Data and Ethics

Ritika Bhasker, *Data Scientist* at *DevResults*

@mostlyinane

datawrites.co

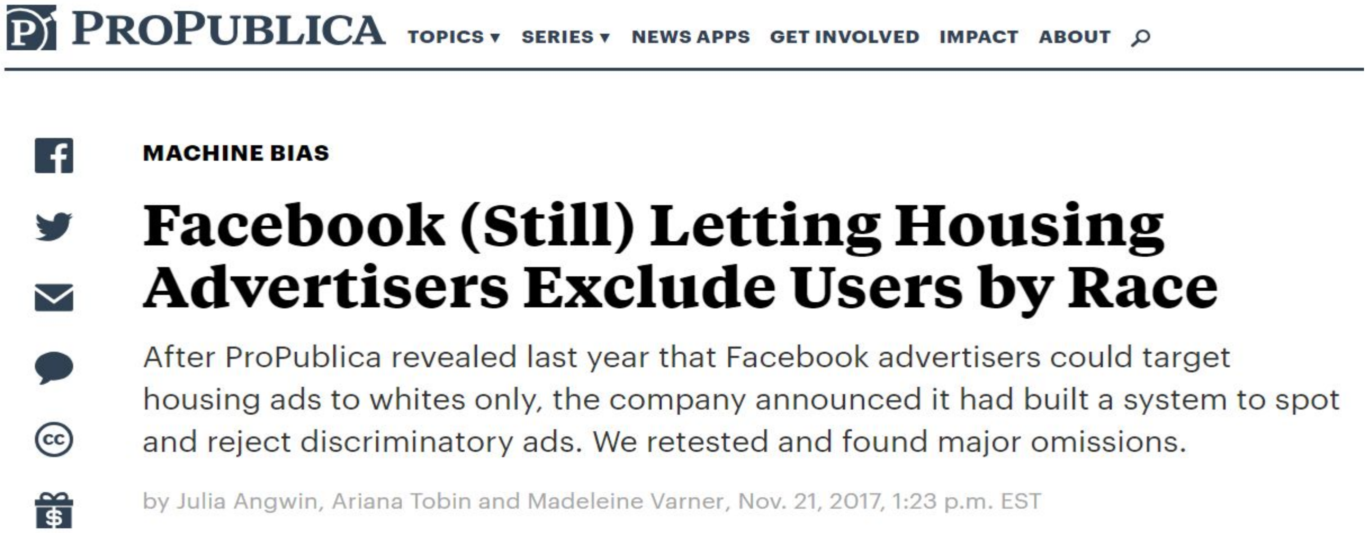
Why does it matter?

- Machine learning is being used to make critical decisions that affect people’s lives
- To make better decisions, everyone is collecting huge amounts of data
- Decisions made using algorithms aren’t often apparent, or well communicated, to the public
- We’re seeing a growing number of data leaks and people are wary about data collection, data retention, and algorithmic decision making.

Why does it matter?

- Fairness:
  - Do our conclusions disproportionately affect certain groups?
- Representation:
  - Does our analysis reinforce existing biases?
- Privacy:
  - What data are we collecting?
- Communication:
  - Does everyone -- from decision makers to those affected by decisions made -- know and understand what just happened?

Case studies 1:



MACHINE BIAS

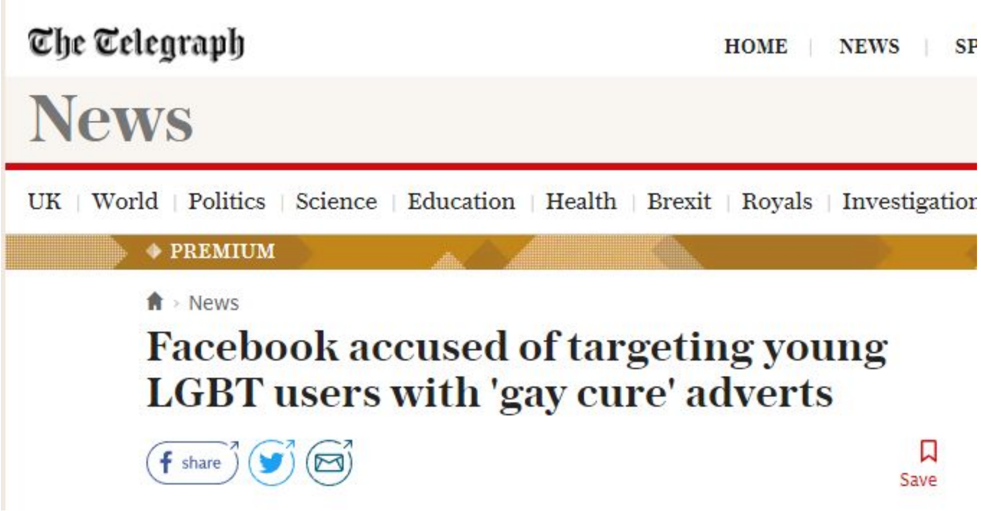
Facebook (Still) Letting Housing Advertisers Exclude Users by Race

After ProPublica revealed last year that Facebook advertisers could target housing ads to whites only, the company announced it had built a system to spot and reject discriminatory ads. We retested and found major omissions.

by Julia Angwin, Ariana Tobin and Madeleine Varner, Nov. 21, 2017, 1:23 p.m. EST

- Why is this problematic?
- How can we fix it?

Case Study 2:



The Telegraph

HOME | NEWS | SP

News

UK | World | Politics | Science | Education | Health | Brexit | Royals | Investigatior

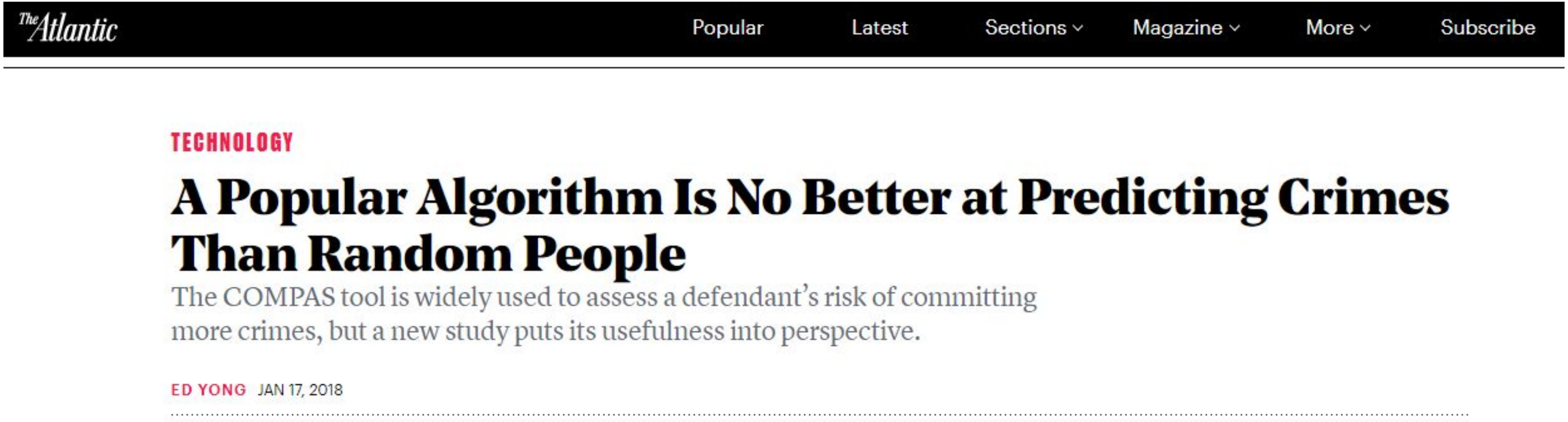
◆ PREMIUM

News

Facebook accused of targeting young LGBT users with 'gay cure' adverts

- Does Facebook have a duty to prevent this?
- How would you prevent this?

Case Study 3:



TECHNOLOGY

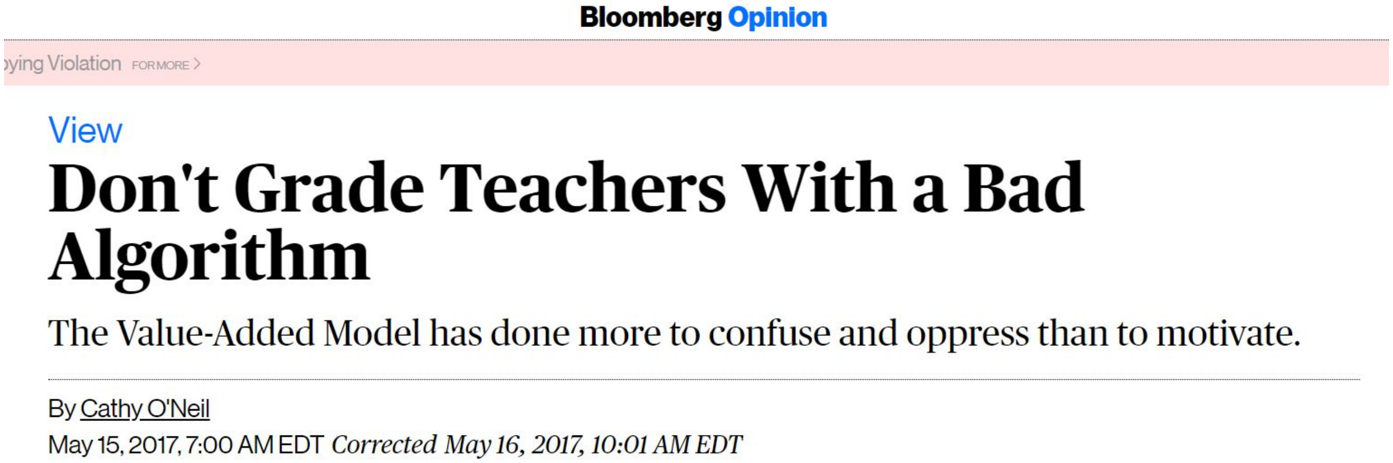
A Popular Algorithm Is No Better at Predicting Crimes Than Random People

The COMPAS tool is widely used to assess a defendant's risk of committing more crimes, but a new study puts its usefulness into perspective.

ED YONG JAN 17, 2018

- Should law enforcement/judges continue using the COMPAS tool?
- How can we make the algorithm perform better?

Case Study 4:



Bloomberg Opinion

ying Violation FOR MORE >

View

Don't Grade Teachers With a Bad Algorithm

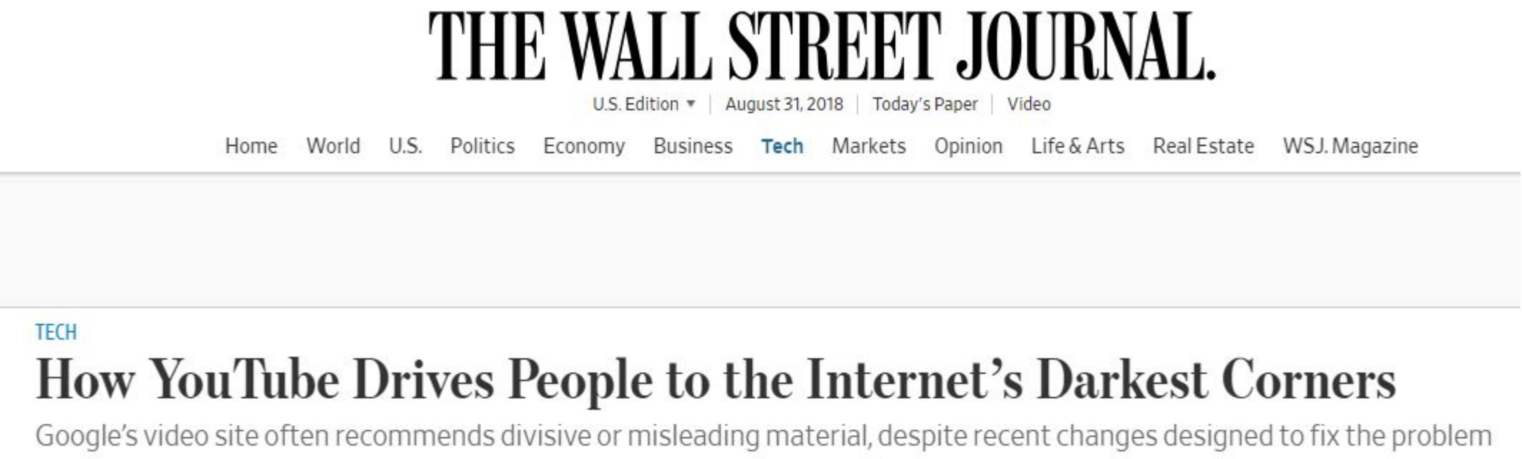
The Value-Added Model has done more to confuse and oppress than to motivate.

By Cathy O'Neil

May 15, 2017, 7:00 AM EDT *Corrected May 16, 2017, 10:01 AM EDT*

- How would you fix the algorithm?
- Should we be using machine learning here?

Case Study 5:



THE WALL STREET JOURNAL.

U.S. Edition ▾ | August 31, 2018 | Today's Paper | Video

Home World U.S. Politics Economy Business Tech Markets Opinion Life & Arts Real Estate WSJ. Magazine

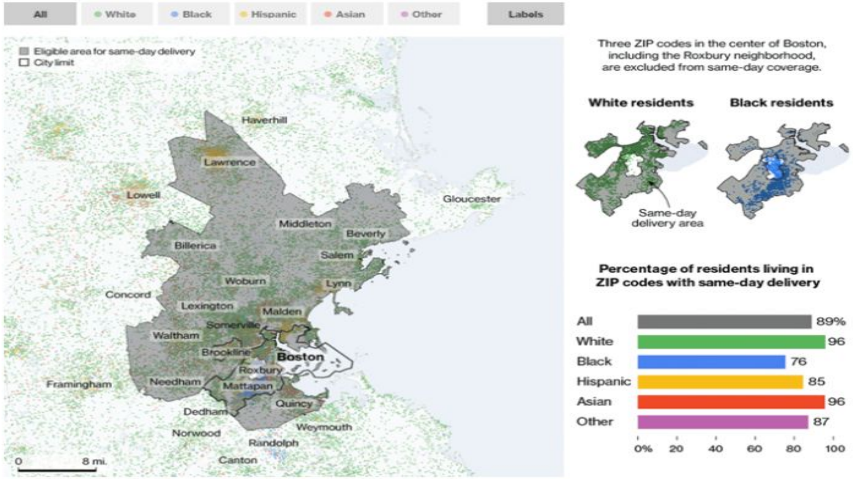
TECH

How YouTube Drives People to the Internet's Darkest Corners

Google's video site often recommends divisive or misleading material, despite recent changes designed to fix the problem

- Is this a problem YouTube has a duty to fix?
- How would you fix this?

Case Study 6:



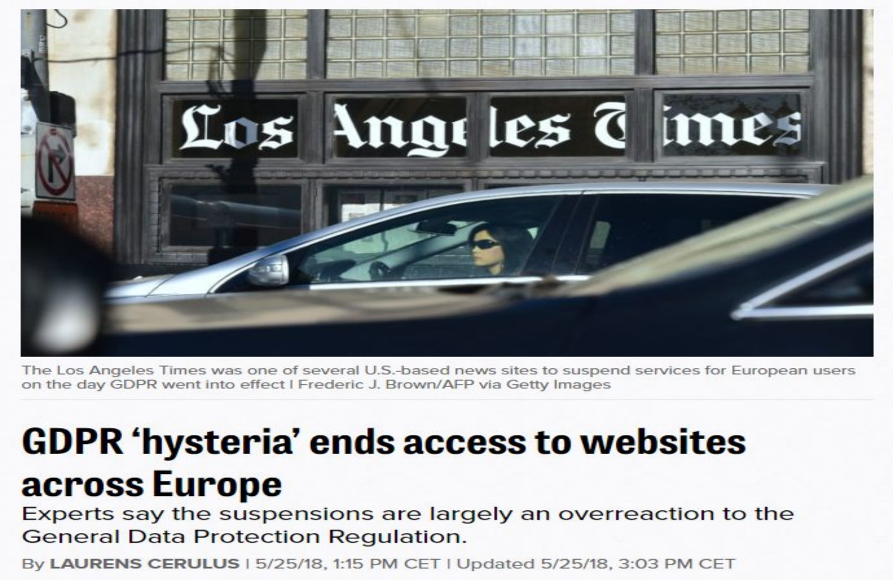
Three ZIP codes in the center of Boston, including the Roxbury neighborhood, are excluded from same-day coverage.

Percentage of residents living in ZIP codes with same-day delivery

MAP COMPARING DELIVERY AREA AGAINST DEMOGRAPHICS IN BOSTON. MAP CREDIT: BLOOMBERG, 2016

- What’s the problem here?
- How do we fix it?

Case Study 7:



The Los Angeles Times was one of several U.S.-based news sites to suspend services for European users on the day GDPR went into effect | Frederic J. Brown/AFP via Getty Images

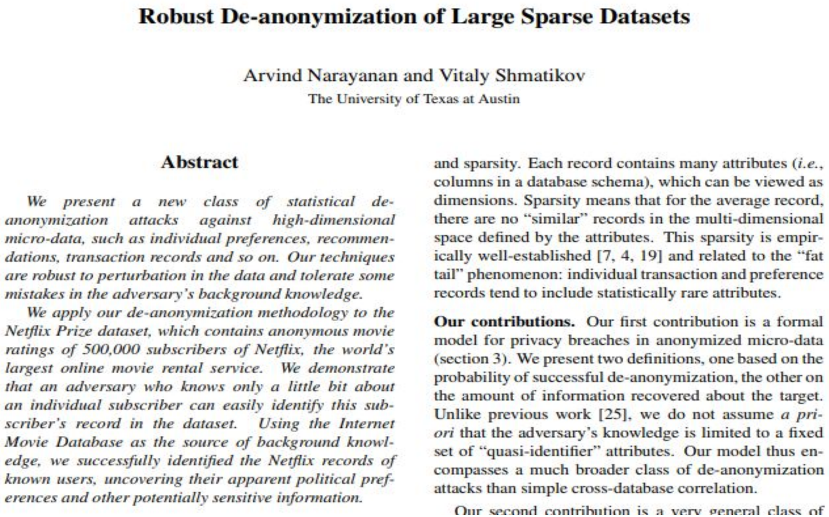
GDPR 'hysteria' ends access to websites across Europe

Experts say the suspensions are largely an overreaction to the General Data Protection Regulation.

By **LAURENS CERULUS** | 5/25/18, 1:15 PM CET | Updated 5/25/18, 3:03 PM CET

- Is this a problem?
- Why is this a problem?

Case Study 8:



Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Abstract

We present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge.

We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

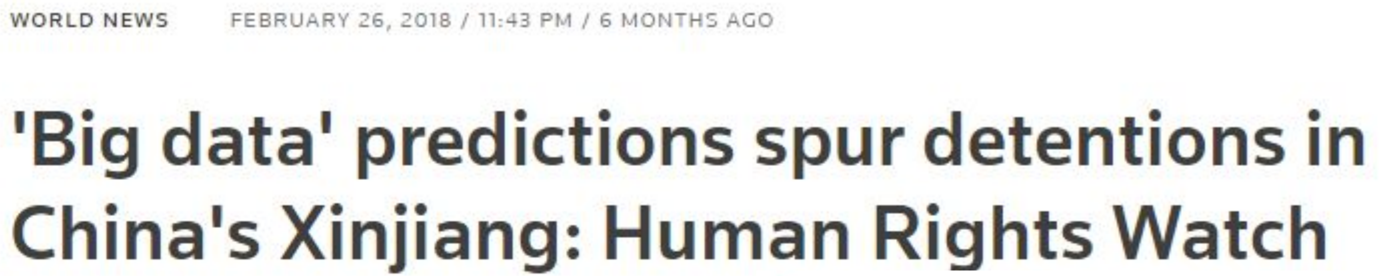
and sparsity. Each record contains many attributes (*i.e.*, columns in a database schema), which can be viewed as dimensions. Sparsity means that for the average record, there are no "similar" records in the multi-dimensional space defined by the attributes. This sparsity is empirically well-established [7, 4, 19] and related to the "fat tail" phenomenon: individual transaction and preference records tend to include statistically rare attributes.

Our contributions. Our first contribution is a formal model for privacy breaches in anonymized micro-data (section 3). We present two definitions, one based on the probability of successful de-anonymization, the other on the amount of information recovered about the target. Unlike previous work [25], we do not assume *a priori* that the adversary's knowledge is limited to a fixed set of "quasi-identifier" attributes. Our model thus encompasses a much broader class of de-anonymization attacks than simple cross-database correlation.

Our second contribution is a very general class of

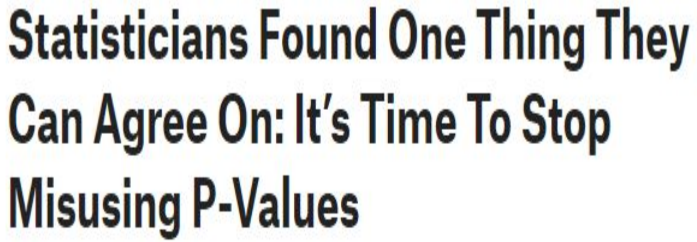
- Where do we draw the line with anonymity?

And then there's:

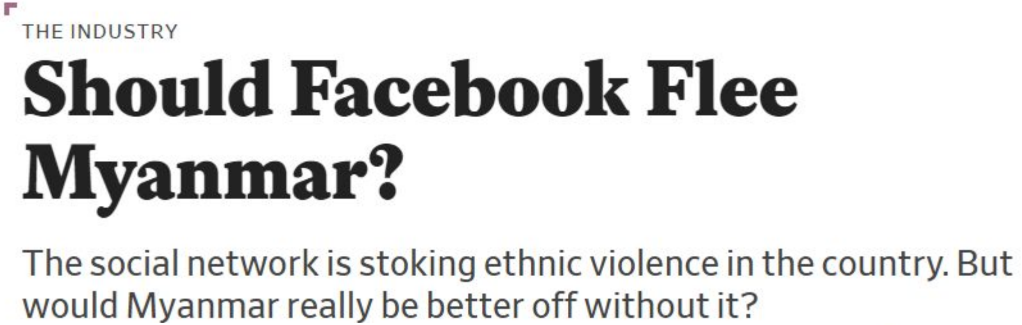


WORLD NEWS FEBRUARY 26, 2018 / 11:43 PM / 6 MONTHS AGO

'Big data' predictions spur detentions in China's Xinjiang: Human Rights Watch



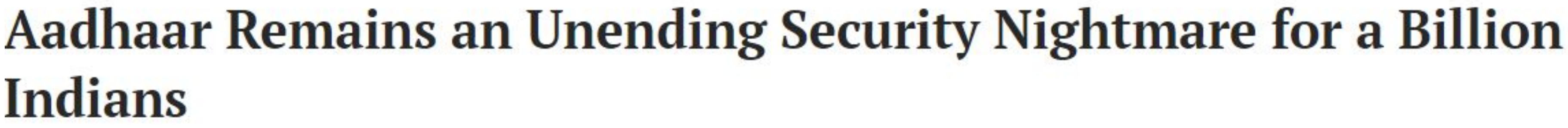
Statisticians Found One Thing They Can Agree On: It's Time To Stop Misusing P-Values



THE INDUSTRY

Should Facebook Flee Myanmar?

The social network is stoking ethnic violence in the country. But would Myanmar really be better off without it?



Aadhaar Remains an Unending Security Nightmare for a Billion Indians

What approaches could data adopt?

- Human rights approach:
  - ICCPR
  - UN
- Medical approach:
  - Hippocratic oath
  - Ethics review board
- Behavioural/Social science approach:
  - IRBs and consent-based approaches
- Journalistic approach:
  - Norms-based approach

Are there laws against this kind of thing?

- GDPR
  - Will be adopted by other governments, too
- California Consumer Privacy Act
  - The most important privacy regulation in the U.S.
- Intellectual Property
- HIPAA
- What else?

What's the takeaway?

- Always question the data
  - Who created the dataset?
  - How was the data generated?
- What data are you collecting?
  - Is the data randomly collected? Is it representative?
- What features are you retaining and what are you discarding?
- What model was chosen and why? What model evaluation metric was chosen and why?
- Who is going to use the algorithm (both anticipated and unanticipated)?
- Is your data secure? Can PII be revealed in any way?

Further reading:

- Weapons of Math Destruction - Cathy O’Neil
- Internet of Garbage - Sarah Jeong
- Twitter and Teargas - Zeynep Tufekci
- Critical questions for big data - danah boyd and Kate Crawford
- Big Data, Machine Learning, and the Social Sciences - Hanna Wallach
- Ethical Guidelines for Statistical Practice - Committee on Professional Ethics of the American Statistical Association
- Journalism as a Professional Model for Data Science - Brian Keegan
- Data Science Ethical Framework - UK Cabinet