETUAL LINE-UP

UNREGULATED POLICE FACE RECOGNITION IN AMERICA

OCTOBER 18, 2016

stateof.ai 2020

Outline

- Success of Machine learning/ AI
- Case studies in negative effects of ML:
  - Job Losses
  - Privacy
  - Bias
  - Manipulation with ML
- **Ethical considerations in other sciences**
- Solutions?
  - Explainability
  - Bias reduction
  - Novel class of algorithms

Medicine

- Ability to do harm

➢ Hippocratic Oath

Genetics

- Eugenics
- Human genetics experiment

➢ Laws, but after a while

Finance

- Fiduciary responsibility
- Insider trading

➢ Financial regulation

Outline

- Success of Machine learning/ AI
- Case studies in negative effects of ML:
  - Job Losses
  - Privacy
  - Bias
  - Manipulation with ML
- Ethical considerations in other sciences
- **Solutions?**
  - Explainability
  - Bias reduction
  - Novel class of algorithms

Explainability

- Explain why?
- Modern machine learning algorithms do not explain their decisions (We have seen it in CNN)
- Approaches:
  - Examples
  - Natural language
  - Counter-example
  - Causal
  - CNN visualization

Bias reduction

- How to reduce bias?
- Better data
- Algorithmic bias reduction

Novel class of algorithms?

- Someone is already working on these?
- Links of interest:
    - Charles Isabell and Michael Littman talk at Neurips 2020
    - Ayanna Howard's course on AI, Ethics and Society

Summary

- ML/AI are important tools
- Can be used for good
- Can be used to do bad
- Important to consider the end-use of your algorithms
- Important to consider bias and overfitting issues
- Important to understand how your data collection will be used in future