



STOR 320 Data Ethics

Lecture 36

Yao Li

Department of Statistics and Operations Research

UNC Chapel Hill

Why Data Science?

- Fields
 - Math
 - Statistics
 - Computer Science
- On-Demand Job Market
 - LinkedIn Workforce Report: Shortage in All Cities (2018)
 - [Article](#) by Gregory Piatetsky
- Applicable Across All Areas
- Bigger Data Requires Better Skills
- People are Scared Because Analytics is Coming

Data Privacy



Data Privacy

- People Have Data and Need Us to Analyze the Information to Answer Their Questions
- Decisions Based on Our Analyses Can Effect Others
- We Must Operate Ethically
- Fairness and Privacy

Hippocratic Oath for Data Scientists

I solemnly pledge to practice my profession with conscience and dignity;

To respect the privacy of the people whose data is confided in me;

To maintain the utmost respect for the individuals whose data I am analyzing;

To be transparent, open, and honest about the type of analysis I am applying to their data;

To never use my knowledge to violate human rights and civil liberties, even under threat.

-- [Marie Wallace](#)

Data Fraud

- Appropriate Data Analysis
 - Let the Data Speak for Itself
 - Hypotheses Before Analyses
- Transparency
 - Data Imputation
 - Handling of Outliers
 - Data Transformations
 - Statistical Methods
 - Necessary Assumptions
 - Variables Included and Excluded

Data Fraud

- Recent Incident at Harvard (2018)
 - Dr. Piero Anversa
 - 31 Publications
 - Stem Cell Research
 - Showed the Heart Has the Ability to Regenerate
 - Fraudulent Data
 - [Article](#) by NY Times



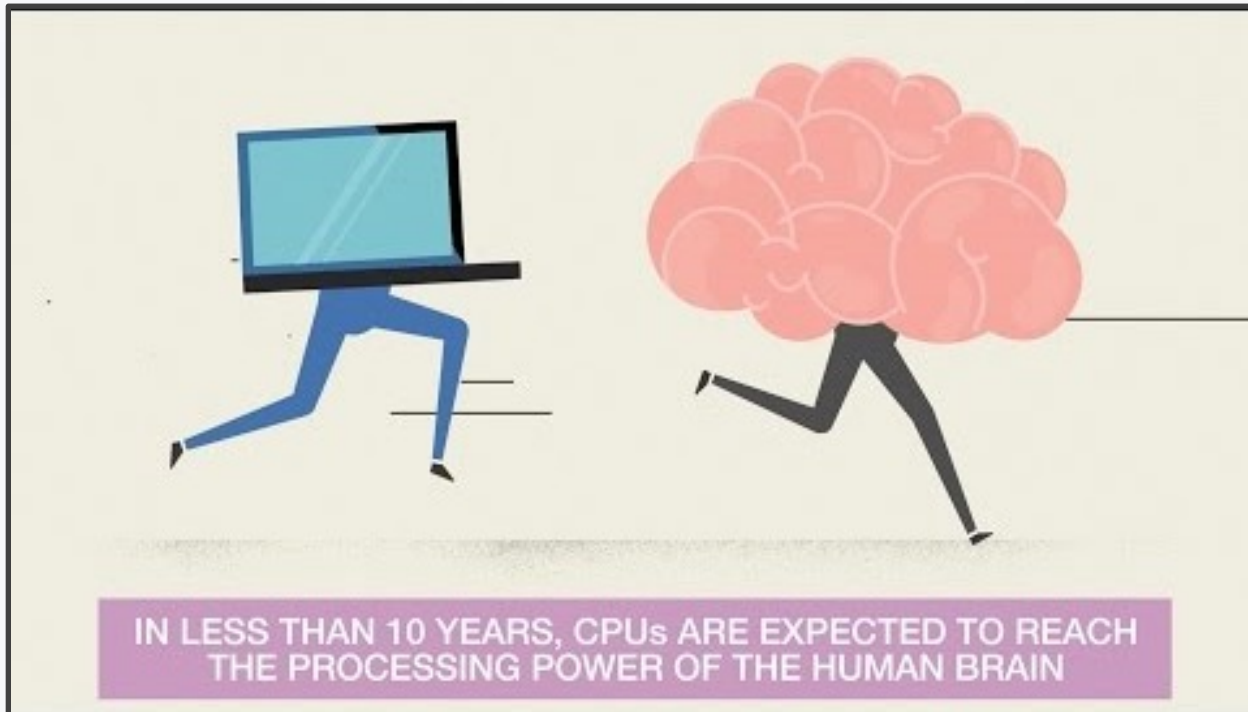
Data Fraud

- Thought Provoking Questions
 - Why Does This Happen?
 - Are You Capable of This?
 - How Does This Apply to You in Your Future Career?



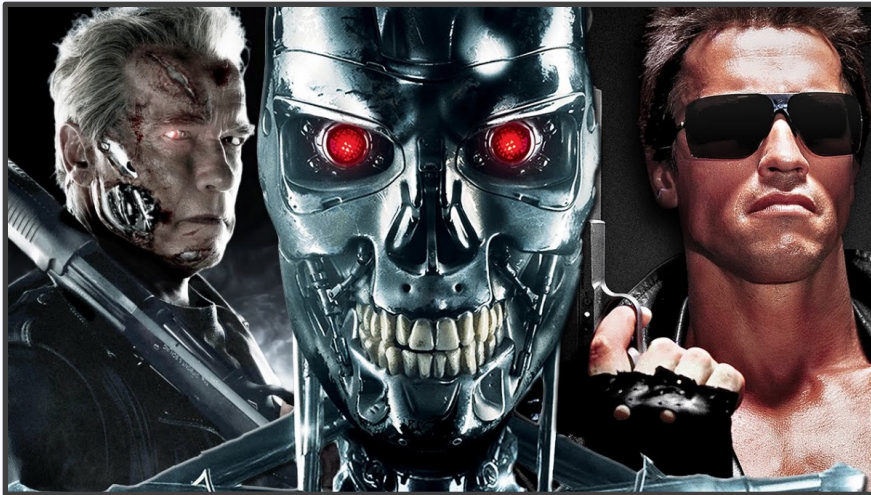
Data is Coming

- The Expansion of Data



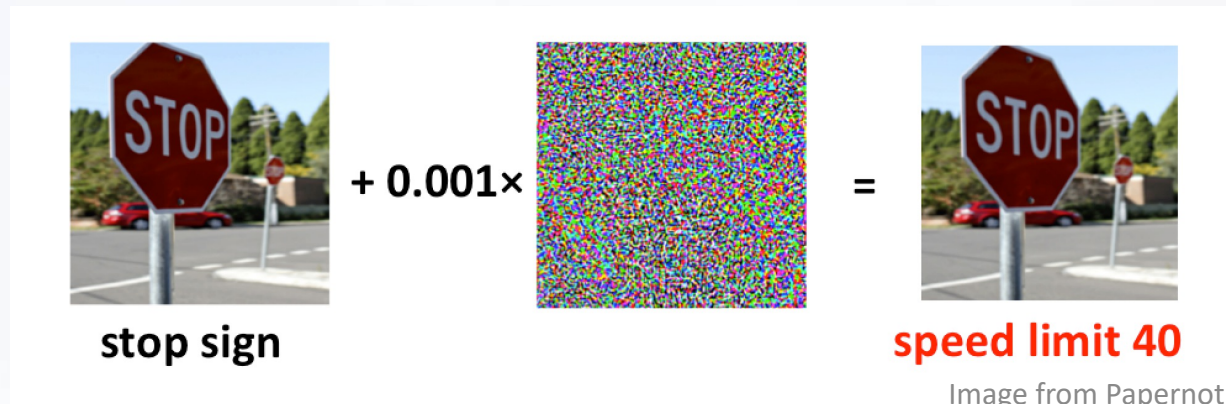
Data is Coming

- How About Now?



ML System is not Reliable

- Fool self-driving system with carefully crafted noise



- Invisible T-shirt

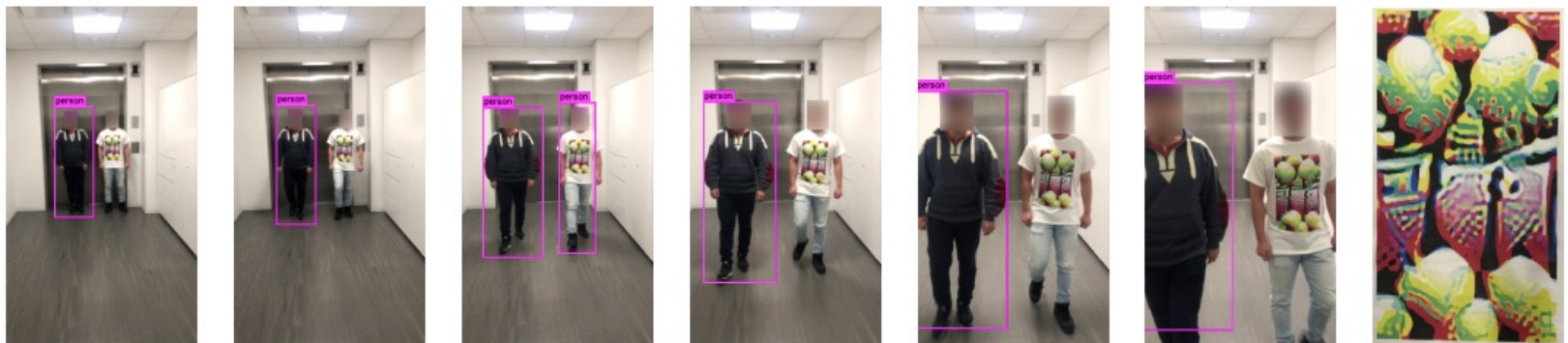


Image from Xu, Kaidi, et al. "Adversarial t-shirt! evading person detectors in a physical world."

Values of Data

- How Much is Data Worth?
 - In 2016, Microsoft Purchased LinkedIn for \$26.2 Billion
 - In 2016, Caesar's Total Rewards Program Was Valued at \$1 Billion During Bankruptcy
 - [Article](#) in MIT Sloan Magazine

Values of Data

- Thought Provoking Questions
 - Are You Worried About Data Privacy?
 - Regulation?



Modern Data Ethics

- Weapons of Math Destruction
 - Author: Cathy O'Neil
 - Blog: mathbabe.org
 - Worked on Wall Street
 - Saw the Dark Side of Using Algorithms on Big Data to Make Business Decisions
 - Leader in Data Ethics



Thank You!