Introduction to Ethics in Data Science

Table of contents

1 Defining Ethics

“The subject matter of ethics is supplied by two intimately related terms, right and good” (Sharp 1928).

Defining Ethics

- Definition:
  - The judgment upon actions which are categorized as "**right**" or "**wrong**"
  - The societal use of the terms good and bad
- Ethics derived from the Greek word *ethos* – spirit of the community.
- Ethicists try to answer "What actions are regarded as right or wrong by the members of society?"

Establishment of Ethics in Science

- The study of ethics dates back to work of Socrates, Plato, and Aristotle
  - Hippocratic Oath: several medical principles such as ``do no harm'' that are still significant today, including consent.
- Ethics in experimental science was a global concern after WWII
  - Nuremberg Code (1947): A set of ethical research principles for human experimentation as a result of the Nazi Nuremberg trials.

💬 Possible Discussion Questions 💬

What daily work-related responsibilities do data scientists have that rely on making an ethical choice?

2 Ethical Concepts in Data Science

> Advancements in computer and data communications technology have resulted in the need to reevaluate the application of ethical principles. - (Weiss 1990) The XXII self-assessment: The ethics of computing

Data privacy

- Data privacy legislation differs across jurisdictions.
- *Data privacy* in the California Consumer Protection Act is the protection of personal information (e.g., Name, email address, purchase and browsing history, location and employment data, IP address, sensitive personal information).

Social justice

- *Social justice* in data science includes practices that promote equity and fairness, focusing on ensuring that data and reports accurately and adequately represent populations. Practicing reflexivity is especially important to identify personal bias while performing data science tasks.

Transparency

- *Transparency* relates to communicating with customers about data collection (design, amount, and duration of storage) and the ease with which other data scientists can access collected information. Transparency related to data is especially important for building trust with customers and other data practitioners.

Consent

- *Consent* is an agreement before data collection that lists a participant's rights to information about a research study or organization's information collection. Participants can terminate their participation and opt out of data use at any time. Changes to informed consent have made it easier for companies to obtain broad consent, which does not require re-consent for the re-use of data (Froomkin 2019) .

Accuracy

- *Accuracy* relates to the characteristics of high-quality data, which are highly complete, valid, and consistent.

Accountability

- *Accountability* relates to a data scientist's quality of work, which delineates decisions and processes to create reproducible work that follows company guidelines and adheres to laws.

Examples

What ethical concepts are relevant to these examples?

- Recently, **Uber** was fined millions by the Netherlands government after a watchdog found that the U.S. company was transferring data about drivers from the Netherlands to the U.S.
- **Target** was sued after a father was able to determine that his daughter was pregnant. A predictive algorithm determined that she was likely pregnant given the items she purchased and sent coupons to the family address for expecting mothers.

Ethics in Data Science

- Greater public access of computers raised ethical concerns
- Ethical issues related to computers (Weiss 1990):
  - Ownership of assets/material (e.g., programs, code, visual material)
  - The degree that programs, software, and computer users are held responsible for outcomes

3 Principles and Legal Frameworks

Legal Frameworks which Regulate Data Science

- There is variation in the laws across jurisdictions relating to privacy protection and permissible data usage.
- Two fundamental common themes in EU and US legislation which are fundamental to data protection :
  - Anti-discrimination rights
  - Personal data protection rights

Legal Frameworks which Regulate Data Science

- Anti-discrimination Rights in the U.S.;
- The Civil Rights Act of 1964 prohibits discrimination based on color, race, sex, religion, or nationality;
- The Disabilities Act of 1990 prohibits discrimination based on disability.

Legal Frameworks which Regulate Data Science

- Personal data protection:
  - In the United States, the Fair Information Practice Principles (1973) delineate principles that agencies use when evaluating information systems, processes, programs, and activities that affect individual privacy
- Since 2017, an increase in proposed laws and regulations related to AI have been proposed
- e.g., The US's Blueprint for an AI Bill of Rights

Data Ethics Principles

- To provide broad guidance,various organizations, e.g., ASA, (American Statistical Association 2022) and government agencies have proposed numerous data ethics principles.
- These principles vary in number, scope, and practicality
- Different corporations may chose to uphold the principles that meet and fit the company's culture due to the **lack of federal legislation**.

Data Ethics Principles (OECD)

Data Ethics Principles (OECD)

1. *Collection Limitation Principle*: There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
2. *Data Quality Principle*: Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

Data Ethics Principles (OECD)

Data Ethics Principles (OECD)

5. *Security Safeguards Principle*: Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.
6. *Openness Principle*: There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity and usual residence of the data controller.

Data Ethics Principles (OECD)

7. *Individual Participation Principle*: An individual should have the right to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him; and to have communicated to him, data relating to him.
8. *Accountability Principle*: A data controller should be accountable for complying with measures which give effect to the principles stated above.

Case Study

In 2017, **Meta** announced that they analyze and flag user-generated content related to suicidal intent or risk (e.g., post and comments). Later, Meta announced that the company used the suicide predictions to notify local authorities and that over 100 wellness checks were conducted by law-enforcement. Meta also announced that the suicide detection program would expand across all other countries. In 2018, Meta revealed that they also scan the content of users' private messages for suicide risk.

💬 Possible Discussion Questions 💬

- In pairs or smaller groups discuss the ethical issues in the previous example or potential for future dilemmas.

4 Ethical Practice

Ethical Practices and Considerations

- Remain **honest** about level of competence related skill, content/subject, potential bias, and timeliness.
- **Understand** responsibilities to ensure that the stakeholders' (i.e., clients and participants) needs are addressed.
- **Communicate** the limitations of the research design and analytic approach. Clearly convey steps, logic, and findings so that other data practitioners and the public can understand them.
- Engage in data science practices processes that **protect** humans and animals and minimize harmful impact.
- **Acknowledge** the work of other data practitioners and researchers that was used.
- **Report** ethical misconduct to the appropriate authority, which will investigate ethical issues and or violations.

💬 Possible Discussion Questions 💬

Algorithms affect the lives of individuals, sometimes in profound ways.

- What are example of algorithms that you've encountered in your life which have influenced your life or your behavior?
- What are the ethical practices and considerations that you think the creator of this algorithm(s) did or did not take into account?

References

American Statistical Association. 2022. “Ethical Guidelines for Statistical Practice.”

Froomkin, A Michael. 2019. “Big Data: Destroyer of Informed Consent.” *BIG DATA*.

OECD. 2002. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD. https://doi.org/10.1787/9789264196391-en.

Sharp, Frank C. 1928. “The Problem of Ethics.” In, 3–20. The Century Philosophy Series. New York, NY: The Century Co.

Weiss, Eric A. 1990. “The XXII Self-Assessment: The Ethics of Computing.” *Communications of the ACM* 33 (11): 110–32. https://doi.org/10.1145/92755.92780.