

Democratization of Data Science

Anil Jain

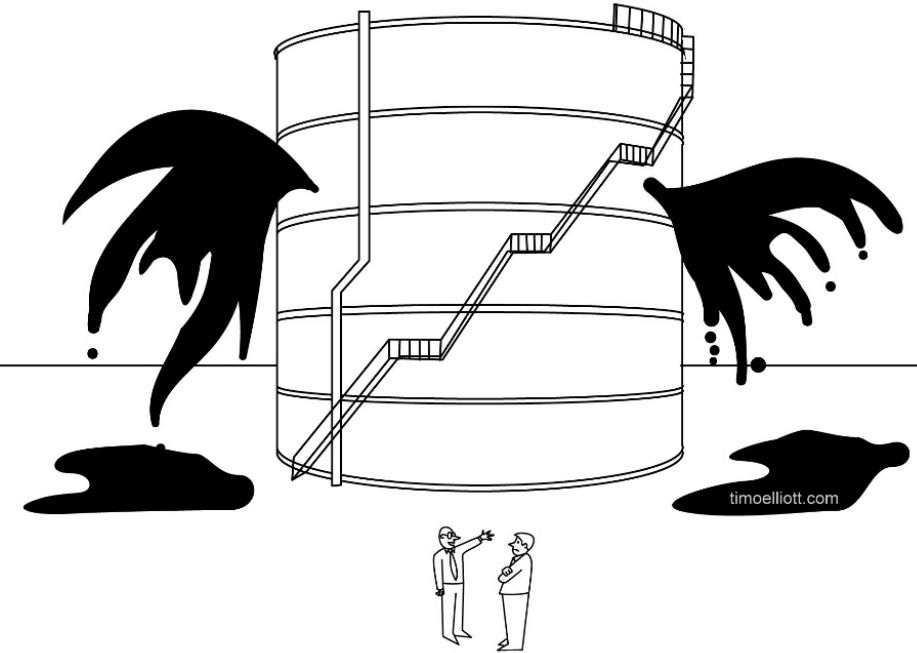
- ***Disperse data science expertise***
- ***Science:*** Easily accessible
- ***Data access:*** Privacy issues, ownership, user consent, data integrity, function creep,....



“It’s free, but they sell your information.”

Personally Identifiable Information (PII)

- Information that can be used to **trace an individual's identity**¹
- **GDPR**² also covers physical, physiological identity, location, online identifier,...
- **Privacy of anonymized data?**³



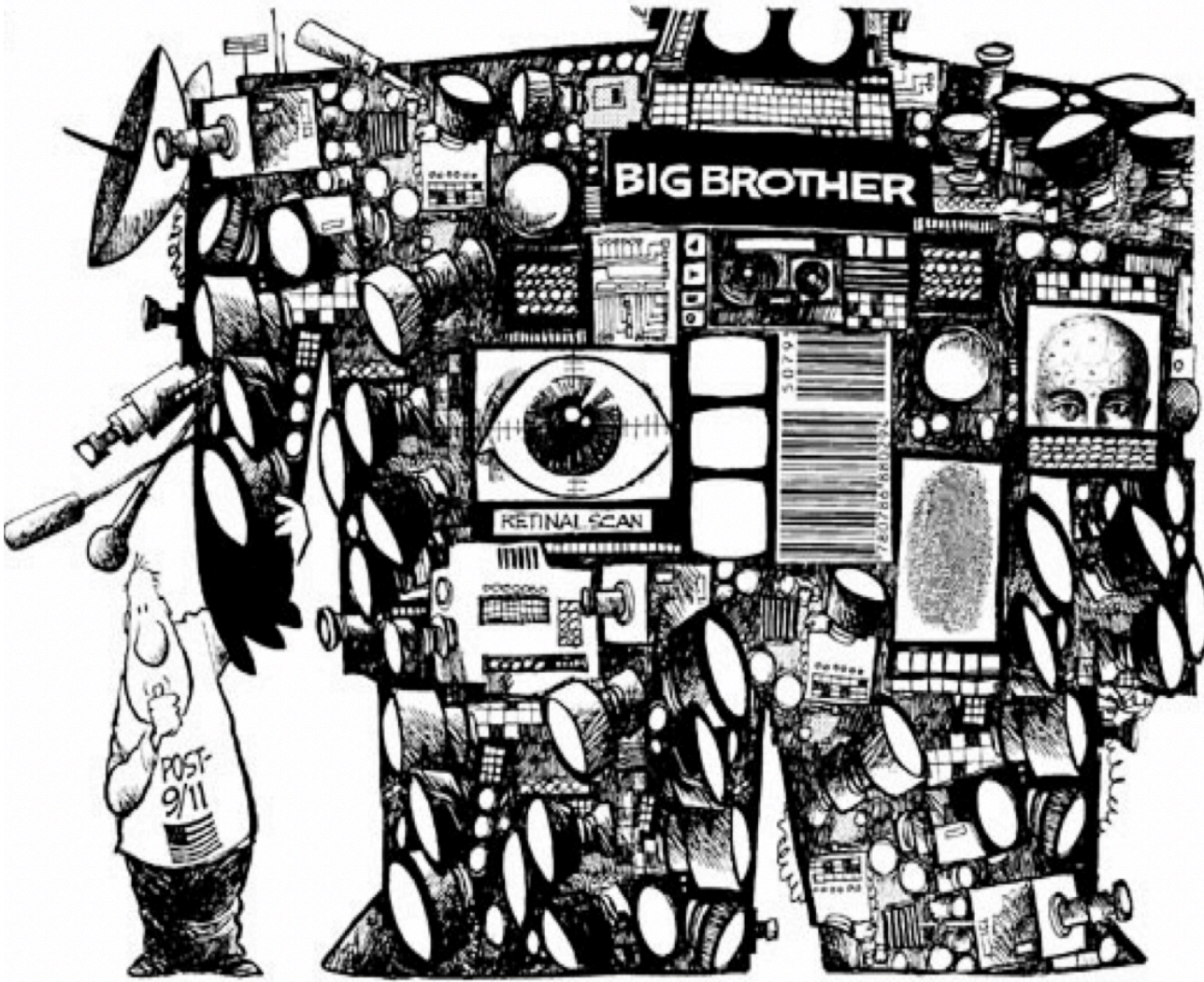
“Data is the new oil? Absolutely – toxic if mishandled!...”

¹OMB Memorandum M-07-1616

² General Data Protection Regulation (<https://eugdpr.org>)

³ “Algorithms can identify 99.98% of Americans with as few as 15 attributes, such as gender, ZIP code or marital status”, **Nature Communications**, 23 July, 2019

Who Can You Trust with Data?



Google Fined **\$22M** (2012)

Facebook Fined **\$5bn** (2019)

SF Bans Face Recognition (2019)

A Hacker Accessed **100M Capital One**
Credit Card Accounts (2019)

Equifax Data Breach Exposing 147M Personal
Info Settlement for **\$425M** (2019)

3 Massachusetts hospitals fined **\$1M**
for HIPAA violations (2019)

Privacy v. Security



A.I. Enhances Health Care, But Is Privacy the Cost? NY Times, August 1, 2019