Actionable Ethics for Data Scientists

DRIVENDATA

Jay Qi
Senior Data Scientist
jay@drivendata.org
@jayyqi

Casey Fitzpatrick, PhD
Machine Learning Engineer
casey@drivendata.org
@caseyfitz

#ODSC

ODSC West • October 29, 2019

Who are we?

DRIVENDATA

Data Science + Social Impact

https://www.drivendata.org • @drivendataorg

Data Science Competitions • Direct Client Engagements • Open Source Projects

https://github.com/drivendataorg

Competition Winners' Code
Cookiecutter Data Science
deon Data Science Ethics Checklist

Agenda

- Introduction (1 hr)
  - Why, What, and How of Data Ethics
  - Checklists and **deon**
  - Learning by Example
- Break (15 min)
- Group Activity: Qualitative Case Study (1 hr)
- Break (15 min)
- Hands-on Coding Exercise: Eviction Data Case Study (1 hr)

Ethics is hard.

Ethics is hard.

There is no free lunch. Tradeoffs are inevitable.

No one right answer. Reasonable people can disagree.

Good intentions aren't enough. Must deal with unintended consequences.

We will not solve all ethical problems today.

We will talk about a **starting point** for incorporating ethics into **practical** data science work.

Why does data ethics matter?

THE WALL STREET JOURNAL.

Why Software Is Eating The World

By Marc Andreessen

August 20, 2011

CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM WIRED

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

The Economist

The world's most valuable resource is no longer oil, but data

Harvard Business Review

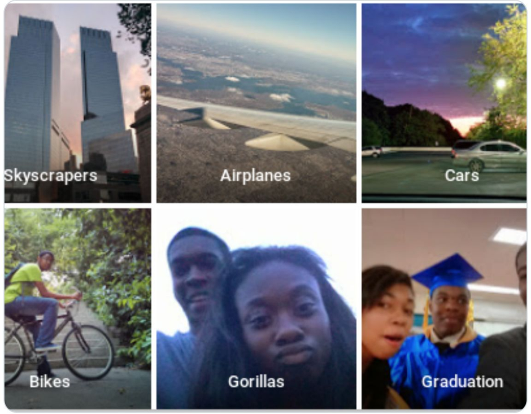
Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

jackyalcine is too young to be this tired
@jackyalcine

Google Photos, y'all f---ed up. My friend's not a gorilla.



6:22 PM · Jun 28, 2015 · Twitter Web Client

The Guardian

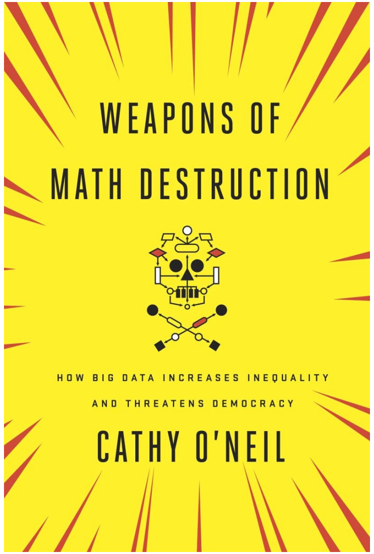
Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

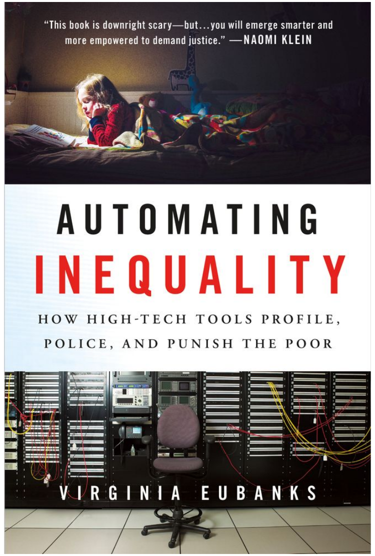
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica
May 23, 2016

ProPublica





How do we think about data ethics?

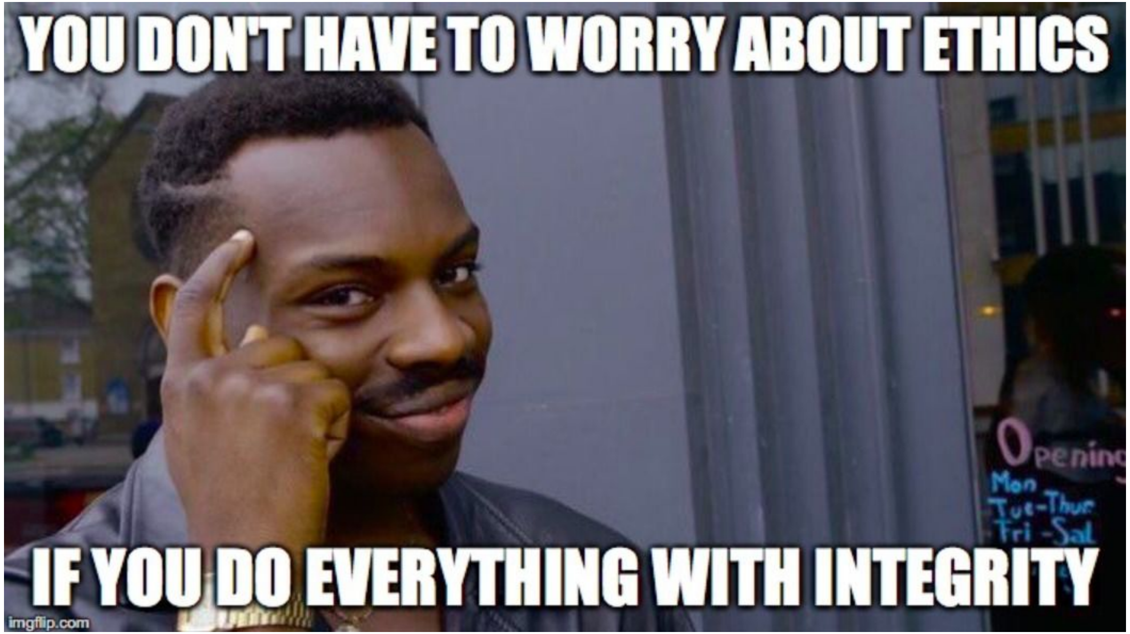
ethics /ˈɛθɪks/ *(plural noun)*

Moral principles that govern a person's behaviour or the conducting of an activity.

https://www.lexico.com/en/definition/ethics
Lexico.com by Oxford University Press, 2019

ethics

Moral principles that govern a person's behaviour or the conducting of an activity.



https://www.resonatepictures.com/friday-meme-integrity/

ethics

Moral principles that govern a person's behaviour or the conducting of an activity.

A set of foundational values and beliefs.

Forming a set of principles for data science in the form of a code or oath is an active area of discussion and development.

This is important area for us all to reflect on, but today’s workshop will not focus on this topic.

ethics

Moral principles that govern a person's behaviour or the conducting of an activity.

Examples of notable efforts in developing principles:

Community-driven

- ACM Code of Ethics (Association for Computing Machinery)
  https://ethics.acm.org/
- Ethical Guidelines for Statistical Practice (American Statistical Association)
  https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx
- Manifesto for Data Practices (data.world & Linux Foundation)
  https://datapractices.org/manifesto/
- Global Data Ethics Project (Data for Democracy)
  https://www.datafordemocracy.org/project/global-data-ethics-project

ethics

Moral principles that govern a person's behaviour or the conducting of an activity.

Examples of notable efforts in developing principles:

Corporate / Industry

- Google’s AI Principles
  https://www.blog.google/technology/ai/ai-principles/
- Microsoft’s AI Principles
  https://www.microsoft.com/en-us/ai/our-approach-to-ai
- Partnership on AI Tenets
  https://www.partnershiponai.org/tenets/

ethics

Moral principles that govern a person's behaviour or the **conducting of an activity**.

Once you have principles, how do you then apply them to your day-to-day practice?

The rest of this workshop will focus on helping you do this.

ethics

Moral principles that govern a person's behaviour or the **conducting of an activity**.

Deon is an open-source command line tool that allows you to easily add an ethics checklist to your data science projects.

deon.drivendata.org

Why an ethics *checklist*?

Inspired by previous work, especially Of Oaths and Checklists by Mike Loukides, Hilary Mason, DJ Patil

Checklists...

- ☑ Connect principles to practice.
- ☑ Can be designed to be actionable. Specific, focused on execution, used repeatedly.
- ☑ Help ensure we don't overlook important work.
- ☑ Embeds considerations into the workflow so that conversations happen even in fast-paced environments.

Key Perspectives

- Checklist items are included in the tool and will be presented in today's workshop. This is ***not*** a definitive or exhaustive set of ethical concerns.
  - ➢ It is meant as a sensible starting point, and we strongly encourage domain-specific custom checklists.

- Decisions on ethical courses of action are not up to data scientists alone.
  - ➢ It is meant to provoke discussion regarding issues that are relevant to data scientists within a larger conversation at an organization.

- Strictly statistical best practices are not included.
  - ➢ This is meant to be above and beyond statistical correctness.

Before we dive in...

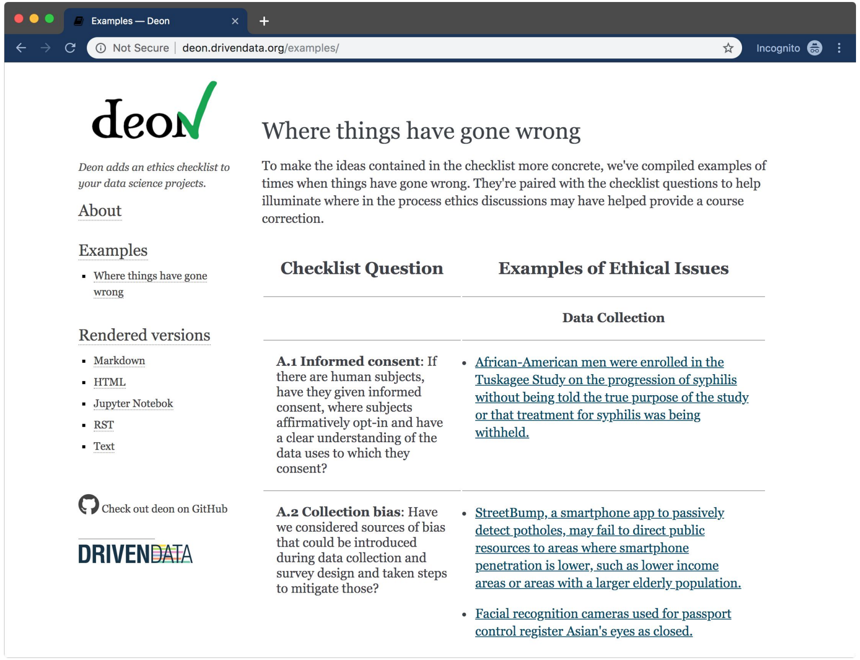
Deon is not the only project with the goal of integrating ethics with day-to-day practice. Here are some others for future reference:

- Ethical OS https://ethicalos.org/
- Ethics & Algorithm Toolkit http://ethicstoolkit.ai/
- Data Practices Courseware https://datapractices.org/courseware/
- Google Responsible AI Practices https://ai.google/responsibilities/responsible-ai-practices/

Deon's Default Checklist

An ethics checklist for data scientists

- Data collection
- Data storage
- Analysis
- Modeling
- Deployment



Examples — Deon

Not Secure | deon.drivendata.org/examples/

Incognito

deon

Deon adds an ethics checklist to your data science projects.

About

Examples

- Where things have gone wrong

Rendered versions

- Markdown
- HTML
- Jupyter Notebok
- RST
- Text

Check out deon on GitHub

DRIVENDATA

Where things have gone wrong

To make the ideas contained in the checklist more concrete, we've compiled examples of times when things have gone wrong. They're paired with the checklist questions to help illuminate where in the process ethics discussions may have helped provide a course correction.

Checklist Question	Examples of Ethical Issues
	Data Collection
A.1 Informed consent: If there are human subjects, have they given informed consent, where subjects affirmatively opt-in and have a clear understanding of the data uses to which they consent?	• African-American men were enrolled in the Tuskagee Study on the progression of syphilis without being told the true purpose of the study or that treatment for syphilis was being withheld.
A.2 Collection bias: Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those?	• StreetBump, a smartphone app to passively detect potholes, may fail to direct public resources to areas where smartphone penetration is lower, such as lower income areas or areas with a larger elderly population. • Facial recognition cameras used for passport control register Asian's eyes as closed.

We believe in the power of examples to bring the principles of data ethics to bear on human experience.

The deon repository includes a list of real-world examples connected with each item in the default checklist.

Data collection

Informed consent

If there are human subjects, have they given informed consent, where subjects affirmatively opt-in and have a clear understanding of the data uses to which they consent?

Collection bias

Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those?

Limit PII exposure

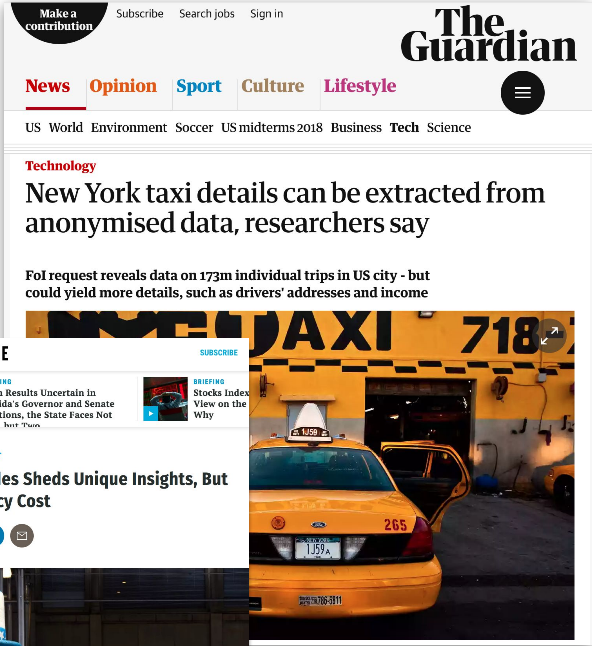
Have we considered ways to minimize exposure of personally identifiable information (PII), for example through anonymization or not collecting information that isn't relevant for analysis?

Data collection

- [ ] **A.3 Limit PII exposure**: Have we considered ways to minimize exposure of personally identifiable information (PII) for example through anonymization or not collecting information that isn't relevant for analysis?

WHERE THINGS HAVE GONE WRONG

Personal information of taxi drivers can be accessed in poorly anonymized dataset (in this case md5 hashing) of 173 million taxi trips released by New York City.


Make a contribution Subscribe Search jobs Sign in

The Guardian

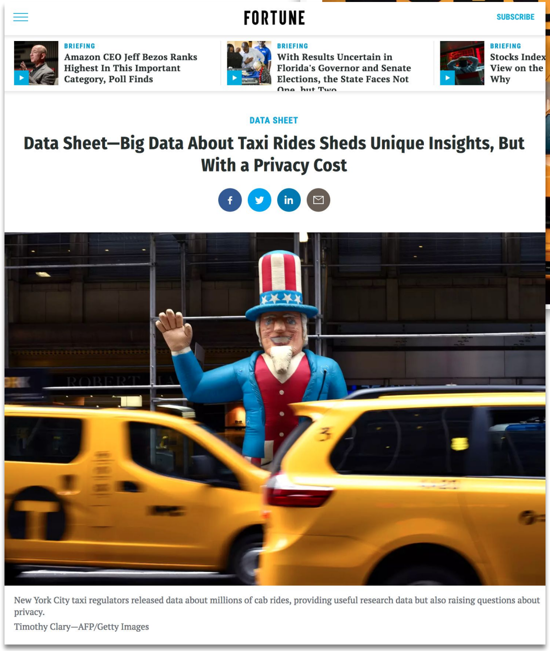
News Opinion Sport Culture Lifestyle

US World Environment Soccer US midterms 2018 Business Tech Science

Technology

New York taxi details can be extracted from anonymised data, researchers say

FoI request reveals data on 173m individual trips in US city - but could yield more details, such as drivers' addresses and income


FORTUNE SUBSCRIBE

BRIEFING Amazon CEO Jeff Bezos Ranks Highest In This Important Category, Poll Finds

BRIEFING With Results Uncertain in Florida's Governor and Senate Elections, the State Faces Not One, but Two

BRIEFING Stocks Index View on the Why

DATA SHEET

Data Sheet—Big Data About Taxi Rides Sheds Unique Insights, But With a Privacy Cost

New York City taxi regulators released data about millions of cab rides, providing useful research data but also raising questions about privacy.

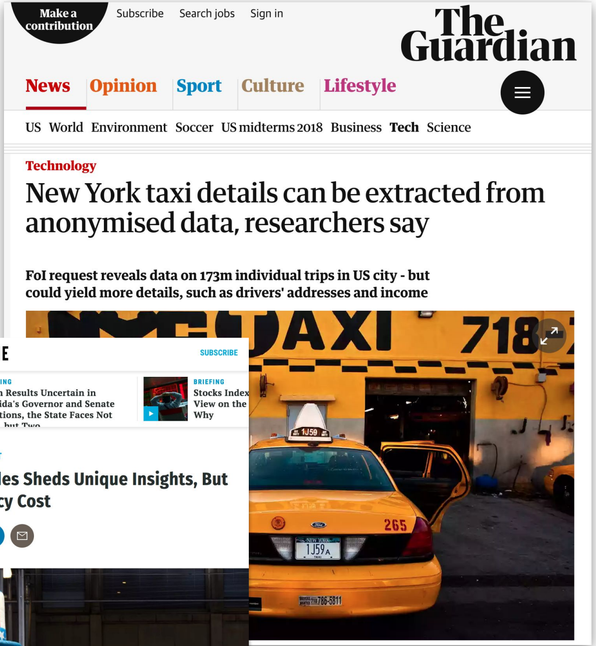
Timothy Clary—AFP/Getty Images

Data collection

A.3 Limit PII exposure: Have we considered ways to minimize exposure of personally identifiable information (PII) for example through anonymization or not collecting information that isn't relevant for analysis?

HOW CHECKING THE BOX MIGHT HAVE HELPED

The formats of medallions–5X55, XX555 or XXX555–and licenses–six or seven-digit numbers starting with a five–bound the size of the set of possible hashes at 24 million. Random ids could be more effective here.



Make a contribution Subscribe Search jobs Sign in

The Guardian

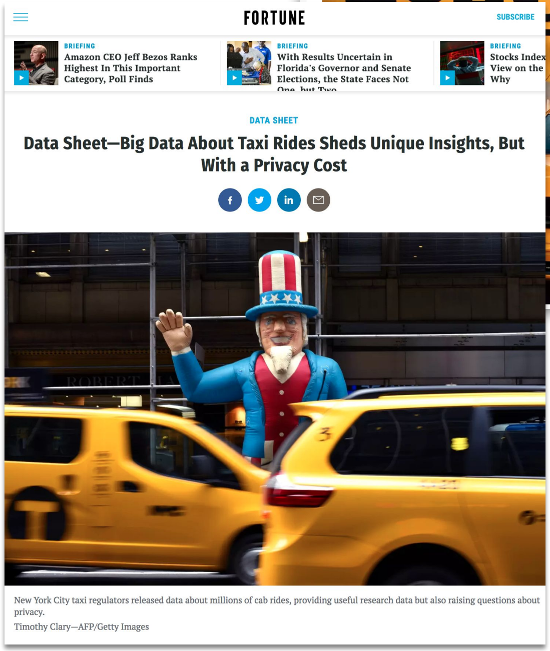
News Opinion Sport Culture Lifestyle

US World Environment Soccer US midterms 2018 Business Tech Science

Technology

New York taxi details can be extracted from anonymised data, researchers say

FoI request reveals data on 173m individual trips in US city - but could yield more details, such as drivers' addresses and income



FORTUNE SUBSCRIBE

BRIEFING Amazon CEO Jeff Bezos Ranks Highest In This Important Category, Poll Finds

BRIEFING With Results Uncertain in Florida's Governor and Senate Elections, the State Faces Not One, but Two

BRIEFING Stocks Index View on the Why

DATA SHEET

Data Sheet—Big Data About Taxi Rides Sheds Unique Insights, But With a Privacy Cost

New York City taxi regulators released data about millions of cab rides, providing useful research data but also raising questions about privacy.

Timothy Clary—AFP/Getty Images

Data storage

Right to be forgotten

Do we have a mechanism through which an individual can request their personal information be removed?

Data retention plan

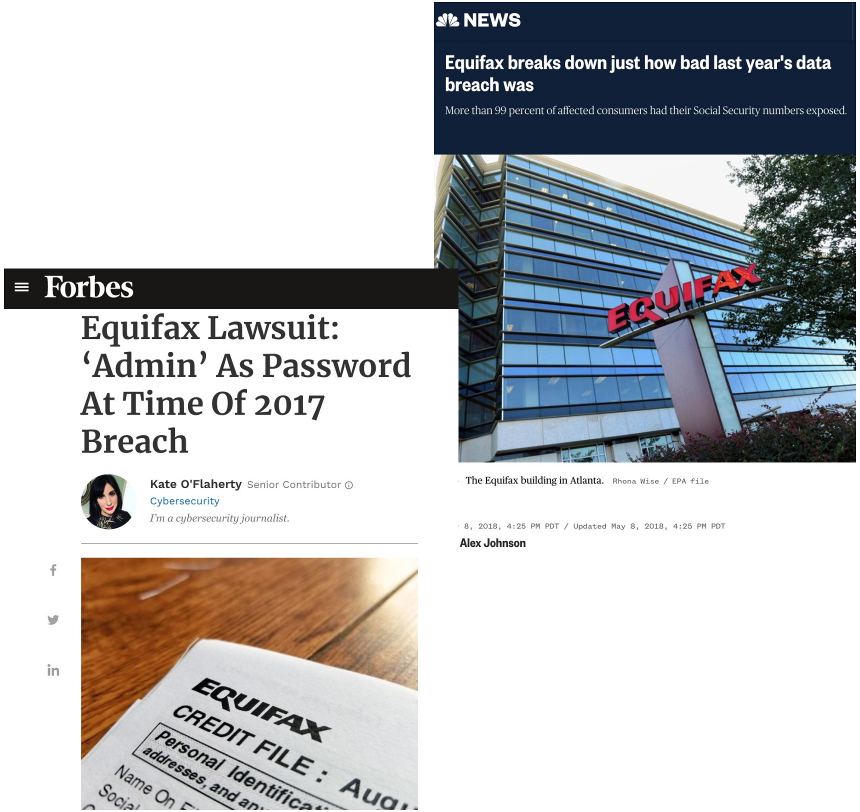
Is there a schedule or plan to delete the data after it is no longer needed?

Data security

Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?

Data storage

- [ ] **B.1 Data security:** Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?



NEWS

Equifax breaks down just how bad last year's data breach was

More than 99 percent of affected consumers had their Social Security numbers exposed.

The Equifax building in Atlanta. Rhona Wise / EPA file

8, 2018, 4:25 PM PDT / Updated May 8, 2018, 4:25 PM PDT

Alex Johnson

Forbes

Equifax Lawsuit: 'Admin' As Password At Time Of 2017 Breach

Kate O'Flaherty Senior Contributor

Cybersecurity

I'm a cybersecurity journalist.

WHERE THINGS HAVE GONE WRONG

Personal and financial data for more than 146 million people was stolen in the Equifax data breach.

Data storage

B.1 Data security: Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?

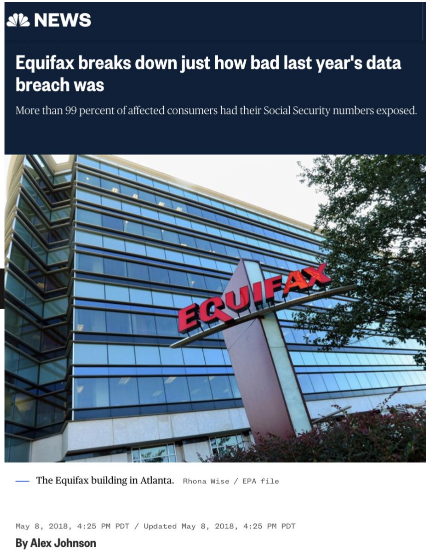
HOW CHECKING THE BOX MIGHT HAVE HELPED

Better passwords, encrypting user data on company servers and mobile applications, or protecting credentials could have avoided or limited the impact of compromised accounts.

Forbes

Equifax Lawsuit: 'Admin' As Password At Time Of 2017 Breach

Kate O'Flaherty Senior Contributor
Cybersecurity
I'm a cybersecurity journalist.



NEWS

Equifax breaks down just how bad last year's data breach was

More than 99 percent of affected consumers had their Social Security numbers exposed.

The Equifax building in Atlanta. Rhona Wise / EPA file

May 8, 2018, 4:25 PM PDT / Updated May 8, 2018, 4:25 PM PDT

By Alex Johnson

Analysis

Honest representation

Are our visualizations, summary statistics, and reports designed to honestly represent the underlying data?

Privacy in analysis

Have we ensured that data with PII are not used or displayed unless necessary for the analysis?

Auditability

Is the process of generating the analysis well documented and reproducible if we discover issues in the future?

Missing perspectives

Have we sought to address blind spots in the analysis through engagement with relevant stakeholders?

Dataset bias

Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases ?

Analysis

- [ ] **C.2 Dataset bias:** Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases (e.g., stereotype perpetuation, confirmation bias, imbalanced classes, or omitted confounding variables)?

WHERE THINGS HAVE GONE WRONG

The geometry of the word2vec embedding space reinforces gender stereotypes. This structure is inherited from the Google News corpus–many of its authors are professional journalists.

MIT Technology Review

Intelligent Machines

How Vector Space Mathematics Reveals the Hidden Sexism in Language

As neural networks tease apart the structure of language, they are finding a hidden gender bias that nobody knew was there.

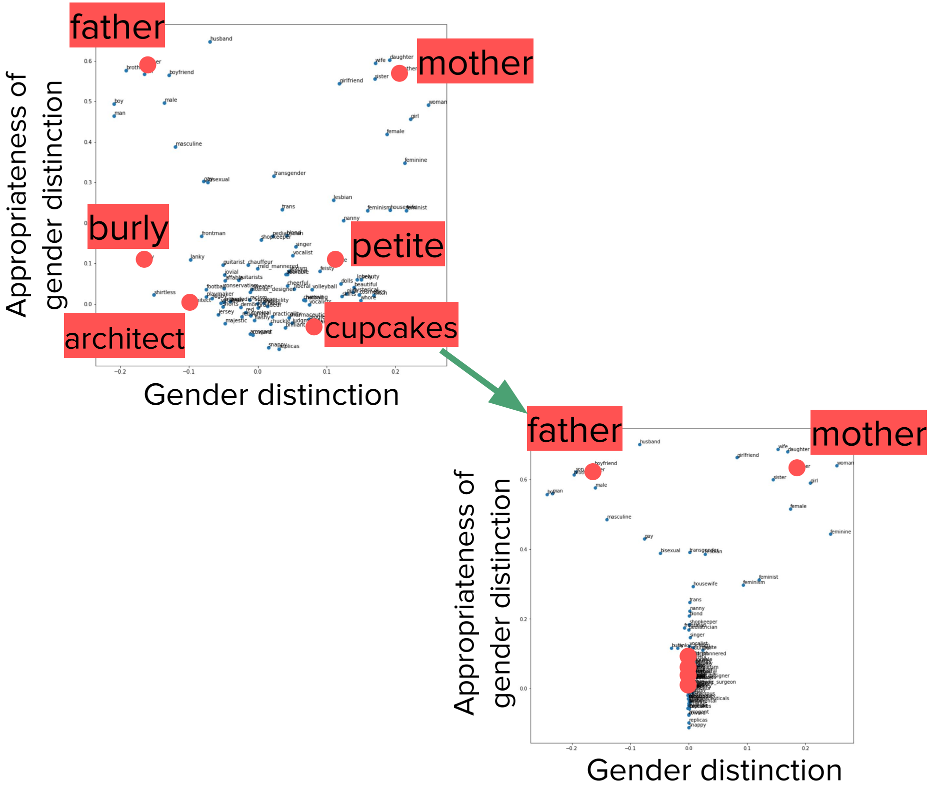
But ask the database “father : doctor :: mother : x” and it will say x = nurse. And the query “man : computer programmer :: woman : x” gives x = homemaker.

Analysis

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Tolga Bolukbasi[1], Kai-Wei Chang[2], James Zou[2], Venkatesh Saligrama[1,2], Adam Kalai[2]
[1]Boston University, 8 Saint Mary's Street, Boston, MA
[2]Microsoft Research New England, 1 Memorial Drive, Cambridge, MA
tolgab@bu.edu, kw@kwchang.net, jamesyzou@gmail.com, srv@bu.edu, adam.kalai@microsoft.com

C.2 Dataset bias: Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases (e.g., stereotype perpetuation, confirmation bias, imbalanced classes, or omitted confounding variables)?



HOW CHECKING THE BOX MIGHT HAVE HELPED

It's important to be aware of biases in the data and techniques we're using. It may be possible to find an alternative approach or transformation. If not, it's still important to factor these considerations into your decision making process.

Modeling

Proxy discrimination

Have we ensured that the model does not rely on variables or proxies for variables that are unfairly discriminatory?

Metric selection

Have we considered the effects of optimizing for our defined metrics and considered additional metrics?

Explainability

Can we explain in understandable terms a decision the model made in cases where a justification is needed?

Communicate bias

Have we communicated the shortcomings, limitations, and biases of the model to relevant stakeholders in ways that can be generally understood?

Fairness across groups

Have we tested model results for fairness with respect to different affected groups (e.g., tested for disparate error rates)?

Modeling

☐ **D.2 Fairness across groups:** Have we tested model results for fairness with respect to different affected groups (e.g., tested for disparate error rates)?

WHERE THINGS HAVE GONE WRONG

Amazon scraps AI recruiting tool that showed bias against women.

Forbes

Are AI Hiring Programs Eliminating Bias Or Making It Worse?

Nicole Martin Contributor
AI & Big Data
I write about digital marketing, artificial intelligence, healthcare innovation, data and privacy concerns.

ur resume GETTY

REUTERS

BUSINESS NEWS
OCTOBER 9, 2018 / 8:12 PM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Modeling

☑ **D.2 Fairness across groups:** Have we tested model results for fairness with respect to different affected groups (e.g., tested for disparate error rates)?

HOW CHECKING THE BOX MIGHT HAVE HELPED

In this case, the model was in fact checked, but too late. Incorporating these ethical considerations into the standard development workflow could have surfaced this particular bias before the tool went into use.

Forbes

Are AI Hiring Programs Eliminating Bias Or Making It Worse?

Nicole Martin Contributor

AI & Big Data

I write about digital marketing, artificial intelligence, healthcare innovation, data and privacy concerns.

ur resume GETTY

REUTERS

BUSINESS NEWS

OCTOBER 9, 2018 / 8:12 PM / A YEAR AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Jeffrey Dastin

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Deployment

Redress

Have we discussed with our organization a plan for response if users are harmed by the results?

Roll back

Is there a way to turn off or roll back the model in production if necessary?

Unintended use

Have we taken steps to identify and prevent unintended uses and abuse of the model and do we have a plan to monitor these once the model is deployed?

Concept drift

Do we test and monitor for concept drift to ensure the model remains fair over time?

Deployment

☐ **E.3 Concept drift:** Do we test and monitor for concept drift to ensure the model remains fair over time?

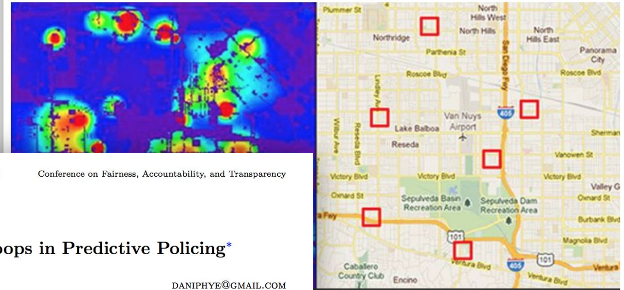
WHERE THINGS HAVE GONE WRONG

Sending police officers to areas of high predicted crime creates a feedback loop, skewing future data as police are repeatedly sent back to the same neighborhoods.

Smithsonian.com SUBSCRIBE

Artificial Intelligence Is Now Used to Predict Crime. But Is It Biased?

The software is supposed to make policing more fair and accountable. But critics say it still has a way to go.



ms that identify potential crime hotspots..

Proceedings of Machine Learning Research 81:1–12, 2018 Conference on Fairness, Accountability, and Transparency

Runaway Feedback Loops in Predictive Policing*

Danielle Ensign DANIPHYE@GMAIL.COM
University of Utah
Sorelle A. Friedler SORELLE@CS.HAVERFORD.EDU
Haverford College
Scott Neville DROP.SCOTT.N@GMAIL.COM
University of Utah
Carlos Scheidegger CSCHEID@CSCHEID.NET
University of Arizona
Suresh Venkatasubramanian† SURESH@CS.UTAH.EDU
University of Utah

Deployment

E.3 Concept drift: Do we test and monitor for concept drift to ensure the model remains fair over time?

HOW CHECKING THE BOX MIGHT HAVE HELPED

Recognize the feedback loop in the design, and change the design to mitigate its impact, for example applying alternative sampling strategies can lead to distributions that account for shortcomings of the basic model.

Ethics in practice

Qualitative Case Study
Group Discussion

Split into groups

Setting up deon and exercise

https://github.com/drivendataorg/odsc-west-2019

Demo

scenario

The San Francisco mayor, London Breed, has hired your team to perform an analysis on the current health of San Franciscans.

She and other policymakers want to understand how healthy (or unhealthy) San Franciscans are in different parts of the city in order to inform health policy interventions.

data

The mayor has formed an agreement with Apple to get data from their HealthKit App, which passively tracks activity using the phone’s sensors, along with other metadata like age, gender, and location.

task

Use this data to produce visualizations helping policymakers understand the current state as well as trends in health across San Francisco neighborhoods.

Consider carefully the ethical implications of your choices.

scenario

Insurance companies have gotten wind of this great dataset and preliminary analysis.

They want to use this fitness data to reward (incentivize) people who are ‘very healthy’ by awarding them credits toward health insurance premiums.

data

Passively collected health data from Android and iPhones which captures activity level, weight, height, and blood pressure, as well as metadata on gender, age, and location.

Data is received in monthly batches. Your initial dataset contains one year of daily data for 1 million people.

task

Design and deploy a model that identifies who should receive credits ($) toward their health insurance.

Consider carefully the ethical implications of your choices.

Scenario 1

https://github.com/drivendataorg/odsc-west-2019

- **Context:** The San Francisco mayor, London Breed, has hired your team to perform an analysis on the current health of San Franciscans. She and other policymakers want to understand how healthy (or unhealthy) San Franciscans are in different parts of the city in order to inform health policy interventions.
- **Data:** The mayor has formed an agreement with Apple to get data from their HealthKit App, which passively tracks activity using the phone's sensors, along with other metadata like age, gender, and location.
- **Task:** Use this data to produce visualizations helping policymakers understand the current state as well as trends in health across San Francisco neighborhoods. Consider carefully the ethical implications of your choices.

Scenario 2

- **Context:** Insurance companies have gotten wind of this great dataset and preliminary analysis. They want to use this fitness data to reward (incentivize) people who are 'very healthy' by awarding them credits toward health insurance premiums.
- **Data:** Passively collected health data from Android and iPhones which captures activity level, weight, height, and blood pressure, as well as metadata on gender, age, and location. Data is received in monthly batches. Your initial dataset contains one year of daily data for 1 million people.
- **Task:** Design and deploy a model that identifies who should receive credits ($) toward their health insurance. Consider carefully the ethical implications of your choices.

Share

Ethics in practice

Eviction Data Case Study

Hands-on Exercise

Some additional helpful tools

These are some tools to help you examine your data and check for things like bias and fairness:

- Google's What If https://pair-code.github.io/what-if-tool/
- Google's Facets https://pair-code.github.io/facets/
- DSSG's Aequitas http://www.datasciencepublicpolicy.org/projects/aequitas/

What did we talk about?

What is data ethics?

Checklists and deon

Power of examples

“The first principle is that you must not fool yourself—and you are the easiest person to fool. So you have to be very careful about that.”

– Richard Feynman

Thank you!

Workshop materials: github.com/drivendataorg/odsc-west-2019

Deon: deon.drivendata.org

Jay Qi
Senior Data Scientist
jay@drivendata.org
@jayyqi

Casey Fitzpatrick, PhD
Machine Learning Engineer
casey@drivendata.org
@caseyfitz

DRIVENDATA
Data Science + Social Impact

https://www.drivendata.org
https://github.com/drivendataorg
@drivendataorg