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

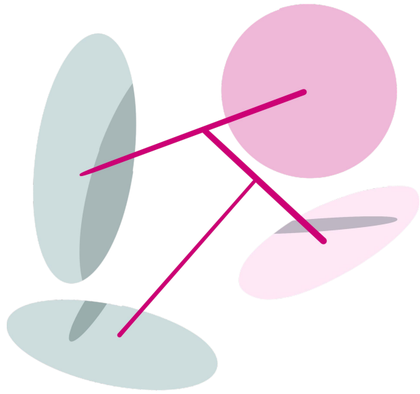
Introduction

01 Ask

02 Inspect

03 Answer

04 Audit



Define your Agent Information Journeys

IN THIS SECTION

Treat your Data Card like a product and define the most important information that it must successfully convey.

INSTRUCTIONS

Learn about what constitutes an Agent Information Journey. Brainstorm, share, and agree upon the top Agent Information Journeys that your Data Card should embody.

OUTCOMES

A initial, prioritized list of AIJs to include in your Data Card template.

ACTIVITY LEVEL

Basic

An Agent Information Journey (AIJ) is the combination of what an agent needs to know about a dataset to make a decision, and questions asked to arrive at an understanding of the dataset for that decision.

Agent

As a(n) ML engineer, I want to know about the source of the dataset, so I ask...

Information needed to make decision

... who owns the dataset?
... how was it collected?
... when was the data collected?
... what are the upstream sources of individual features?
... how was the data sampled?
... what were the inclusion and exclusion criteria?

Questions that establish an understanding about the source of the dataset.

Agent Information Journeys =

Perspective

Individuals reading the Data Card(s)

Think about:
Roles, agendas, foundations, and frameworks in domains.

Lens

Statements articulating what they want to learn from Data Card(s).

Think about:
Information needed to complete tasks and make decisions about the dataset and its use.

Scope

Follow-up questions that get readers closer to the Lens.

Think about:
Probing specifics about a dataset from multiple angles, such as a root cause analysis.

perspective

As a(n) ML engineer, I want to know
about the source of the dataset, so I ask…

lens

… who owns the dataset?
… how was it collected?
… when was the data collected?
… what are the upstream sources of individual features?
… how was the data sampled?
… what were the inclusion and exclusion criteria?

scopes

EXAMPLE

perspective

As a(n) ML engineer,

lens

I want to know the source of the dataset, so I ask…

… who owns the dataset?
… how was it collected?
… when was the data collected?
… what are the upstream sources of individual features?
… how was the data sampled?
… what were the inclusion and exclusion criteria?

scopes

EXAMPLE

perspective

As a(n) data analyst,

lens

I want to know the structure of the dataset, so I ask…

… what is the data format?
… what is modality of the dataset?
… how many features are there in the dataset?
… how many features are engineered?
… which features are strongly correlated?
… if there are any dependencies in the structure?

scopes

EXAMPLE

perspective

lens

As a(n) AI designer,
I want to know about the representation in the dataset, so I ask...

... is this data about target users?
... does it contain demographic information?
... which population groups are represented in the dataset?
... what are the most common represented groups?
... which groups are minorities in this dataset?
... does this data represent users of my product?

scopes

EXAMPLE

perspective

As a(n) policy aide,

lens

I want to know how the data might be misused, so I ask…

… what was the intended use of the dataset?
… what application prompted the dataset creation?
… what are known dangerous or risky applications of the dataset?
… what is the risk to specific groups?
… how do intended uses of this dataset impact constituencies?
… how can one ask for recourse?
… what safety nets can and should be build?

scopes

30
sec

Try it yourself

As a(n) _____,
I want to know _____
_____,
so I ask…

… _____?
… _____?
… _____?

perspective

lenses

scopes

12 min

Your Name: Write Top 6 AIJs of your Data Card(s)

✍️ As a [Perspective], I want to know [Lens], so I ask [Scopes]:

1. Question (Attribute)
2. Follow up Question (Evidence)
3. Follow up Question (Rationale)

✍️ As a [Perspective], I want to know [Lens], so I ask [Scopes]:

1.
2.
3.

✍️ As a [Perspective], I want to know [Lens], so I ask [Scopes]:

1.
2.
3.

✍️ As a [Perspective], I want to know [Lens], so I ask [Scopes]:

1.
2.
3.

✍️ As a [Perspective], I want to know [Lens], so I ask [Scopes]:

1.
2.
3.

✍️ As a [Perspective], I want to know [Lens], so I ask [Scopes]:

1.
2.
3.

15
min

Cluster your AIJs

Checklist

YOU SHOULD NOW HAVE

–

- ✔ A clear understanding of AIJs and how to frame them
- ✔ Explored the AIJs of all your agents
- ✔ Agreed upon a set of AIJs to include in your Data Card and measure over time

#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 ↗.