Data Ethics Checklist

Domain: Data Ethics
MADS SIADS593 Ethics – Case Study
By André Buser, Jul 2021

Table of Contents

Introduction

This **data ethics checklist** is designed to provide guidance for data science projects conducted within our departments.

To ensure ethical practices are adhered to, a checklist was chosen as the preferred method over an oath. This decision was made based on the drawbacks of oaths outlined by Loukides et al. (2018, chap. 2), such as their one-time nature and lack of connection to ethical principles and theories. Furthermore, Gawande's (2009) research demonstrated that the implementation of checklists can effectively eliminate errors in basic processes.

The checklist follows a **four-stage phased approach**, based on Paul Resnick's (2021) recommendations and the Cross-Industry Standard Process for Data Mining (CRISP-DM) reference model (Chapman et al., 2000, p. 10):

CRSIP-DM Project Phases	Simplified Phases
Business understanding	Problem formulation
Data understanding	Data collection and cleaning
Data preparation	
Modeling	Modeling and analysis
Evaluation	
Deployment	Presentation and deployment

To ensure that the checklist remains **user-friendly** and that **critical topics** are given priority, a **maximum of three** data ethics questions should be addressed and discussed at each project phase before proceeding to the next stage.

The following sources were consulted and considered to develop the checklist:

- SIADS 501: Being a Data Scientist (BDS) [Link]
- Ethics and Data Science (EDS), Chapter 2, Of Oaths and Checklists [Link]
- Data Ethics Framework (DEF), UK Government [Link]
- Community Principles on Ethical Data Practices (PEDP) [Link]
- Principles for Accountable Algorithms and a Social Impact Statement for Algorithms (The Fairness, Accountability, and Transparency in Machine Learning group (FAT/ML)) [Link]

Purpose

The primary aim of this checklist is to promote ethical data practices by addressing fundamental values and principles such as transparency, accountability, fairness, and explainability. These values and principles are based on established guidelines such as the UK “Data Ethics Framework” (2020) and "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms" (n.d.):

Transparency & Auditability

- Ensure complete, open, and understandable project documentation to make actions, processes, and data are available for review (“Data Ethics Framework,” 2020).
- Enable the review of the algorithm’s behavior through disclosure of complete and accurate information (“Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” n.d.).

Accountability & Responsibility

- Implement effective governance and oversight mechanisms for the project to ensure that the values and principles are followed.
- Document the project risks, decisions, and actions to improve transparency and auditability.

Fairness

- Mitigate **biases** that may influence the model’s outcome and ensure that the project and its outcomes respect the dignity of individuals, are just, non-discriminatory, and consistent with the public interest, including human rights and democratic values (“Data Ethics Framework,” 2020).
- Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g., race, sex, etc.) (“Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” n.d.).
- Ensure that
    - **federally protected classes** (“Protected Class,” n.d.),
    - **protected characteristics** (“Equality Act 2010,” 2010),
    - **special categories of personal data** (“Art. 9 GDPR – Processing of special categories of personal data,” n.d.),
    - or **proxies** of those,

  are not used or used with the required measures and protection because we want to judge persons involved by what they do and not by the circumstances they cannot control (Sandvig, 2021). If used, ensure to highlight, and explain the situation.

Explainability & Accuracy

- Ensure that algorithmic decisions, as well as any data driving those decisions, can be explained to end-users and other stakeholders in non-technical terms (“Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” n.d.).
- Identify, log, and articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst-case implications can be understood and

inform mitigation procedures (“Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” n.d.).

Checklist

Important: All questions should be considered in the context to identify, avoid, minimize, or mitigate potential harm to our auditees (divisions, functions, down to individuals). Incorrect conclusions and allegations must be minimized. All automated decision-making must be verified before or during an audit or security investigation.

Project Organization

Question:	**What senior/external oversight is there for your project?**
Principle:	Accountability & Responsibility
Rational:	Identify the required steering committee members to ensure proper support for escalations and decisions.
Concepts addressed:	Governance and Project Structure
Source:	Data Ethics Framework (DEF) [Link]

Question:	**What are the governance mechanisms that enable domain experts to challenge your project?**
Principle:	Accountability & Responsibility
Rational:	Clarify the governance and accountability right from the beginning incl. questions such as who has the power to decide on necessary changes to the algorithmic system during the design stage, pre-launch, and post-launch? Identify and engage with the required business partners in data privacy, information security, ethics, and compliance.
Concepts addressed:	Governance and Project Structure
Source:	Data Ethics Framework (DEF) [Link]

Problem Formulation Phase

Question:	**How are the results used and could the misuse of the data/algorithm or poor design of the project contribute to reinforcing social and ethical problems and inequalities?**
Principle:	Fairness
Rational:	To clarify the intended use of the results and to ensure we are heading in the right direction when developing the algorithm / executing the project. For example, is the right problem be addressed? To ensure that there is a clear articulation of the problem before the start of the project. Identifying potential risks and issues (harm) to individuals and/or groups at a very early stage of the project.
Concepts addressed:	**Harm without intent:** Could we introduce harm to individuals? Which group could face negative consequences because of the project? Consequences predictable by experts vs. discoverable by users.
Source:	Data Ethics Framework (DEF) [Link]

Question:	**What kind of mechanisms can you put in place to prevent this from happening?**
Principle:	Fairness, Accountability & Responsibility
Rational:	This allows for early planning and implementation of corrective or preventive actions. Risks and decisions (act, mitigate or accept) need to be documented.
Concepts addressed:	Harm without intent, bias toward the majority, consequences predictable by experts vs. discoverable by users
Source:	Data Ethics Framework (DEF) [Link]

Data Collection and Cleaning

Question:	**How was the collected/obtained?**
Principle:	Fairness, Accountability & Responsibility
Rational:	Clarify whether the data has been collected/obtained legally or within the terms and conditions. Have the users consented to the specific use of their data (in this project)? Be aware of (statistical) bias types during data collection and sampling.
Concepts addressed:	Informed consent, sampling bias
Source:	Data Ethics Framework (DEF) [Link]

Question:	**What sources of error do you have identified during the data cleaning and how will you mitigate their effect?**
Principle:	Fairness, Transparency & Auditability
Rational:	To be specified.
Concepts addressed:	Bias toward the majority, challenges of classification
Source:	Principles for Accountable Algorithms and a Social Impact Statement for Algorithms (FAT/ML) [Link]

Question:	**Has altered data (during data cleaning) been labelled?**
Principle:	Transparency & Auditability
Rational:	It is important to flag the changed data to make the process more transparent and to increase reproducibility.
Concepts addressed:	Reproducibility
Source:	SIADS 501: Being a Data Scientist (BDS) [Link]

Modeling and Analysis

Question:	**How has the data been used to train a model been assessed for potential bias?**
Principle:	Transparency & Auditability
Rational:	Consider whether the data might (accurately) reflect **the biased historical practice** that you do not want to replicate in the model (historical bias). The data might be a biased **misrepresentation of historical practice**, e.g., because only specific categories of data were recorded correctly in a format accessible to the project (selection bias).
Concepts addressed:	Bias toward the majority, challenges of classification, cumulative disadvantage
Source:	Data Ethics Framework (DEF) [Link]

Question:	**What measures have been taken to mitigate bias?**
Principle:	Fairness, Transparency & Auditability
Rational:	The model can be tested for fairness concerning different user groups and disparate error rates among different user groups. Google's What-If tool can be used for that.
Concepts addressed:	Bias toward the majority, challenges of classification, cumulative disadvantage, Google's What-If tool
Source:	Data Ethics Framework (DEF) [Link]

Presentation and Deployment

Question:	**How much of your system / algorithm can you explain to your users and stakeholders?**
Principle:	Explainability & Accuracy
Rational:	To be specified.
Concepts addressed:	Bias toward the majority, challenges of classification
Source:	Principles for Accountable Algorithms and a Social Impact Statement for Algorithms (FAT/ML) [Link]

Question:	**How much of the data sources can you disclose?**
Principle:	Explainability & Accuracy
Rational:	To be specified.
Concepts addressed:	Bias toward the majority, challenges of classification
Source:	Principles for Accountable Algorithms and a Social Impact Statement for Algorithms (FAT/ML) [Link]

Question:	**Has a lesson learned session been planned of conducted?**
Principle:	Explainability & Accuracy
Rational:	To share all the learnings from this project incl. risks, decisions, and how the various pitfalls/harms/issues/risks were identified and addressed.
Concepts addressed:	Knowledge sharing (new)
Source:	Data Ethics Framework (DEF) [Link]

References

Art. 9 GDPR – Processing of special categories of personal data, n.d. . Gen. Data Prot. Regul. GDPR. URL https://gdpr-info.eu/art-9-gdpr/ (accessed 7.26.21).

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., 2000. Cross-Industry Standard Process for Data Mining 76.

Data Ethics Framework [WWW Document], 2020. . GOV.UK. URL https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020 (accessed 7.26.21).

Equality Act 2010 [WWW Document], 2010. URL https://www.legislation.gov.uk/ukpga/2010/15/contents (accessed 7.26.21).

Loukides, M., Mason, H., Patil, D., 2018. Ethics and Data Science.

Principles for Accountable Algorithms and a Social Impact Statement for Algorithms [WWW Document], n.d. URL https://www.fatml.org/resources/principles-for-accountable-algorithms (accessed 7.26.21).

Protected Class [WWW Document], n.d. . Thomson Reuters Pract. Law. URL https://content.next.westlaw.com/Document/Ibb0a38daef0511e28578f7ccc38dcbee/View/FullText.html?transitionType=Default&contextData=(sc.Default) (accessed 7.12.21).

Resnick, P., 2021. Introduction to the Four Stage Pipeline - Lecture [WWW Document]. Coursera. URL https://www.coursera.org/learn/siads501/home/welcome (accessed 7.26.21).

Sandvig, C., 2021. Cumulative Disadvantage and Protected Classes [WWW Document]. Coursera. URL https://www.coursera.org/learn/siads503/home/week/2 (accessed 7.12.21).