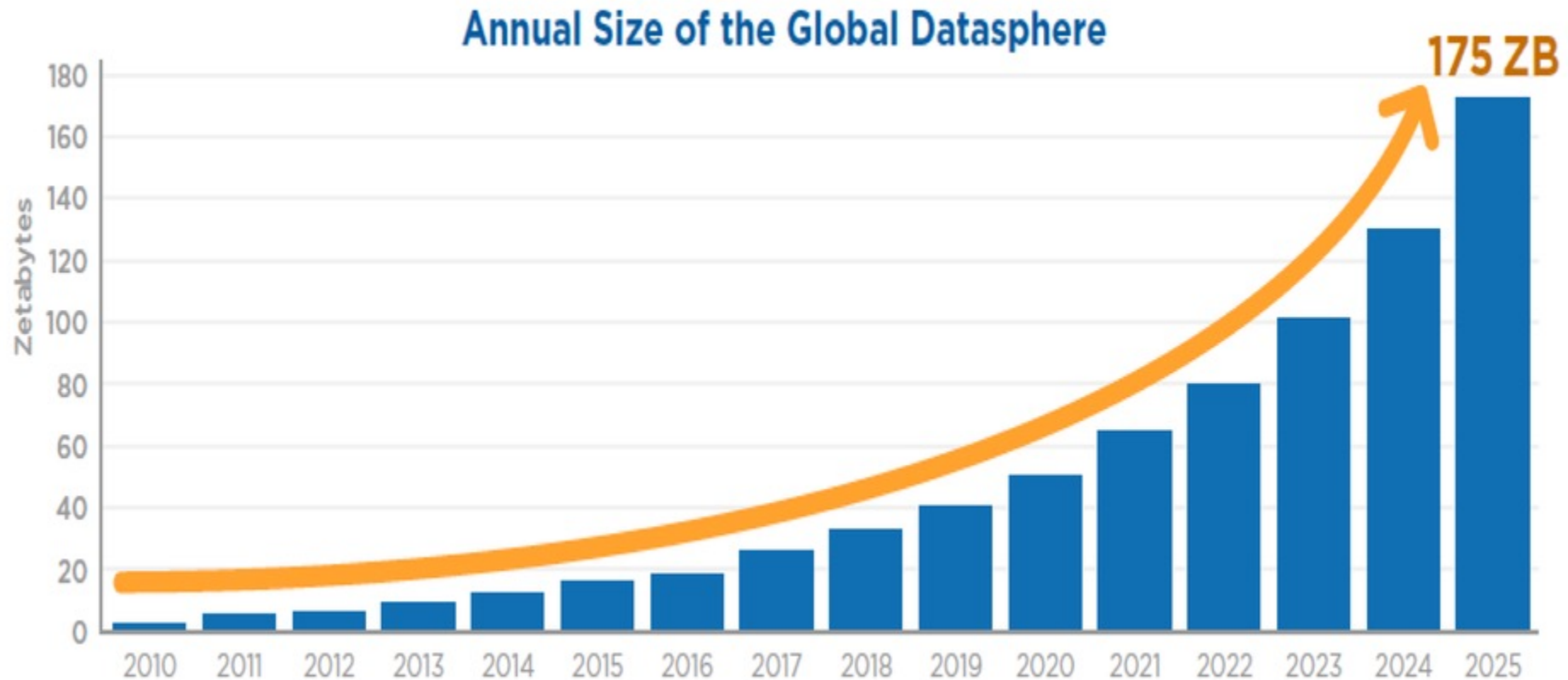# Data

Akitaka Matsuo

Department of Government

Data

- Data is everywhere
    - Government statistics
    - Internet
        - Social media text and graphics
        - Trace data
            - GPS
            - IoT
        - Protest

University of Essex

Growth of Data

Annual Size of the Global Datasphere

175 ZB

Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

Source: https://www.forbes.com/sites/tomcoughlin/2018/11/27/175-zettabytes-by-2025/

University of Essex

What is Big Data

- Three V's of Big Data
  - **Volume:** SNS posts, IoT devices
  - **Velocity:** Fast growing data, timely incorporation in analysis
  - **Variability:** Textual, still image, video, network

- In this course we learn how to work with "data"

University
of Essex

Example: Covid-19 and misinformation

HEALTHCARE • CORONAVIRUS

Did Covid-19 Vaccines Cause Coronavirus Delta Variants? Here's What The Timing Says

Bruce Y. Lee Senior Contributor ⓘ

I am a writer, journalist, professor, systems modeler, computational and digital health expert, avocado-eater, and entrepreneur, not always in that order.

Follow

Aug 7, 2021, 12:56pm EDT

- We can trace how information spread
  - Who shared misinformed tweets?
  - What do they tweet?
  - Whom do they follow?
  - etc

University of Essex

But we need to be aware the nature of data

Mitigating Gender Bias in Natural Language Processing: Literature Review

Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, William Yang Wang

📄 PDF　　66 Cite　　🔍 Search　　🎥 Video

Abstract

As Natural Language Processing (NLP) and Machine Learning (ML) tools rise in popularity, it becomes increasingly vital to recognize the role they play in shaping societal biases and stereotypes. Although NLP models have shown success in modeling various applications, they propagate and may even amplify gender bias found in text corpora.

- The machine learning could lead to the reproduction of existing discrimination
  - Employability
  - Crime profiling
  - Educational attainment
- What could have done to prevent the problem?
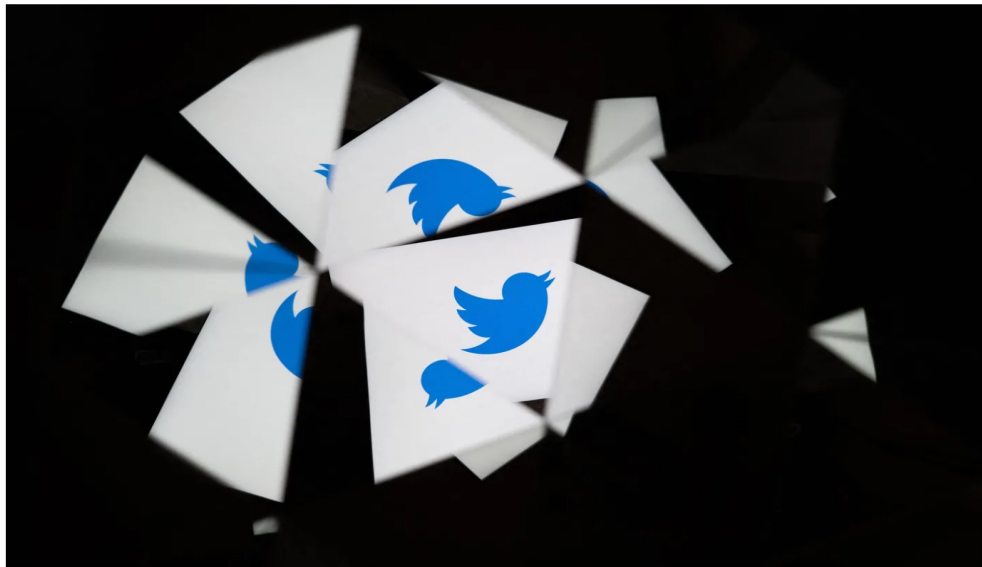 - *The Alignment Problem,* Brian Christian, 2020, Norton

University of Essex

Also, Twitter is gone

Twitter just closed the book on academic research / Twitter was once an indispensable resource for academic research. That's changed under Elon Musk.

By **Justine Calma**, a science reporter covering the environment, climate, and energy with a decade of experience. She is also the host of the Hell or High Water podcast.
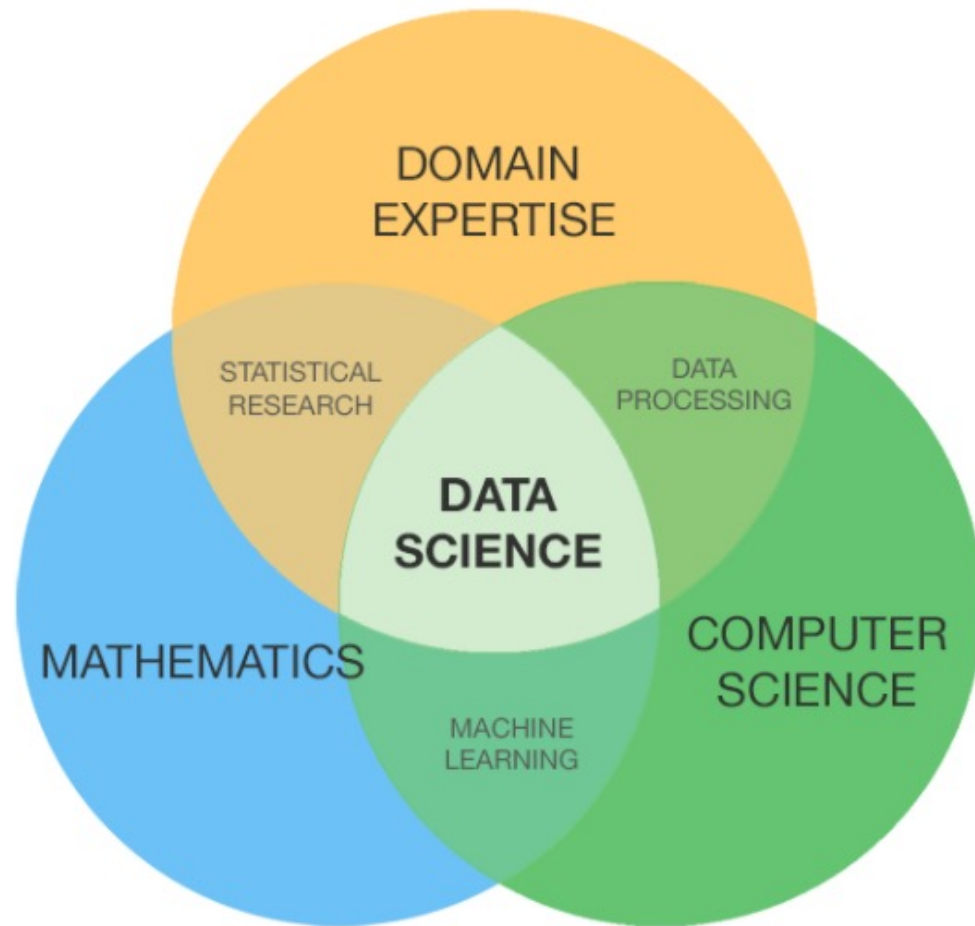
May 31, 2023, 2:19 PM GMT+1 | 💬 23 Comments / 23 New

The twitter's logo is pictured on screen reflected by mirrors in Mulhouse, eastern France on May 30, 2023. Photo by SEBASTIEN BOZON/AFP via Getty Images

Twitter was once a mainstay of academic research — a way to take the

- The data is there, but its accessibility sometimes relies on the discretion of the company
- Misinformation is harmful, but we may or may not be able to know if the data are not accessible.

University of Essex

Data science

- Data science is the combination of three
- We are social scientist and there should be something we can do the best