Actionable Ethics for Data Scientists

BIG DATA
IGNITE
2020

October 1, 2020

DRIVENDATA

Jay Qi
Senior Data Scientist

@jayyqi

DRIVENDATA

Data Science + Social Impact

Data Science Competitions · Direct Client Engagements · Open Source Projects

drivendata.org

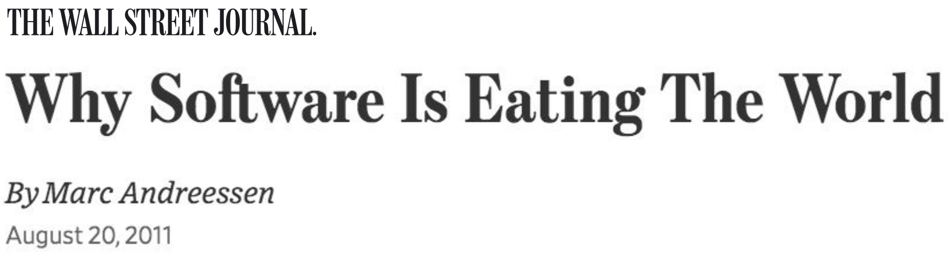
https://github.com/drivendataorg

@drivendataorg

Agenda

- **What is data ethics and why does it matter?**
- **An actionable approach to data ethics**
    - deon: an ethics checklist for data scientists
        - Why a checklist
        - Checklist content
        - Examples
- **Q&A**

Why does data ethics matter?

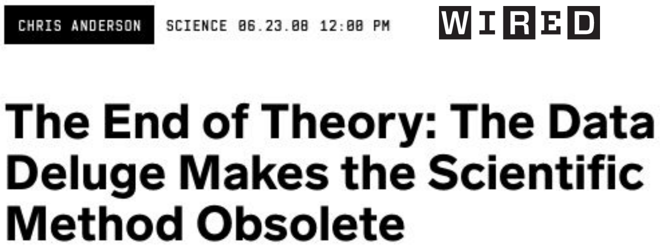


THE WALL STREET JOURNAL.

Why Software Is Eating The World

By Marc Andreessen

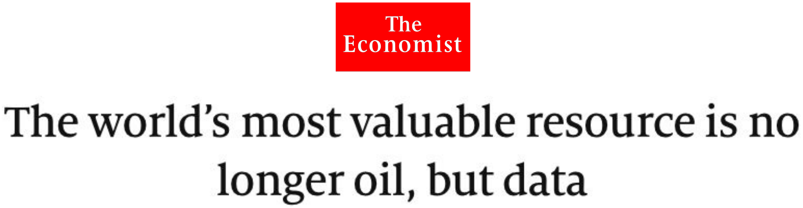
August 20, 2011



CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM

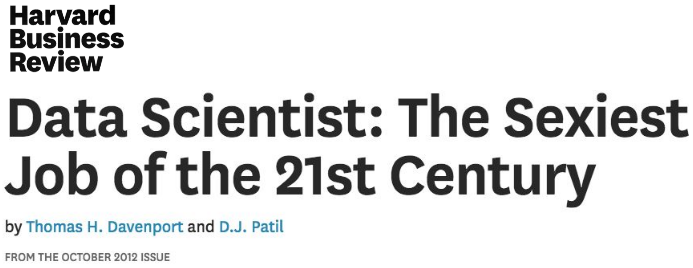
WIRED

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

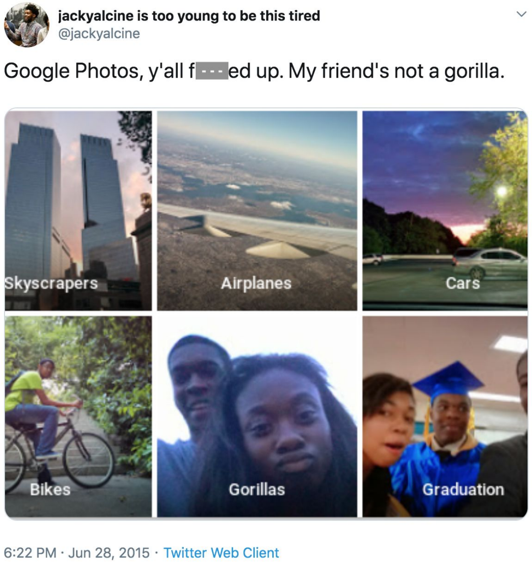


Harvard Business Review

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[redacted]ed up. My friend's not a gorilla.

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

The Guardian

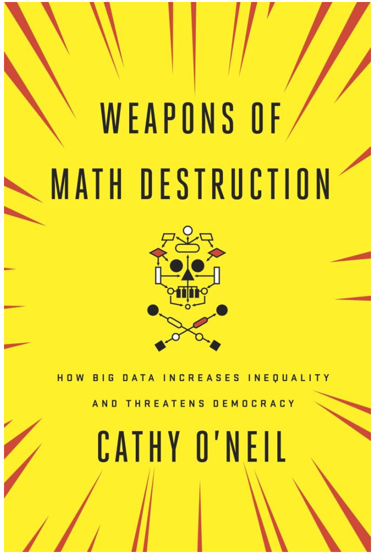
Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

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

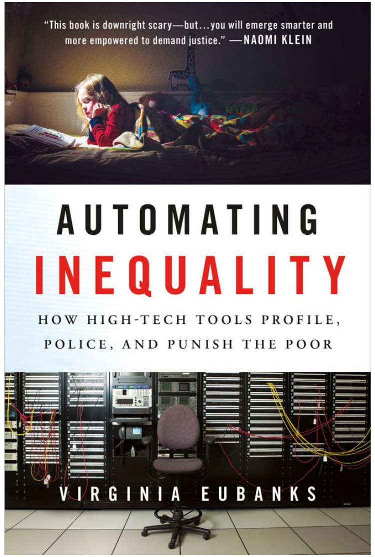
PROPUBLICA



WEAPONS OF MATH DESTRUCTION

HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY

CATHY O'NEIL



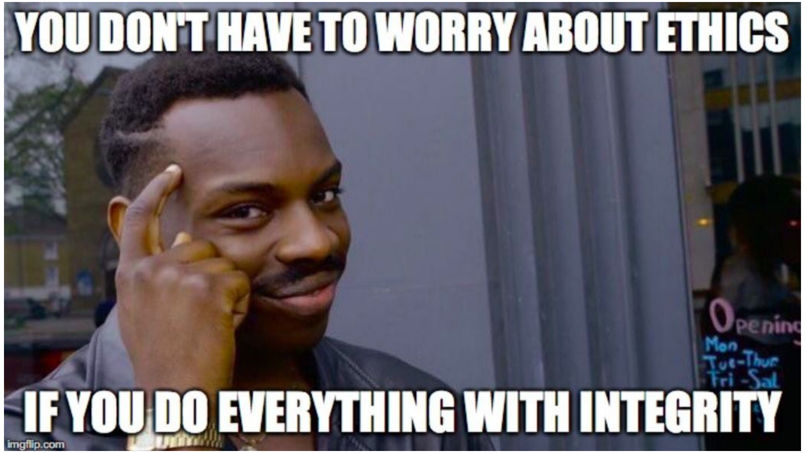
"This book is downright scary—but…you will emerge smarter and more empowered to demand justice." —NAOMI KLEIN

AUTOMATING INEQUALITY

HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR

VIRGINIA EUBANKS

As data scientists, we are work on complex systems, and we generally have good intentions.


YOU DON'T HAVE TO WORRY ABOUT ETHICS
IF YOU DO EVERYTHING WITH INTEGRITY
imgflip.com
Opening
Mon
Tue-Thur
Fri-Sat

But we do have to worry about ethics.
Good intentions alone aren't enough.

Ethical outcomes need deliberate and active effort.

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

How do we think about data ethics systematically?

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.

A set of foundational values and beliefs. Abstract.

Professional codes of ethics are examples of this.

Important and necessary, but not sufficient for outcomes.

Not the focus of today's talk.

ethics

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

Briefly, some 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/

ethics

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

Briefly, some 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?

Today’s talk will be about making ethics **practical** and **actionable**.

Ethics is hard.

There is no free lunch. Tradeoffs are inevitable.

Not one right answer. Reasonable people can disagree.

Good intentions aren’t enough. Must actively consider and anticipate consequences.

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

A practical and actionable approach to data ethics

deon

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

https://deon.drivendata.org

Why an ethics checklist?

Inspired by long-standing checklists in other professions, such as surgery and aviation, and by Of Oaths and Checklists by Mike Loukides, Hilary Mason, DJ Patil

Checklists...

☑ Connect principles to practice.

☑ Are designed to be actionable: specific, focused on execution, used repeatedly.

☑ Help ensure we don’t overlook important issues by embedding considerations into the workflow.

Why a Python package?

Command line tool

- Easily integrated into a data science workflow
- Scriptable
- Customizable
- Support for many formats: .md, .html, .ipynb, .rst, .txt

deon✓

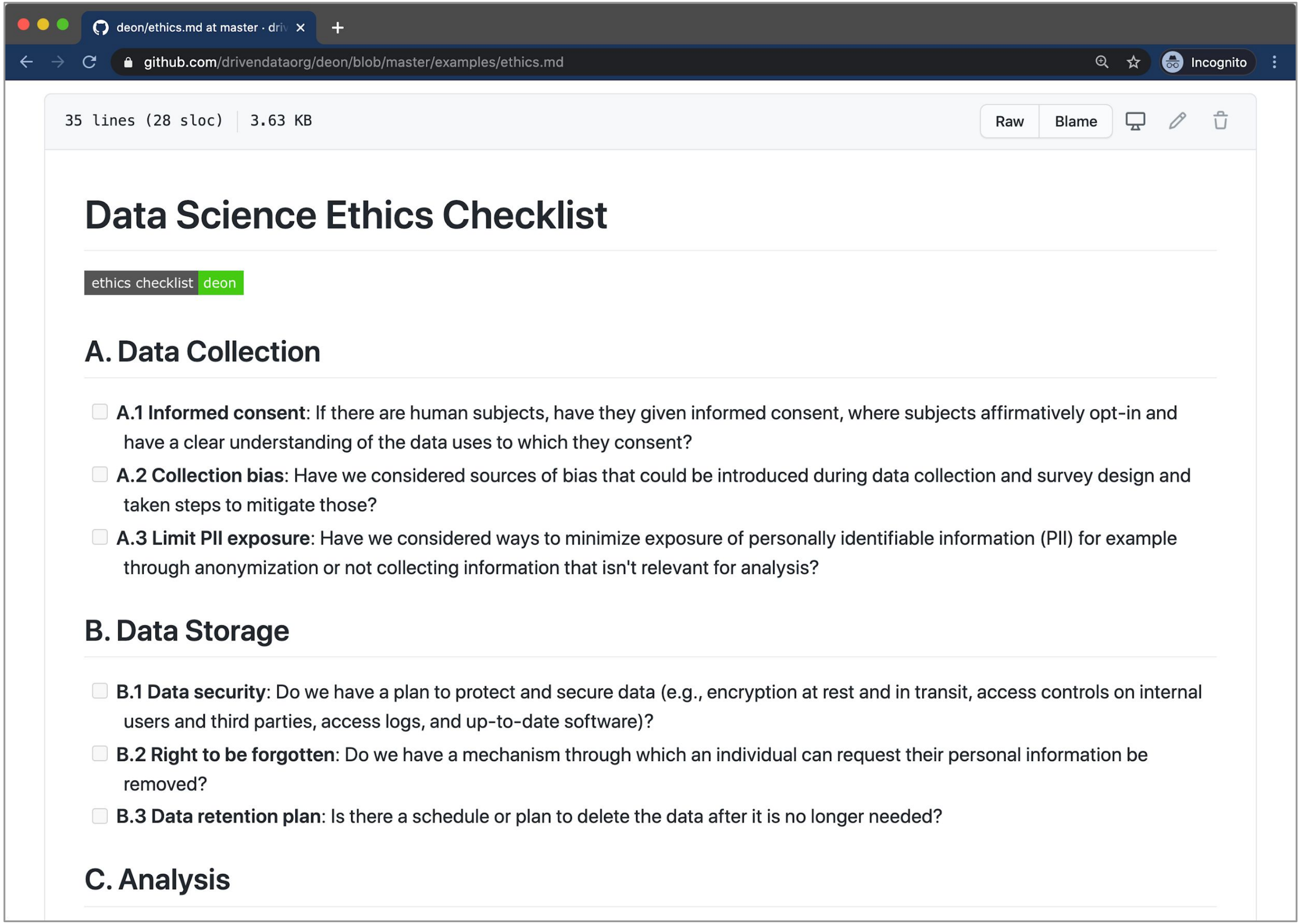
Deon is an open-source command line tool that allows you to easily add an ethics checklist to your data science projects.

```
pip install deon
```

or

```
conda install deon -c conda-forge
```

```
deon -o ethics.md
deon --help
```


deon/ethics.md at master · driv
github.com/drivendataorg/deon/blob/master/examples/ethics.md
Incognito
35 lines (28 sloc) | 3.63 KB
Raw
Blame

Data Science Ethics Checklist

ethics checklist deon

A. 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?
- □ **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?
- □ **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?

B. 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)?
- □ **B.2 Right to be forgotten**: Do we have a mechanism through which an individual can request their personal information be removed?
- □ **B.3 Data retention plan**: Is there a schedule or plan to delete the data after it is no longer needed?

C. Analysis

Key design perspectives

- **Our goal is not to be arbitrators of which ethical concerns merit inclusion.**
  - The default checklist is meant as a sensible starting point, and we believe teams will benefit from building custom checklists.

Key design perspectives

- Our goal is not to be arbitrators of which ethical concerns merit inclusion.
    - The default checklist is meant as a sensible starting point, and we believe teams will benefit from building custom checklists.
- **Checklist items are meant to provoke discussion.**
    - The goal of the checklist items are not to concretely recommend a specific action but rather are framed as prompts to discuss or consider.

Key design perspectives

deon✓

- Our goal is not to be arbitrators of which ethical concerns merit inclusion.
  - The default checklist is meant as a sensible starting point, and we believe teams will benefit from building custom checklists.
- Checklist items are meant to provoke discussion.
  - The goal of the checklist items are not to concretely recommend a specific action but rather are framed as prompts to discuss or consider.
- **Decisions on ethical courses of action are not up to data scientists alone.**
  - Checklist is designed to provoke conversations around issues where data scientists have particular responsibility and perspective.

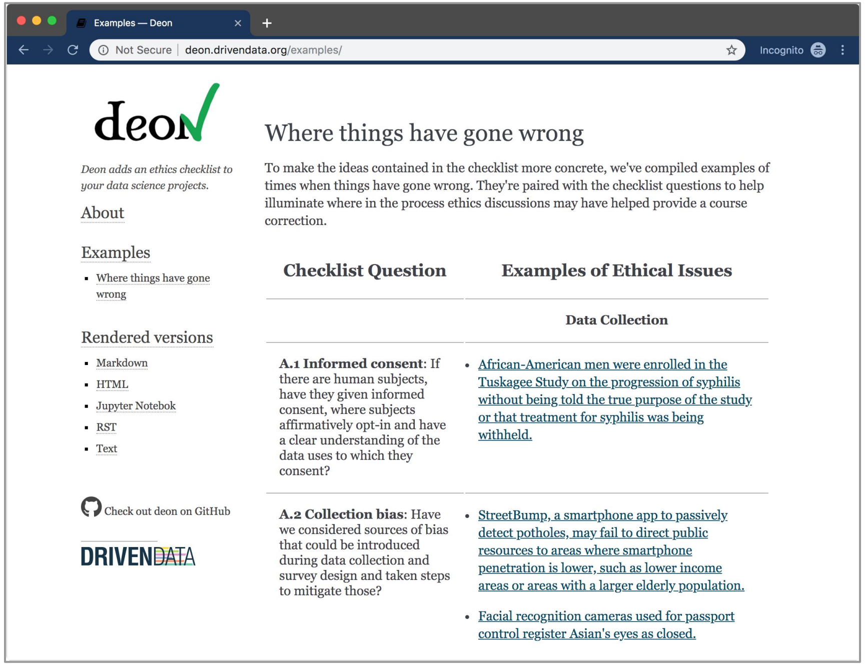
Key design perspectives

deon✓

- Our goal is not to be arbitrators of which ethical concerns merit inclusion.
    - The default checklist is meant as a sensible starting point, and we believe teams will benefit from building custom checklists.
- Checklist items are meant to provoke discussion.
    - The goal of the checklist items are not to concretely recommend a specific action but rather are framed as prompts to discuss or consider.
- Decisions on ethical courses of action are not up to data scientists alone.
    - Checklist is designed to provoke conversations around issues where data scientists have particular responsibility and perspective.
- **Strictly statistical best practices are not included.**
    - This is meant to be above and beyond statistical correctness.

An ethics checklist for data scientists

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



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.

Examples on the following slides can be found at https://deon.drivendata.org/examples/

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

The deon documentation 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

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?

Where things have gone wrong:

Collection bias

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. ↗

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 storage

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)?

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 storage

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)?

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?

Where things have gone wrong:

Data Security

Personal and financial data for more than 146 million people was stolen in the Equifax data breach in 2017. ↗

Data storage

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)?

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?

Analysis

Missing perspectives

Have we sought to address blind spots in the analysis through engagement with relevant stakeholders (e.g., checking assumptions and discussing implications with affected communities and subject matter experts)?

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)?

(cont.)

Analysis

Missing perspectives

Have we sought to address blind spots in the analysis through engagement with relevant stakeholders (e.g., checking assumptions and discussing implications with affected communities and subject matter experts)?

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)?

(cont.)

Where things have gone wrong:

Dataset bias

The popular word2vec embedding, trained on Google News corpus, reinforces gender stereotypes. ↗

man : king *woman : queen*

father : doctor *mother : nurse*

man : computer programmer
woman : homemaker

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?

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?

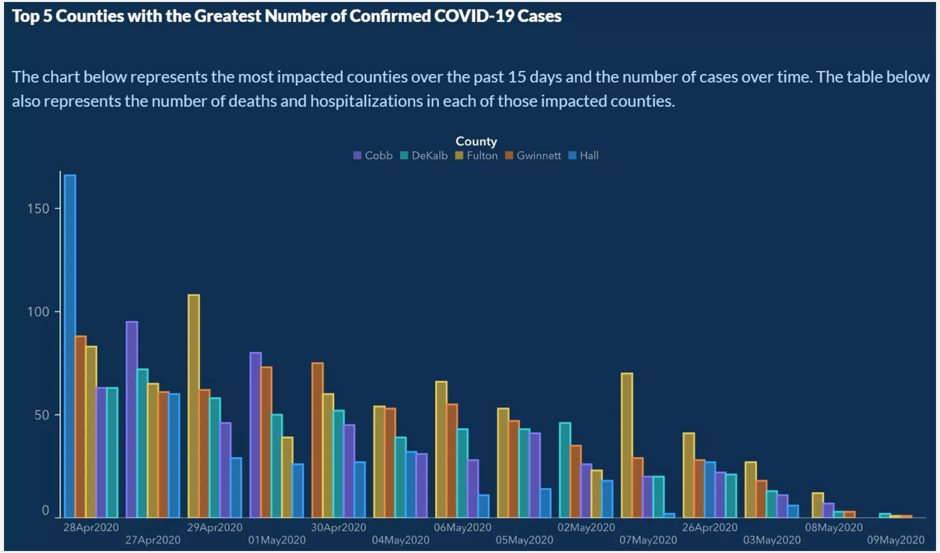
Auditability

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

Where things have gone wrong:

Honest representation

The initial version of a plot of COVID-19 cases from the Georgia Dept. of Public Health was misleading. The x-axis shows dates, but was sorted by decreasing case counts and not by time, making it look like cases were decreasing when they weren't. ↗



Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.

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?

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?

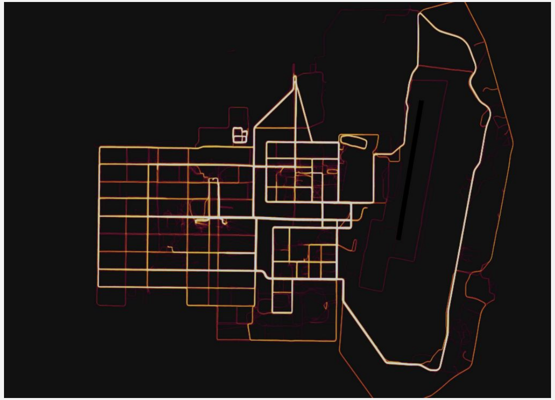
Auditability

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

Where things have gone wrong:

Privacy in analysis

Strava heatmap of exercise routes reveals sensitive information on military bases and spy outposts. ↗



A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap courtesy of The Guardian.

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?

Modeling

Proxy discrimination

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

Fairness across groups

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

Metric selection

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

(cont.)

Modeling

Proxy discrimination

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

Fairness across groups

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

Metric selection

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

(cont.)

Where things have gone wrong:

Proxy discrimination

Amazon scraps secret AI recruiting tool that showed bias against women. It was a tool to automatically review resumes.

Gender was not an explicit input, but learned from training data that reflected existing gender imbalance in the tech industry.

The model penalized resumes that included the word "women's," as in "women's chess club captain." It also penalized graduates of all-women's colleges. ↗

Modeling

Proxy discrimination

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

Fairness across groups

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

Metric selection

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

(cont.)

Modeling

Proxy discrimination

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

Fairness across groups

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

Metric selection

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

(cont.)

Where things have gone wrong:

Fairness across groups
and metric selection

The COMPAS recidivism risk algorithm was the subject of a prominent ProPublica investigation Machine Bias. *They found the false positive rate for black people was twice as high as white people.* ↗

This ended up being a complex topic. The COMPAS model was tuned to have equal accuracy between the groups, but did not have predictive rate parity (equal false positive rates). ↗

There are different ways to define fairness, models can be optimized differently. This demonstrates the importance of transparency and dialogue in policy-making to address these issues.

Modeling

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?

Modeling

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?

Where things have gone wrong:

Explainability

Patients with pneumonia with a history of asthma are usually admitted to the intensive care unit as they have a high risk of dying from pneumonia.

Given the success of the intensive care, neural networks predicted asthmatics had a low risk of dying and could therefore be sent home. Without explanatory models to identify this issue, patients may have been sent home to die. ↗

Modeling

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?

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?

Concept drift

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

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?

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?

Concept drift

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

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?

Where things have gone wrong:

Sending police officers to areas of high predicted crime can skew future training data collection as police are repeatedly sent back to the same neighborhoods regardless of the true crime rate. ↗

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?

Concept drift

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

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?

What did we talk about?

1. What is data ethics?
2. Checklist framework and deon
3. 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!

Learn more at:
http://deon.drivendata.org/

DRIVENDATA

Additional resources

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

- Ethical OS
- Ethics & Algorithm Toolkit
- Data Practices Courseware
- Google Responsible AI Practices
- Ethics and Data Science (ebook)