Dimensions

Participatory activities for purposeful, transparent, and people-centric dataset documentation

Dimensions

Dimensions evaluate usefulness

DIMENSIONS

High-level descriptions of the different types of judgements readers make, yielding directional insights into the usefulness and readability of the Data Card.

Can this Data Card help readers arrive at an informed conclusion about the dataset?

Dimensions evaluate usefulness of the Data Card in making assessments about accountability / utility / quality / impact & consequences / risks & recommendations of the dataset.

Accountability

Statements that express different stakeholders' reflective, reasonable, and systematic decisions regarding the trust in the dataset.

EXAMPLE AREAS

Authorship	Responsibility
Maintenance	Intentions

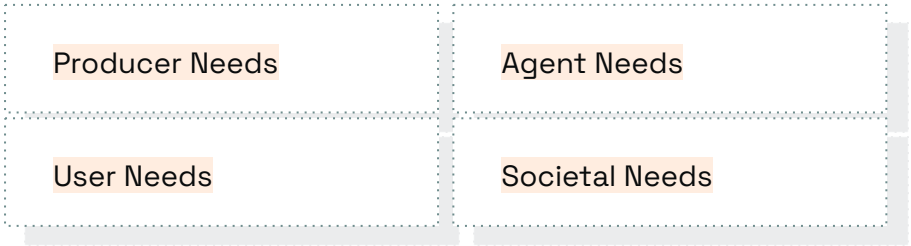
AS A [PERSPECTIVE], I WANT TO KNOW

–

- → about the publishers of the dataset
- → the access restrictions and policies of the dataset
- → the paradigm (updates, refreshes, additions to the data) of the dataset
- → explanations and motivations for creating the dataset

Utility

Provides details that satisfy readers' responsible decision-making process needs, and establish use cases suitability as it pertains to their goals.



AS A [PERSPECTIVE], I WANT TO KNOW

–

- → definitions and explanations for technical terms used in the documentation (metrics, scores, industry-specific terms, acronyms)
- → expectations around using the dataset with other datasets or tables (feature engineering, joining, sampling, comparative analysis, etc.)
- → the intended applications of the dataset

Quality

Summarizes the rigor, integrity and completeness of the dataset, communicated in a manner that's accessible to many readers.

AS A [PERSPECTIVE], I WANT TO KNOW

–

→ if there any known patterns (correlations, biases, skews) within the dataset

→ any dataset validation processes, explanations, and their results

→ what privacy and security measures were applied to the dataset

Validity	Reliability
Integrity	Reproducibility

Impact & Consequences

Information that help readers achieve their desirable outcomes when using and managing the dataset, acknowledging consequences that could negatively impact their goals.

AS A [PERSPECTIVE], I WANT TO KNOW

–

→ the past usage and associated performance of the dataset (eg. models trained)

→ policies associated with the dataset (e.g. licensing)

→ if there any known patterns (correlations, biases, skews) within the dataset

Efficacious	Group Benefit
Relevance	Implications

Risk and Recommendations

Makes readers aware of known and potential risks associated with the dataset, stemming from representation, use, or context of use.

AS A [PERSPECTIVE], I WANT TO KNOW

–

→ the safety (risks, limitations, and trade-offs) of using the dataset

→ any socio-cultural, geopolitical, or economic representation of people in the dataset

→ if there are attributes missing from the dataset or its documentation

Risks	Recommendations
Group Harms	Mitigation

The Data Cards Playbook ↗ by Google Research ↗ is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. You are free to share and adapt this work under the appropriate license terms ↗.