Cheatography

Community Principles on Ethical Data Practices Cheat Sheet

by [deleted] via cheatography.com/2754/cs/17342/

Introduction

The Community Principles on Ethical Data Practices are being developed by people from the data science community in conjunction with data science organizations. These principles focus on defining ethical and responsible behaviors for sourcing, sharing and implementing data in a manner that will cause no harm and maximize positive impact. The goal of this initiative is to develop a community-driven code of ethics for data collection, sharing and utilization that provides people in the data science community a standard set of easily digestible, recognizable principles for guiding their behaviors..

Source: https://datapractices.org/community-principles-on-ethical-data-sharing/

Values

Fairness: Understand, mitigate and communicate the presence of bias in both data practice and consumption.

Benefit: Set people before data and be responsible for maximizing social benefit and minimizing harm.

Openness: Practice humility and openness. Transparent practices, community engagement, and responsible communications are an integral part of data ethics.

Reliability: Ensure that every effort is made to glean a complete understanding of what is contained within data, where it came from, and how it was created. Extend this effort for future users of all data and derivative data.

Principles

1. Consider (if not collect) informed and purposeful consent of data subjects for all projects, and discard resulting data when that consent expires.
2. Make best effort to guarantee the security of data, subjects, and algorithms to prevent unauthorized access, policy violations, tampering, or other harm or actions outside the data subjects' consent.
3. Make best effort to protect anonymous data subjects, and any associated data, against any attempts to reverse-engineer, de-anonymize, or otherwise expose confidential information.
    - This includes all intermediate results, working with individuals or companies to help them maintain the anonymity of all data and parties involved, and supporting the rights to explanation, recourse, and rectification for any data subjects impacted by data work.
4. Practice responsible transparency as the default where possible, throughout the entire data lifecycle.
    - This includes providing enough context and documentation to enable other trained practitioners to understand and evaluate the use of data.

Social Responsibility



Sponsored by **ApolloPad.com**
Everyone has a novel in them. Finish Yours!
https://apollopad.com

Principles Continued

5. Foster diversity by making efforts to ensure inclusion of participants, representation of viewpoints and communities, and openness. The data community should be open to, welcoming of, and inclusive of people from diverse backgrounds.

■ This can be achieved by: being conscious of, and owning the results of actions, regardless of intent; promoting the voices of marginalized groups; acknowledging and self-checking privilege; accepting checks of privilege by others in good faith, and using privilege to advocate for equity.

■ The data community will not remain silent when witnessing others behaving in a manner that is not accessible, open, welcoming and inclusive.

6. Acknowledge and mitigate unfair bias throughout all aspects of data work.

■ This includes but is not limited to providing details and methodologies around data collection, processing and storage, and actively working to identify and disclose bias in algorithms, training data, and test data.

7. Hold up datasets with clearly established provenance as the expected norm, rather than the exception.

■ As a data collector, be responsible for recording provenance; as a data publisher, be responsible for propagating provenance; as a data scientist, be responsible for reviewing, considering, and declaring what is known about data provenance.

■ Provenance is a living part of data work and can evolve with the project and all reasonable efforts should be made to understand and pass on provenance work.

8. Respect relevant tensions of all stakeholders as it relates to privacy and data ownership.

9. Take great care to communicate responsibly and accessibly.

Principles Continued (cont)

■ This includes: acknowledging and disclosing caveats and limitations to the process and outputs; considering and providing clear opportunities for feedback from all stakeholders; considering and discussing whether something should be done (not just if it can be done); and clearly communicating who may be impacted, and how they are impacted, in order to minimize any potential harm from data work.

10. Ensure that all data practitioners take responsibility for exercising ethical imagination in their work, including considering the implication of what came before and what may come after, and actively working to increase benefit and prevent harm to others.

Sponsored by **ApolloPad.com**
Everyone has a novel in them. Finish Yours!
https://apollopad.com