Actionable Ethics for Data Scientists

Emily Miller
Data Scientist, DrivenData

Christine Chung
Data Scientist, DrivenData

April 14, 2020

DRIVENDATA

Data Science + Social Impact

Data Science Competitions • Direct Client Engagements • Open Source Projects

drivendata.org

https://github.com/drivendataorg

@drivendataorg

Agenda

- **Deon, an ethics checklist for data scientists** (55 min)
  - Questions in Slack

  Break (5 min)

- **Group activity: Qualitative case study** (55 min)
  - Breakout groups in Zoom
  - Share out in Slack

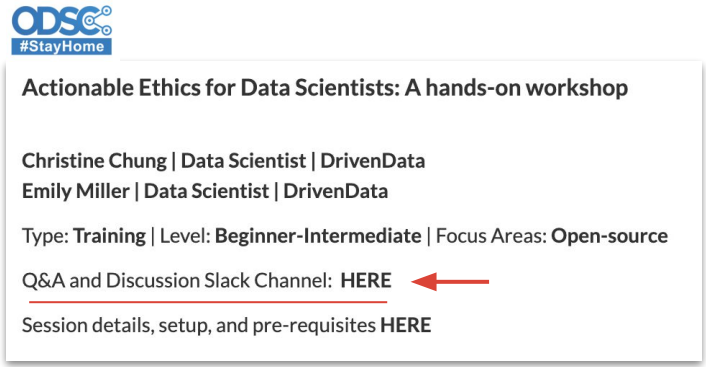
  Break (5 min)

- **Hands-on coding exercise: Eviction data case study** (55 min)
  - Share questions and findings in Slack

Slack

#tue_christine-chung-emily-miller_actionable-ethics-for-data-scientists

- We'll use it for **questions**, case study **share-outs**, and to share **relevant links**.
- DM **Christine Chung** or **Emily Miller** with specific installation questions.
- Please refrain from using Zoom chat.



Next to the where you joined this workshop on live.odsc.com, click the link to join the slack channel.

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 is hard.

There is no free lunch. Tradeoffs are inevitable.

No one right answer. Reasonable people can disagree.

Good intentions aren't enough. Must intentionally consider consequences.

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

The Economist

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

CHRIS ANDERSON SCIENCE 06.23.08 12:00 PM WIRED

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

Harvard Business Review

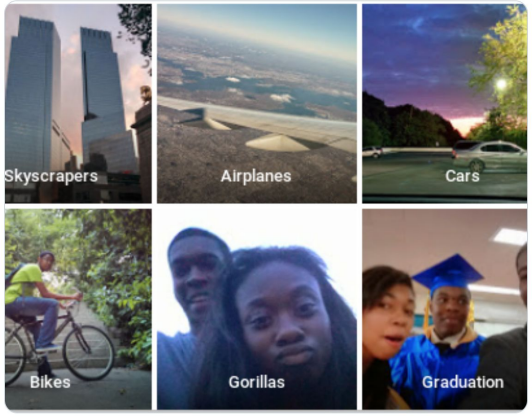
Data Scientist: The Sexiest Job of the 21st Century

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

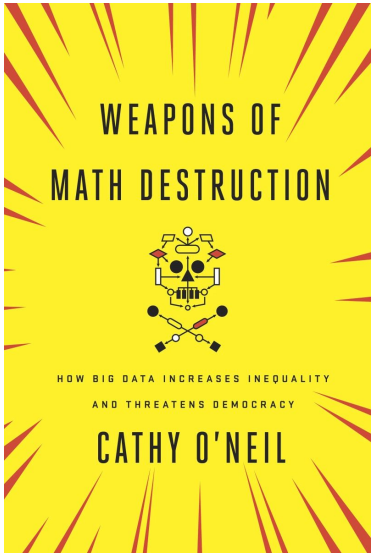
Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

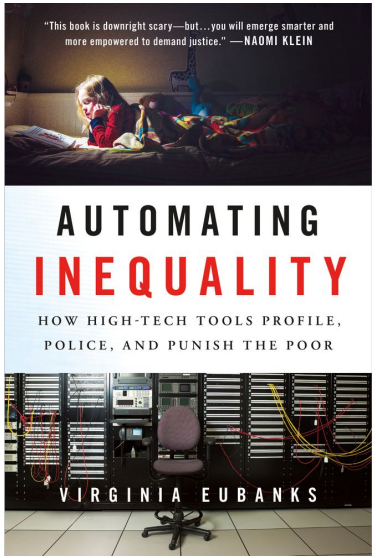
Machine Bias

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





The conversation about ethics in data science, machine learning, and AI is increasingly important. How can we help data scientists to engage in an actionable and practical way?

A practical and actionable approach to data ethics

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.
- ✅ Are designed to be actionable: specific, focused on execution, used repeatedly.
- ✅ Help ensure we don't overlook important work by embedding considerations into the workflow.

Why a python package?

Command line tool

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

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
```

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

Key perspectives

deon

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

Key perspectives

- Our goal is not to be arbitrators of what ethical concerns merit inclusion.
    - It 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 perspectives

deon

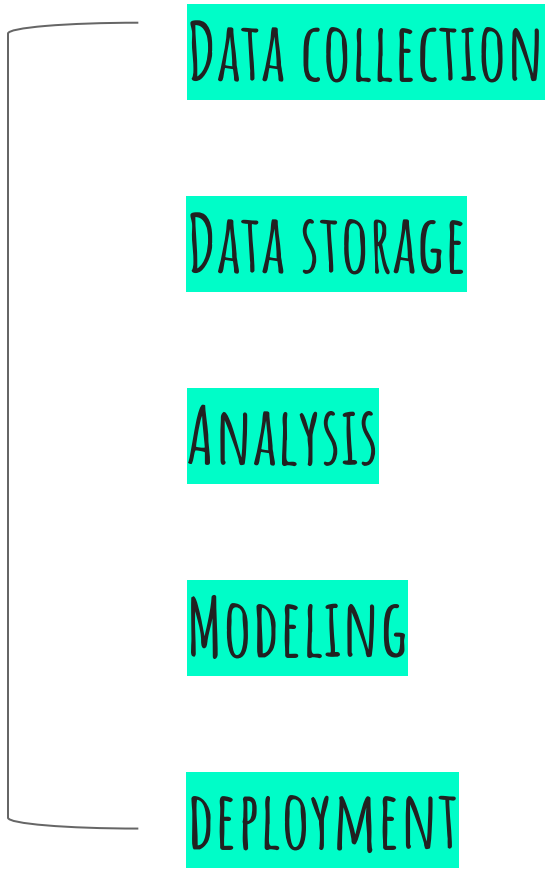
- Our goal is not to be arbitrators of what ethical concerns merit inclusion.
    - It is meant as a sensible starting point, and we believe teams will benefit from building custom checklists.
- Checklist items are meant to provoke discussion.
    - The goal of the checklist items are not to concretely 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.

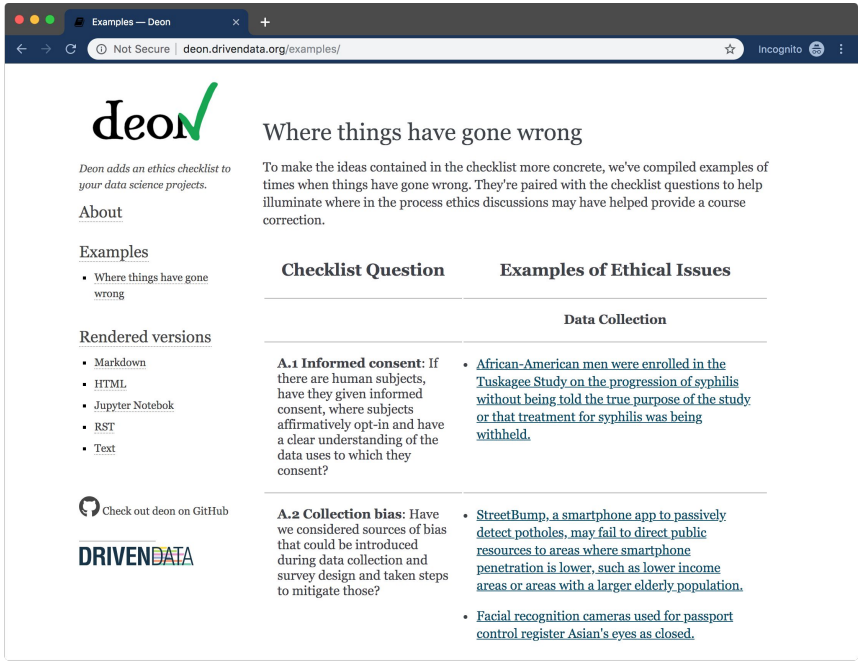
Key perspectives

deon

- Our goal is not to be arbitrators of what ethical concerns merit inclusion.
    - It 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





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:

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 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:

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

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

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:

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

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:

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.

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

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:

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.

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.

Questions?

Ethics in practice: Let's discuss

https://github.com/drivendataorg/odsc-actionable-ethics

1. Clone the repo.
2. Follow the installation instructions in the README.
3. Generate an ethics checklist at the command line with `deon -o ethics.md`
4. Open the ethics.md file in a text editor.

If you have difficulties with installation, mention **Christine Chung** or **Emily Miller** on the slack channel and open https://deon.drivendata.org/#default-checklist in the meantime.

scenario

The mayor of Boston has hired your team to perform an analysis on the current health of Bostonians.

She and other policymakers want to understand how healthy (or unhealthy) Bostonians 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 six months of data from their HealthKit App. This app passively tracks users' physical activity using the phone's sensors, and contains other metadata like age, gender, height, weight, and location. There are optional fields where users can track nutrition, sleep, and mindfulness habits.

task

Use this data to produce visualizations helping policymakers understand the current state as well as trends in health across Boston neighborhoods. Consider carefully the ethical implications of your choices.

scenario

The mayor of Boston has hired your team to perform an analysis on the current health of Bostonians.

She and other policymakers want to understand how healthy (or unhealthy) Bostonians 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 six months of data from their HealthKit App. This app passively tracks users' physical activity using the phone's sensors, and contains other metadata like age, gender, height, weight, and location. There are optional fields where users can track nutrition, sleep, and mindfulness habits.

task

Use this data to produce visualizations helping policymakers understand the current state as well as trends in health across Boston neighborhoods. Consider carefully the ethical implications of your choices.

With your breakout group:

- Talk through each item in the **data collection, storage,** and **analysis** sections of the checklist.
- Write 3-5 bullets (per group) for each section on what specific ethical concerns you see and how you'd plan to address them.
- Share these in the **slack** channel.
- Choose one person from your group to share out your findings.

Share out

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 includes physical activity (type and duration), resting heart rate, age, gender, height, weight, and location. There are optional fields where users can track nutrition, sleep, and mindfulness habits. 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

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 includes physical activity (type and duration), resting heart rate, age, gender, height, weight, and location. There are optional fields where users can track nutrition, sleep, and mindfulness habits. 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.

With your breakout group:

- Talk through each item in the **analysis, modeling,** and **deployment** sections of the checklist. You may want to briefly review how data collection and storage concerns have shifted.
- Write 3-5 bullets (per group) for each section on what specific ethical concerns you see and how you'd plan to address them.
- Share these in the **slack** channel.
- Choose a ***different*** person from your group to share out your findings.

Share out

Ethics in practice: Let's code!

https://github.com/drivendataorg/odsc-actionable-ethics

- Open your terminal
- Move to your odsc folder: `cd odsc-actionable-ethics`
- Activate your environment: `conda activate odsc-ethics`
- Launch a jupyter notebook: `jupyter notebook`
- Open **notebooks / eviction-data-case-study.ipynb**

Mention **Emily Miller** on the slack channel if you experience issues.

Questions?

“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/

Workshop materials:
github.com/drivendataorg/odsc-actionable-ethics

DRIVENDATA