Ethics in Data-Driven Industries

What might it mean to “do” ethics?

Emanuel Moss - Data & Society, CUNY Graduate Center
emanuel@datasociety.net

Friederike Schüür - Cityblock Health
friederike.schueuer@gmail.com

EPIC 2019

Case Studies

[Part 1]

Case Studies [Part 1]

For each case:

- What are the risks from a Technical Perspective?
- What are the risks from a Systems Perspective?
- What are the risks from a Socio-Technical Perspective?

Case Studies

https://docs.google.com/document/d/1E0QFadJuMuDWetaO6Qv7mv4JoR82CCMjWlv5hXE9Mlg/edit?usp=sharing

Case Study: Facebook News Feed

Determines which friends' posts or publisher content you see, in what order

Determined in part by content of a post, but also who else clicks on what, and how similar they are to you

- What are some of the data that might be used to train the News Feed Algorithm? What are some of the real-world impacts of this Algorithm?
- Identify potential ethical risks using Technical, Systems, and Socio-technical Perspectives

http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html

Case Study: Algorithmic Hiring / HireVue

Examines facial expressions, eye contact, "enthusiasm", and vocal style

Compares interviewees' data with that gathered from an employer's employees to generate an "employability score".

Determines who gets later-stage interviews and plays a role in final hiring decisions

- What are some limitations in this product? What are some of the real-world impacts of using HireVue?
- Identify potential ethical risks using Technical, Systems, and Socio-technical Perspectives

Case Study: Resource Allocation in Healthcare

Determines the amount of resources dedicated to a patient.

Determined in part by medical history and past utilization of medical and social services including incurred cost.

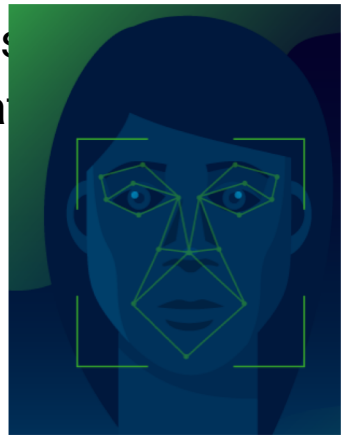
- Data used to train such algorithms: medical claims data, electronic health record (EHR) data.
- What are risks from a technical, systems, and sociotechnical perspective?

Case Study: Facial Recognition in Atlantic Towers

Determines whether or not someone can enter an Atlantic Towers apartment. Atlantic Towers is rent-controlled, and any modifications to the entryway must be approved by a municipal authority. Atlantic Towers residents are majority non-white.

Uses biometric fingerprinting to match a facial image with a databas
permitted residents. Residents submit an initial face scan for the da

Data is only shared between the vendor and the landlord.



- What are some limitations in this system? What are some of the real-world impacts?
- Identify potential ethical risks using Technical, Systems, and Socio-technical Perspectives

https://gothamist.com/news/is-it-facial-or-is-it-racial-brooklyn-tenants-fight-proposed-facial-recognition-tech

Recommendations & Solutions

Babysteps toward the apocalypse

How might we *decrease the likelihood* that organization that work with data and develop technology build products that *accidentally* harm others?

Technical Perspective

Entirely internal to the technical operations of an AI product or feature

Technical Perspective

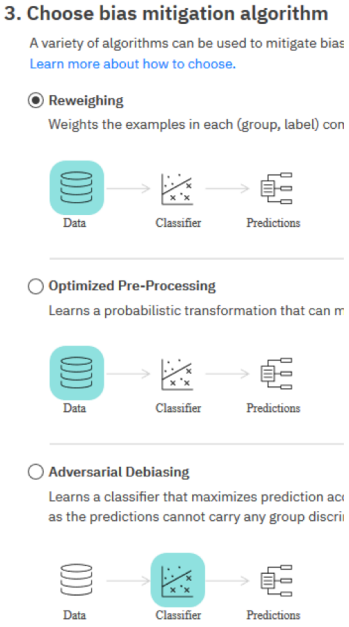
Entirely internal to the technical operations of an AI product or feature

Focuses attention on:

- Pre-processing of data
- The algorithm, accuracy/fairness tradeoffs
- Statistical fixes to outcomes to minimize disparate effects

IBM Fairness 360™

- Auto-generates a suite of fairness metrics
- Applies a statistical 'solution' from a menu of possible fixes
- Updates metrics




```
In [3]: metric_orig_train = BinaryLabelDatasetMetric(dataset_orig_train,
                                                    unprivileged_groups=unprivileged_groups,
                                                    privileged_groups=privileged_groups)
        display(Markdown("#### Original training dataset"))
        print("Difference in mean outcomes between unprivileged and privileged groups = %f" % metric_orig_train.mean_difference())
```


https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/

Datasheets for Datasets

Motivation for dataset creation

Dataset composition

- How many instances of how many different categories?
- Is the data source static or dynamic?

Data collection process

- Was data collected with sensors? Scraped from a public source like Reddit?
- When was the data collected?
- Is it exhaustive or a sample of a larger set of available data?

Data Preprocessing

- What cleaning/preprocessing was done?
- Is un-preprocessed data available?

Data Distribution

- What license is data distributed under?

Dataset Maintenance

- How often is data updated?

Legal and Ethical Considerations

- Was data collected with informed consent?
- Could it expose people to harm or legal action?
- Offensive or inappropriate content?

https://arxiv.org/abs/1803.09010

Model Cards for Model Reporting

Model Details

- Responsible organization
- Model data/version/type
- Training parameters
- License
- Contact information

Intended Use

- Primary intended uses/users
- Out-of-scope use cases

Factors - social categories legible to the model

Metrics

- Model performance measures
- Decision thresholds

Evaluation Data

- Datasets used
- Preprocessing details

Training Data

Quantitative analysis

- Fairness and accuracy metrics
- Precision, Recall, etc.

Ethical Considerations

Caveats and Recommendations

https://arxiv.org/abs/1810.03993

Systems Perspective

Systems Perspectives take into account where data comes into the product and how it flows out to produce changes in the real world.

Consequence Scanning

1. What are the intended and unintended consequences of this product or feature?
2. What are the positive consequences we want to focus on?
3. What are the consequences we want to mitigate?

Consequence Scanning

1. What are the intended and unintended consequences of this product or feature?
2. What are the positive consequences we want to focus on?
3. What are the consequences we want to mitigate?

- **Act**: Consequences within the control of the participants to act upon.
- **Influence**: Consequences which are out of your control but you can influence the outcome of.
- **Monitor**: Consequences completely beyond your control, but could affect your product and so you should understand better and monitor.

https://www.doteveryone.org.uk/project/consequence-scanning/

Impact Assessment

"To direct the FTC to require entities … to conduct automated decision system impact assessments and data protection impact assessments"

Socio-Technical Perspective

Technology is not neutral

Human action shapes technology

Technology affords humans certain capabilities to act differently within society

Socio-Technical Perspective

- Data and data products may not scale as easily and well as assumed
- Requires processes to identify missing voices and engage advocacy groups
- UX can be a mechanism for bringing in different voices
- Requires bringing one's "whole self" to work
- Embraces the full scope of diversity and inclusion, not thin versions of D&I

Ethics in Organizations

Owning Ethics

1. New Hires, New Job Titles
    - Tech companies are hiring high-ranking positions with "ethics" in the job title.
2. Higher Visibility
    - Ethics panels, workshops, and conferences are now a standard part of the tech industry ecosystem.
3. Move to Implementation
    - There has been a clear shift from arguing that "ethics is important" to establishing concrete practices.

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Silicon Valley Logics

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Silicon Valley Logics

- Meritocracy

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Silicon Valley Logics

- Meritocracy
- Tech Solutionism

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Silicon Valley Logics

- Meritocracy
- Tech Solutionism
- Market Fundamentalism

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External
- Upside vs. Downside

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External
- Upside vs. Downside
- B2B vs. B2C vs. Hardware

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External
- Upside vs. Downside
- B2B vs. B2C vs. Hardware
- Product vs. Legal

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External
- Upside vs. Downside
- B2B vs. B2C vs. Hardware
- Product vs. Legal
- Individual vs. Corporate

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External
- Upside vs. Downside
- B2B vs. B2C vs. Hardware
- Product vs. Legal
- Individual vs. Corporate
- Measurable vs. Non-measurable Impacts

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Key Tensions

- Internal vs. External
- Upside vs. Downside
- B2B vs. B2C vs. Hardware
- Product vs. Legal
- Individual vs. Corporate
- Measurable vs. Non-measurable Impacts
- User vs. Non-user Impacts

https://datasociety.net/output/owning-ethics-corporate-logics-silicon-valley-and-the-institutionalization-of-ethics/

Case Studies

[Part 2]

Case Studies [Part 2]

For each case:

- How would you address risks from a Technical Perspective?
- How would you address risks from a Systems Perspective?
- How would you address risks from a Socio-Technical Perspective?

What in your view are good mechanisms to identify and mitigate risk?

How should decisions be made about how to identify and mitigate risk?

Discussion

Building towards a robust practice that ...

How might we *decrease the likelihood* that organization that work with data and develop technology build products that *accidentally* harm others?