Best Practices in Ethical Data Collection

Towards Identity & Accountability

Victor Zimmermann

KÖNIGSWEG

PyData Berlin New Year 2021 Meetup

Personal primer.

I am not the most qualified person to talk about this subject.

- I am a white man from a somewhat privileged background.
- I have not experienced discrimination first-hand.
- I come to this with a pretty removed point of view.

As a result, there may be some distance between what we discuss here, and the actual impact of the measures presented.
Every point made here should be considered within the context of the millions of people affected by algorithmic decision making every day.
(Also: I am not a philosopher, so please try not to be mean if I misrepresent your favourite theory.)

Here are a few things that are important, but not part of this talk:

- This talk is not about machine ethics.
  (AIs as moral decision makers, e.g. self-driving cars. [Mis18])
- This talk is not about algorithmic bias.
  (How an algorithm interacts with and amplifies inequality.)
- This talk is not about algorithmic regulation.
  (Should there be government oversight in data science?)
- This talk is not about bullshit.
  (''Medium, message and misinformation.'' [BW20])

The Basics

- Theory of moral behaviour.
- Ethics is a discursive field.
- "[Ethics] invites us to think about what kind of people we want to be. [...] But we also have to learn how to live with the consequences of deciding wrong" [GKZ19]
- Three perspectives for this talk:
  - Teleological ethics.
  - Deontology.
  - Virtue ethics.

- Teleological ethics:
  "Objectives and purposes, impact assessment, costs and benefits."[1]
  Documentation, data sheets and impact statements.
- Deontological ethics:
  "Duties and rights, intrinsic moral values."[1]
  Red lines, regulation, ethics codes.
- Virtue ethics:
  "Striving for a good, happy life."[1]
  Embedded ethics, open source, aspirations.

[1] All definitions translated from [GKZ19].

Consequential Collections

- All data is biased.
- Machine learning is really good in finding bias.
- Historical (and inherent!) biases are still represented by proxy in "unproblematic" data sets.
- Documentation and transparency of collection methods can help mitigate unfair outcomes down the line.

- Deliberate documentation of data collection process.
- Parallel to spec sheets for electronic components, provide data sheets for data sets.
- General machine learning data set survey: [Geb+20]
- Specific natural language processing questions: [BF18]
- Questions should be answered before, during and after data collection.

- **Motivation**

  ″For what purpose was the data set created?″

  ″Who funded the creation of the data set?″

- **Composition**

  ″Does the data set contain all possible instances or is it a sample(not necessarily random) of instances from a larger set?″

  ″Are there recommended data splits (e.g., training, development/validation, testing)?″

  ″Are there any errors, sources of noise, or redundancies in the data set?″

  ″Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the data set?″

- Collection Process

  ″Who was involved in the data collection process (e.g., students,crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?″

  ″Does the data set relate to people?″

  ″Did the individuals in question consent to the collection and use of their data?″
- Preprocessing/cleaning/labeling

  ″Was the ′raw′ data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?″
- Uses

  ″What (other) tasks could the data set be used for?″

  ″Are there tasks for which the data set should not be used?″

- **Distribution**
  "Will the data set be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the data set was created?"
- **Maintenance**
  "Who is supporting/hosting/maintaining the data set?"
  "If others want to extend/augment/build on/contribute to the data set, is there a mechanism for them to do so?"

- Standardised documentation for data sets (in the form of data sheets or alternative formats) asks the right questions at the right time.
- It prevents miscommunication by forcing you to anticipate problems down the line.
- You know the data better than anyone working with it, so make sure that knowledge is not lost.
- By providing as much information as possible, you allow *(force)* users to make their own value judgements, instead of putting out fires until their standards are met.

- Most major machine learning conferences (e.g. NeurIPS) started to ask researchers to provide impact statements with their papers.
- Impact statements should fill a similar role to data sheets, so consult them whenever available.
  e.g. if you use a new(er) algorithm during preprocessing, check the impact statement for any unintended perturbations to the data that might affect your specific use case.

Deontological Data Science

Ethical Data Science from data sheets works under the assumption that more knowledge allows for a better calculation of (moral) costs and benefits.
It does not necessarily answer the following questions:

- What kind of red lines are there?
- How are they enforced? Are they enforceable?
- How do we as a profession want to be perceived? By clients? By society?

These questions guided the urgent rise to ethics codes in Medicine after WW2, and should still guide any researcher working with fundamentally human data.
So what would (or what should) an ethics code for data scientists look like?

Ethics codes: outward facing goals

Outward facing goals from Metcalf (2014):

1. **protect vulnerable populations who could be harmed by the profession's activities**
2. protect/enhance the good reputation of and trust for the profession
3. **establish the profession as a distinct moral community worthy of autonomy from external control and regulation**
4. **provide a basis for public expectations and evaluation of the profession**
5. serve as a basis for adjudicating disputes among members of the profession and between members and non-members
6. create institutions resilient in the face of external pressures
7. **respond to past harms done by the profession**

Protect vulnerable populations who could be harmed by the profession's activities.

- Viewing data protection as an ethical duty towards your customer.
- Inherent compliance with data protection standard vs. adjustment to meet baseline.
- Can you justify paying your Turkers starvation wages?

Establish the profession as a distinct moral community worthy of autonomy from external control and regulation.

- Promotion of ethics standards at conferences and in the wider profession.
- Support of programs like DISC, PyLadies that further diversity (and self-reflection!) in Data Science.

Provide a basis for public expectations and evaluation of the profession.

- Publish and promote ethics code.
- Open data and open source as default.

Respond to past harms done by the profession.

- Take a pick.
- Keep up to date, try to hire a diverse staff, acknowledge the size of the task…

Ethics codes: inward facing goals

Inward facing goals from Metcalf (2014):

1. **providing guidance when existing inexplicit norms and values are not sufficient, that is guidance for a novel situation**
2. reducing internal conflicts, that is, strengthen the sense of common purpose among members of the organization
3. satisfying internal criticism from members of profession
4. create generalized rules for individuals and organizations that have responsibilities for important human goods
5. establish role-specific guidelines that instantiate general principles as particular duties
6. establish standards of behavior toward colleagues, students/trainees, employees, employers, clients
7. strengthen the sense of common purpose among members of the organization
8. **deter unethical behavior by identifying sanctions and by creating an environment in which reporting unethical behavior is affirmed**
9. provide support for individuals when faced with pressures to behave in an unethical manner

Providing guidance when existing inexplicit norms and values are not sufficient, that is guidance for a novel situation.

- Continual review of ethics code.
- Establish a process that can act with some authority on ethical matters (that isn't HR).

Deter unethical behavior by identifying sanctions and by creating an environment in which reporting unethical behavior is affirmed.

- Establish consequences of unethical behaviour beforehand.
- Can someone report unethical behaviour without risking their job?
- Are sanctions helpful? (E.g. journalism codes often completely lack enforcement.)

Ethics codes

- No matter what project, what team, what company, you should have some ethics code that defines your values and red lines.
- If nothing else, this is preceded by 1.) talking about values you hold about your field, and 2.) agreeing on these values in your team, which is valuable in and of itself.
- This code does not need to meet every point in that list.
- An ethics code should be a mission statement first, regulation second. The goal is to make ethical decision-making a central part of the process, not another compliance step.

Virtual Virtues

- What is a good doctor?
- What is a good data scientist?
- A good doctor is not good because the profession and the daily procedures forces them to be.
- So how do we move from better processes and structures to better professionals?
- And what does a better professional look like if there are no clear accepted role models?

Ethics gap

- Five principles for AI in society: beneficence, nonmaleficence, autonomy, justice, and explicability.
- (Alignment with biomedical ethics, in contrast to tech & information ethics.)
- Still, "[t]he gap between principles and practice is large."[Mor+19]
- Ethical education of data scientist almost exclusively from case studies (e.g. Cambridge Analytica, bias in search engines/embeddings) [BR20]

- Treat data ethics as any other data science skill (i.e. coding, statistics, logical reasoning) that grows through training.
- Make a conscious (and concrete!) effort, to engage in virtuous behaviour.
- Practice data ethics on a day-to-day basis.

Developing code

Opportunities for ethical training when developing code [BR20]:

- Am I re-using other's code responsibly?

  Deliberations about sharing fosters reflection on citizenship, responsibilities, generosity.
- Am I providing the necessary credit for the work of others?

  Deliberation about providing credit fosters reflection on gratitude and humility.
- Am I going to share my code with others?

  Deliberation about misuse fosters reflection on responsibility to community and future users.
- Could my code be misused?

  Deliberation about misuse fosters reflection on societal responsibilities.

Contextual, ethical and 'big picture' ethical conundrums [BR20]:

- Individual, institutional, community ownership
  Who (should) own the code? (Open Science/open software)
- National and international copyright and patent
  How does one manage conflicting priorities? (Limits of ownership of digital artifacts)
- Coding community norms and requirements
  How does one manage conflicting priorities to self and communities? (Free and open source software)
- Social norms and requirements
  How does any code contribute to the broader body of the digital landscape? (FFP (fabrication, falsification, plagiarism) in research)

- Conscious documentation and ethics codes help to hold institutions accountable and force ethical considerations.
- Internalising and practicing the underlying virtues is much harder.
- Just like learning a language or playing the guitar, make everyday the conscious effort to become a *better* data scientist by reflecting on your work and making the effort.
- Engage in open data, read up on the issues in your field and find moral exemplars you can aspire to.

Thank you!

Slides, resources and contact info:

vz@koenigsweg.com

axtimhaus.eu

gitlab.com/axtimhaus

@dieaxtimhaus

linkedin.com/in/vzim

Further reading

- AI Ethics Guidelines Global Inventory (Algorithm Watch)
- Ethics Codes: History, Context and Challenges (Metcalf, 2014)
- Datasheets for Datasets (Gebru, et al., 2018)
- Timnit Gebru and Emily M. Bender in general
- Su Lin Blodgett, most notably Blodgett, et al. (2016)
- Applying Virtue to Ethics (Annas, 2014)

References

[BF18] Emily M Bender and Batya Friedman. "Data statements for natural language processing: Toward mitigating system bias and enabling better science". In: *Transactions of the Association for Computational Linguistics* 6 (2018), pp. 587–604.

[BR20] Louise Bezuidenhout and Emanuele Ratti. "What does it mean to embed ethics in data science? An integrative approach based on the microethics and virtues". In: *AI & SOCIETY* (2020), pp. 1–15.

[BW20] Carl T Bergstrom and Jevin D West. *Calling bullshit: the art of skepticism in a data-driven world*. Random House, 2020.

[Fac+20] Danica Facca, Maxwell J. Smith, Jacob Shelley, Daniel Lizotte, and Lorie Donelle. "Exploring the ethical issues in research using digital data collection strategies with minors: A scoping review". In: *PLOS ONE* 15.8 (Aug. 2020), pp. 1–17. URL: `https://doi.org/10.1371/journal.pone.0237875`.

[FC19] Luciano Floridi and Josh Cowls. "A unified framework of five principles for AI in society". In: *Harvard Data Science Review* 1.1 (2019).

[Geb+20] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III au2, and Kate Crawford. *Datasheets for Datasets*. 2020. arXiv: `1803.09010 [cs.DB]`.

[GKZ19] Petra Grimm, Tobias O. Keber, and Oliver Zöllner, eds. *Digitale Ethik: Leben in vernetzten Welten*. Reclam, 2019.

[Met14] Jacob Metcalf. ''Ethics codes: History, context, and challenges''. In: *Council for Big Data, Ethics, and Society* (2014).

[Mis18] Catrin Misselhorn. *Grundfragen der Maschinenethik*. Reclam, 2018.

[Mit+20] Margaret Mitchell, Dylan Baker, Nyalleng Moorosi, Emily Denton, Ben Hutchinson, Alex Hanna, Timnit Gebru, and Jamie Morgenstern. "Diversity and Inclusion Metrics in Subset Selection". In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. AIES '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 117–123. ISBN: 9781450371100. URL: `https://doi.org/10.1145/3375627.3375832`.

[Mor+19] Jessica Morley, Luciano Floridi, Libby Kinsey, and Anat Elhalal. "From what to how. An overview of AI ethics tools, methods and research to translate principles into practices". In: *arXiv preprint arXiv:1905.06876* (2019).