
The Data Cards Playbook

A toolkit for purposeful and people-centric dataset documentation for transparency in AI systems.

<https://pair-code.github.io/datacardsplaybook/>

#datacardsplaybook



THE DATA CARDS PLAYBOOK

Introduction

01 Ask

02 Inspect

03 Answer

04 Audit



THE DATA CARDS PLAYBOOK

Heuristics Evaluation

IN THIS SECTION

Review your template against accepted usability principles.

INSTRUCTIONS

As a group, evaluate your template using our heuristics worksheet. For each criteria, give a score between 0 to 10, and take note of where the questions, prompts, and the structure of the template can be improved.

OUTCOMES

Scores help introduce telemetry and stay on track, comments help capture context and action items. Use these to identify critical next steps in improving the usability of your Data Card template.

ACTIVITY LEVEL

Basic

Heuristics for Transparency Artifacts

Intelligible

The design and content of the template is effective, relevant, and easy-to-understand for a majority of experts and non-experts who are documenting the dataset.

Comprehensive

The template is easy for producers to create and express their mental models of the dataset, how it came into being, and what is important to know before using it.

Consistent

The template follows platform and industry conventions, maintains consistency within itself and across other similar templates, especially for implementing at scale.

Concise

The design and content of the template reduces vast and complex information into meaningful, digestible bits of relative importance that address the needs of both novices and experienced readers.



HEURISTIC:

Intelligible

The design and content of the template is effective, relevant, and easy-to-understand for a majority of experts and non-experts who are documenting the dataset.

CRITERIA:

Understandable

Information required can be easily understood by expert and non-expert producers.

 comments

Relevant

Prompts, guidelines, answer structures, and examples included in the template are relevant and actionable for a majority of producers.

 comments

Effective

A variety of producers can produce appropriate answers to a majority of questions in the template with little confusion.

 comments

Total Score
total / 30

 score: 0-10

 score: 0-10

 score: 0-10

HEURISTIC:

Comprehensive

The template is easy for producers to create and express their mental models of the dataset, how it came into being, and what is important to know before using it.

CRITERIA:

Purposeful

Template solicits information that establishes context for the dataset and related models, and is helpful to all stakeholders.

 comments

Complete

Questions are coherent and complete, appropriately describing all stages in a dataset (and related model's) lifecycle.

 comments

Depth

Questions and prompts are human-readable, and clearly state the fidelity of information or specification.

 comments

Total Score
total / 30

 score: 0-10

 score: 0-10

 score: 0-10

HEURISTIC:

Consistent

The template follows platform and industry conventions, maintains consistency within itself and across other similar templates, especially for implementing at scale.

CRITERIA:

Recognizable

Sections are organized in a logical order such that producers can recognize context and where to provide information.

 comments

Standard

Uses industry-standard terms and describes deviations from standards or offers additional context where relevant.

 comments

Clear

The same term means the same concept, every time it's used.

 comments

Total Score
total / 30

 score: 0-10

 score: 0-10

 score: 0-10

HEURISTIC:

Concise

The design and content of the template reduces vast and complex information into meaningful, digestible bits of relative importance that address the needs of both novices and experienced readers.

Total Score
total / 30

CRITERIA:

Graspable

The relative meaning and importance of keywords, key value pairs, and visual summaries are easy to grasp and therefore communicate.

 comments

 score: 0-10

Glanceable

The goals of questions and expectations of responses are clear at a glance.

 comments

 score: 0-10

Contextual

The template solicits background knowledge and context at abstractions designed for understanding, *and* without sacrificing the nature and nuance of the dataset.

 comments

 score: 0-10

Final checks

YOU SHOULD NOW HAVE

—

- ✔ Evaluated both content and design of your Data Card template for usability
- ✔ Identified quick-fixes and longer-term opportunities to refine your template
- ✔ Arrived at considerations for implementing your Data Card template at scale



#datacardsplaybook



[The Data Cards Playbook](#) [↗] is an adaptable toolkit of participatory activities, conceptual frameworks, and guidance that support Responsible AI practices for transparency in dataset documentation.

If you've adapted, implemented, or have feedback for this guidance, we'd love to hear from you at <https://github.com/pair-code/datacardsplaybook> [↗].

Find the complete playbook at
<https://pair-code.github.io/datacardsplaybook> [↗]



The [Data Cards Playbook](#) [↗] by [the People + AI Research Initiative](#) [↗] at [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](#) [↗].