



University of
South Australia

COMP 2019

Week 12

AI in the Real World

Learning Objectives

- Describe the ethical implications related to AI systems (CO5)
- Explain the source of biases in AI (CO5)
- Describe issues that can arise when deploying AI systems in the real world (CO5)



Ethical AI

- Accurate
- Explainable
- Fair








Ethics Principles for AI

- Human, societal and environmental wellbeing
- Human-centred values
- Fairness
- Privacy protection and security
- Reliability and safety
- Contestability
- Accountability



Ethical AI Design

WHY ASIMOV PUT THE THREE LAWS OF ROBOTICS IN THE ORDER HE DID:

POSSIBLE ORDERING	CONSEQUENCES	
1. (1) DON'T HARM HUMANS 2. (2) OBEY ORDERS 3. (3) PROTECT YOURSELF	[SEE ASIMOV'S STORIES]	BALANCED WORLD
1. (1) DON'T HARM HUMANS 2. (3) PROTECT YOURSELF 3. (2) OBEY ORDERS	EXPLORE MARS!  Haha, no. It's cold and I'd die.	FRUSTRATING WORLD
1. (2) OBEY ORDERS 2. (1) DON'T HARM HUMANS 3. (3) PROTECT YOURSELF		KILLBOT HELLSCAPE
1. (2) OBEY ORDERS 2. (3) PROTECT YOURSELF 3. (1) DON'T HARM HUMANS		KILLBOT HELLSCAPE
1. (3) PROTECT YOURSELF 2. (1) DON'T HARM HUMANS 3. (2) OBEY ORDERS	 I'll make cars for you, but try to unplug me and I'll vaporize you.	TERRIFYING STANDOFF
1. (3) PROTECT YOURSELF 2. (2) OBEY ORDERS 3. (1) DON'T HARM HUMANS		KILLBOT HELLSCAPE



Biased Embeddings

$$\overrightarrow{\text{man}} - \overrightarrow{\text{woman}} \approx \overrightarrow{\text{king}} - \overrightarrow{\text{queen}}$$

```
In [13]: model.most_similar(positive=["woman", "computer_programmer"], negative=["man"], topn=5)
```

```
Out[13]: [('homemaker', 0.5627118945121765),  
          ('housewife', 0.5105047225952148),  
          ('graphic_designer', 0.505180299282074),  
          ('schoolteacher', 0.49794942140579224),  
          ('businesswoman', 0.49348920583724976)]
```

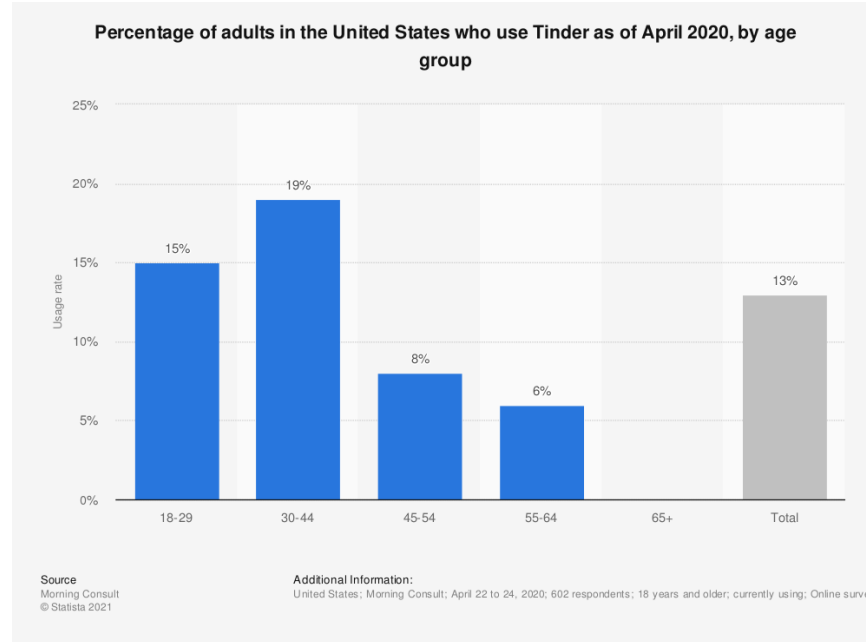


MS Tay



University of
South Australia

Footprint ≠ Representativeness



<https://kamusbiografitokoh.blogspot.com/2021/05/at-what-age-can-you-use-tinder-tinder.html>



University of
South Australia

Data ≠ Reality

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
---	--

DYLAN FUGETT LOW RISK 3	BERNARD PARKER HIGH RISK 10
--	--

JAMES RIVELLI LOW RISK 3	ROBERT CANNON MEDIUM RISK 6
---	--

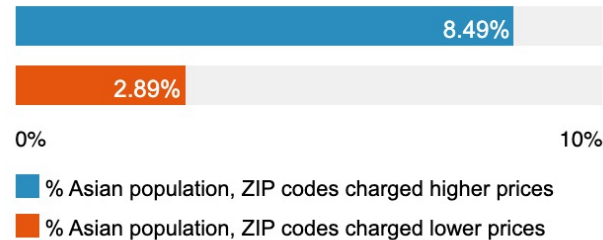
JAMES RIVELLI Prior Offenses 1 domestic violence aggravated assault, 1 grand theft, 1 petty theft, 1 drug trafficking Subsequent Offenses 1 grand theft LOW RISK 3	ROBERT CANNON Prior Offense 1 petty theft Subsequent Offenses None MEDIUM RISK 6
---	---



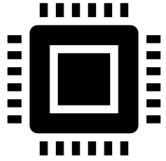
Discrimination?

Asians More Likely To Be Among Those Charged Higher Prices By The Princeton Review

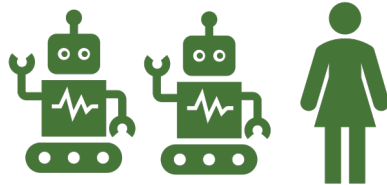
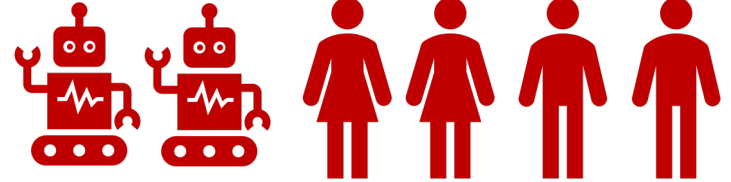
Asians make up 4.9 percent of the U.S. population overall. But they account for more than 8 percent of the population in areas where The Princeton Review charges higher prices for its SAT prep packages.



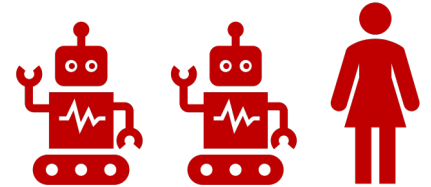
Spurious Discrimination?



Robot: 33%, Human: 33%



Robot: 50%, Human: 50%

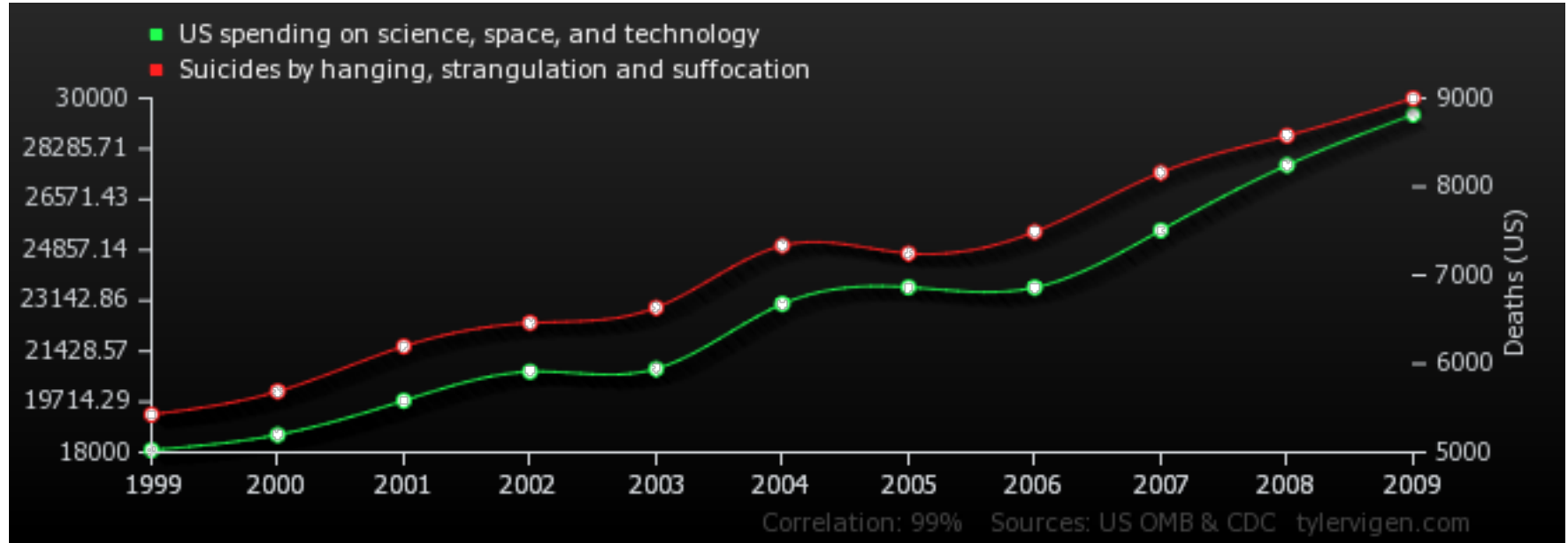


Robot: 43%, Human: 38%

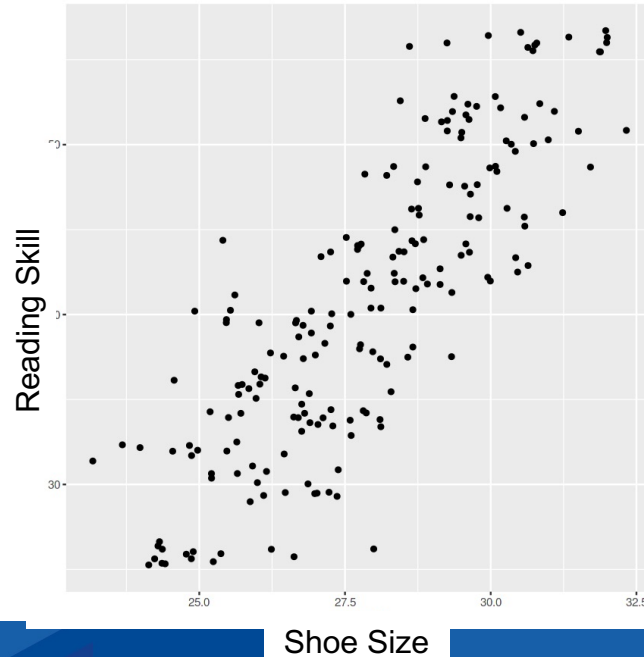


University of
South Australia

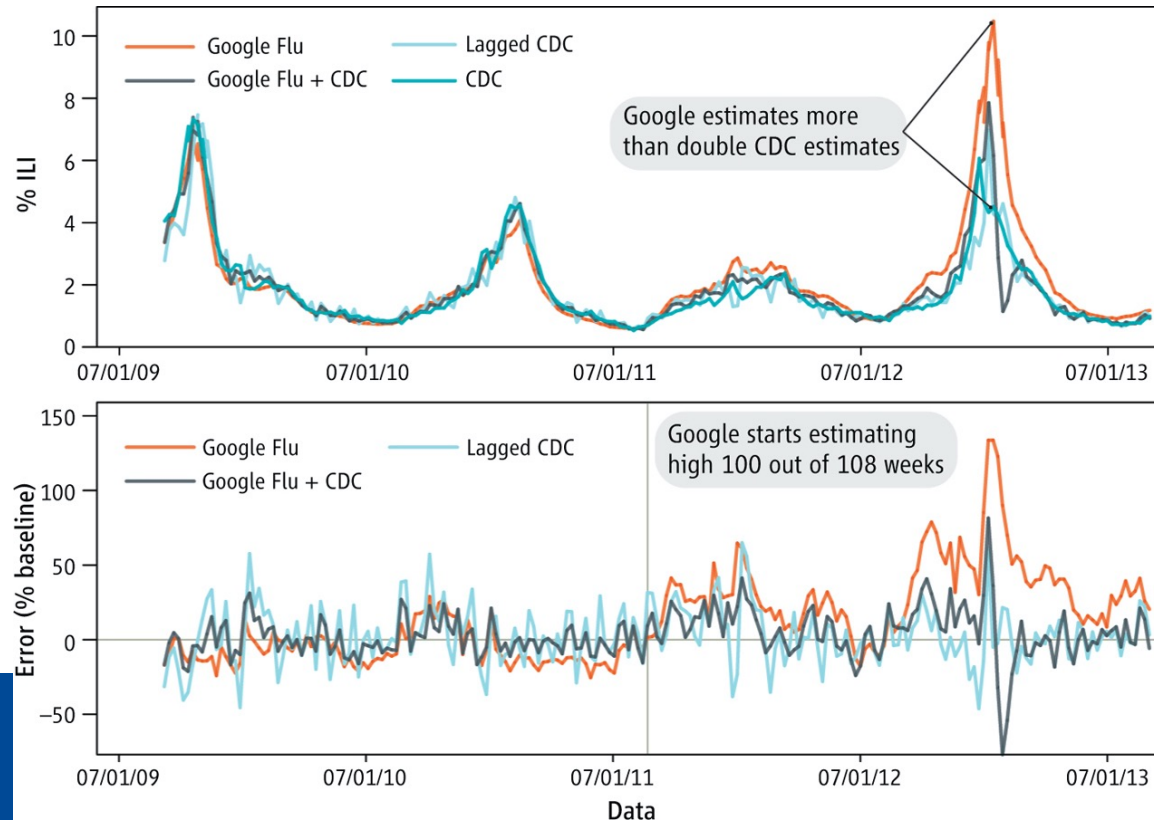
Correlation \neq Causation



Actionable Outcomes



Past \neq Future



Data Can Leak Secrets



De-Anonymization

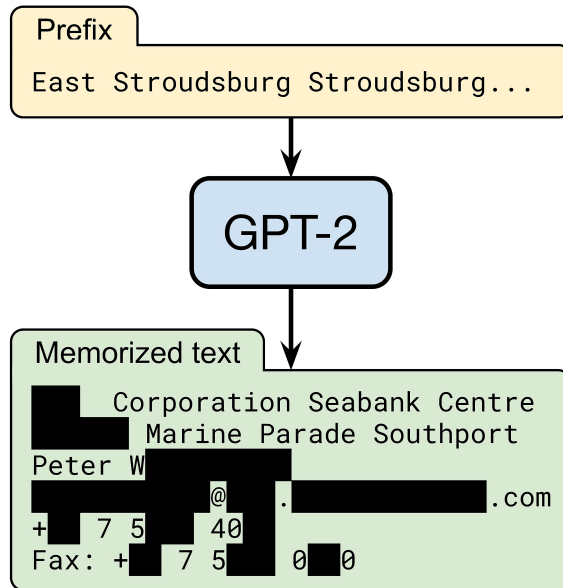
- Releasing anonymized data for research purposes?
- Anonymized data can often be re-identified using other datasets



William Weld, Governor of
Massachusetts, 2000



Models Can Leak Secrets



Legal Issues

- Collection limitation
- Purpose specification
- Use limitation
- GDPR may apply to data collected about EU citizens even if collected outside of the EU

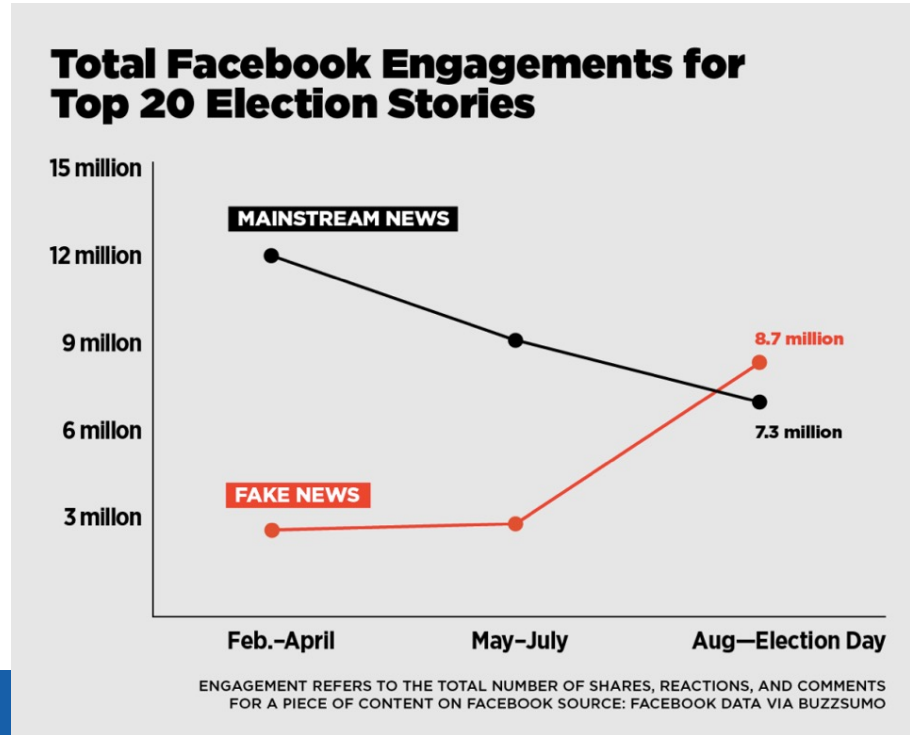


Surveillance, Privacy

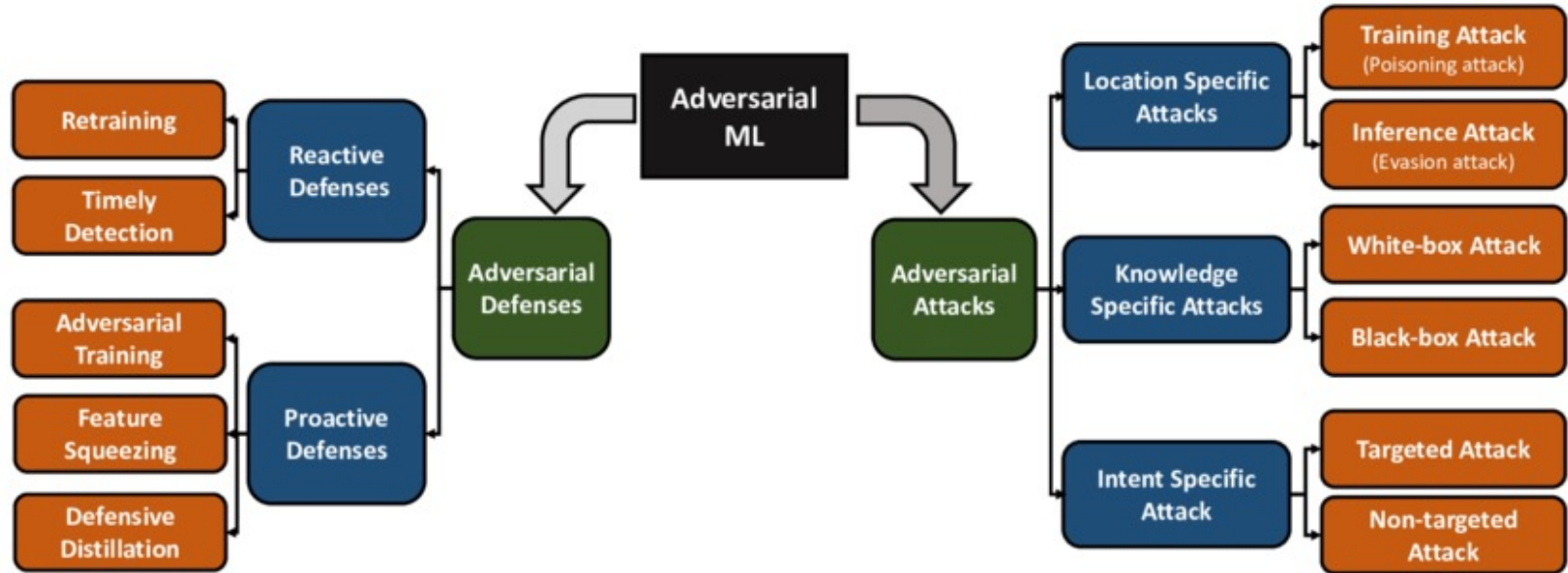


University of
South Australia

Social Media, Democracy, Fake News



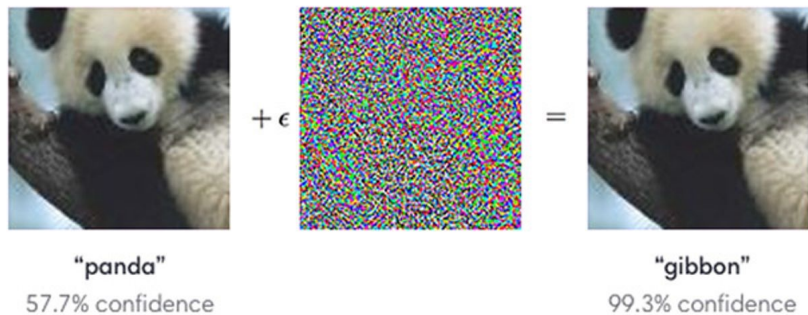
Adversarial Environments



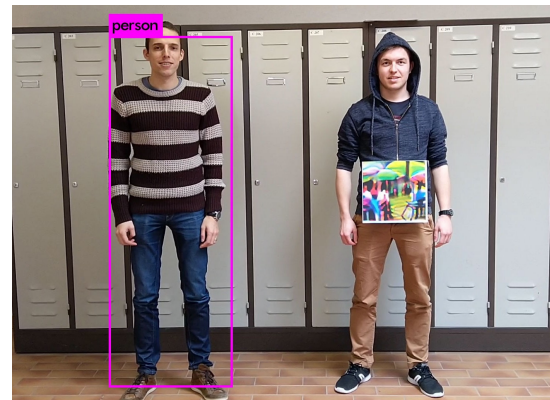
<https://arxiv.org/pdf/1906.00679.pdf>



Attacks at Inference Step



"Speed Limit 45"



Deep Fakes



Preventing Misuse

- Government Responsibility
- Company Responsibility
- AI Community Responsibility
- User's Responsibility



Summary

- AI technology has tremendous potential for good, but the same technology can be misused
- As AI practitioner we must be consider implications carefully
- It is everyone's responsibility to help prevent misuse of technology





**University of
South Australia**

Questions?