The Data Cards Playbook

Create purposeful, transparent,
and people-centric dataset documentation

https://pair-code.github.io/datacardsplaybook/
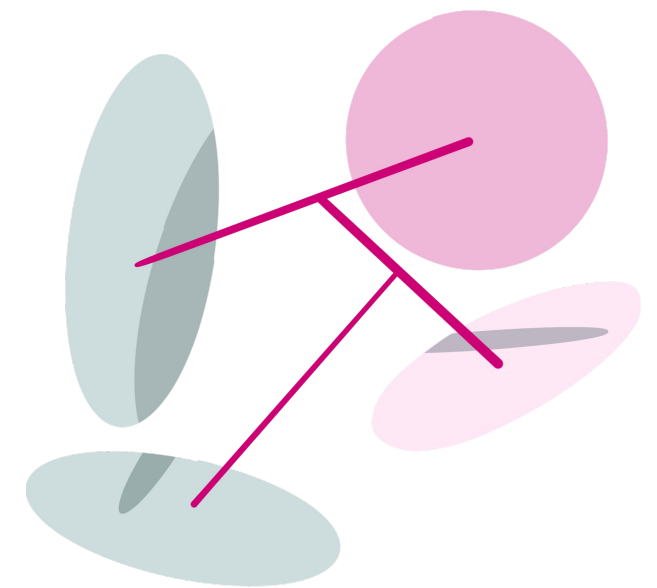
#datacardsplaybook

Questions with Optics

IN THIS SECTION Learn how to frame questions that can be used on a broad class of datasets to support agents in their information journeys.

INSTRUCTIONS

Understand the difference between Telescopic, Periscopic, and Microscopic Questions and enforce structure and relationships into questions to include in your Data Card(s).

OUTCOMES

Questions that agents might ask to arrive an understanding of the dataset to include in your Data Cards.

ACTIVITY LEVEL

Basic

Ask a series of questions in succession to make sense of datasets.

Use "Scopes" to frame questions and reveal obvious, non-obvious, visible, and invisible aspects of datasets. Scopes occur in sets of three: **telescopic, periscopic, and microscopic**.
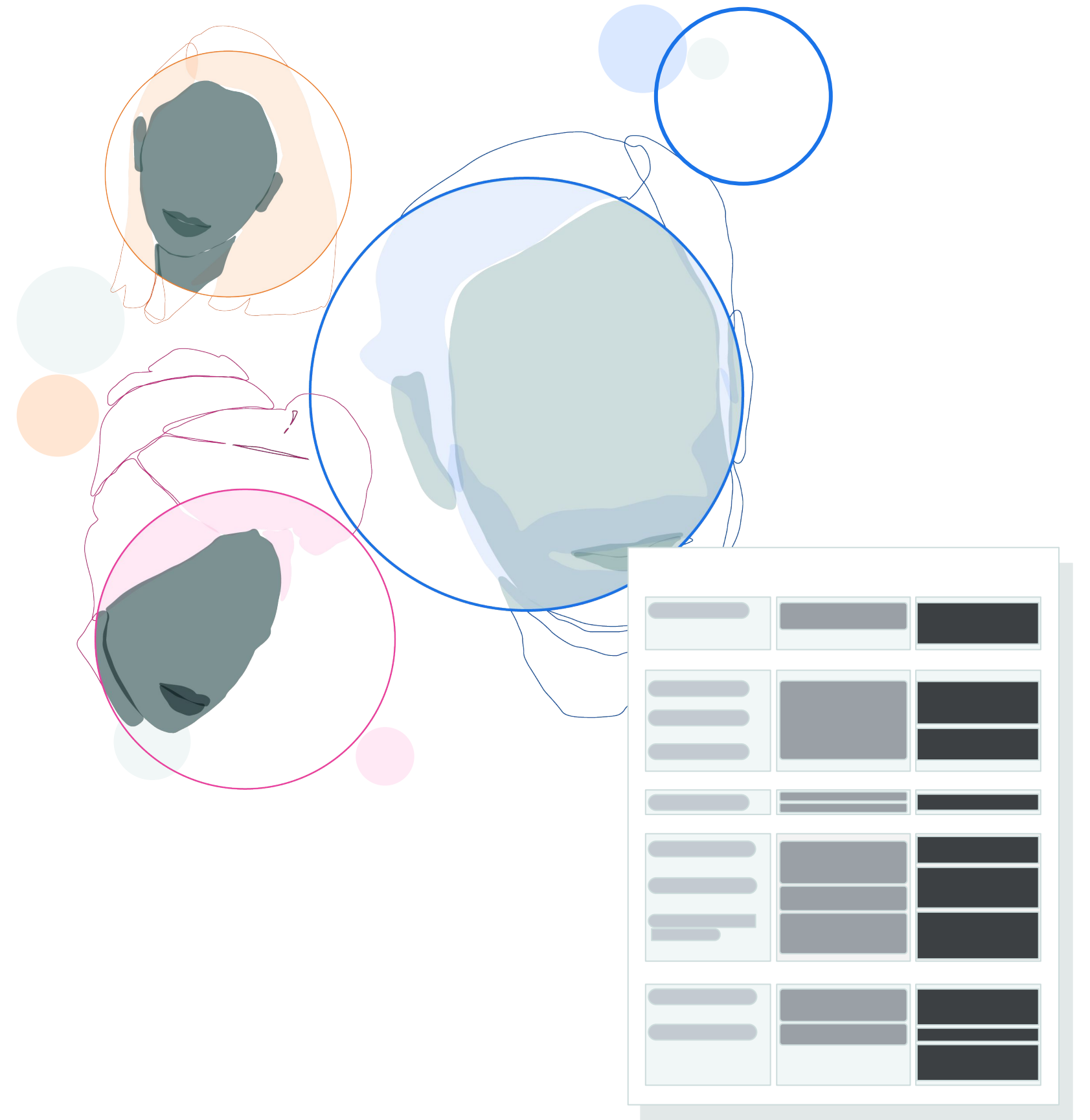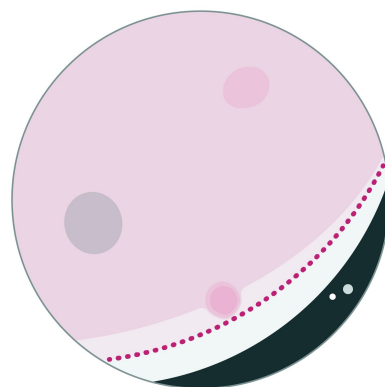
Types of scopes

Telescopic

Questions about attributes or characteristics commonly found across multiple datasets.

For example: Does this dataset **contain S/PII?**

Data Cards with just telescopes will describe obvious and not add any distinct value, but telescopes tell agents where to begin looking.
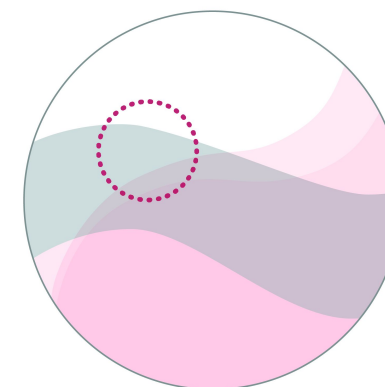
Periscopic

Questions about attributes or observations specific to the dataset being documented.

For example: Which **features in the dataset contain S/PII?**

Data Cards with just periscopes can get overly technical, without any details on context, relevance, or importance.
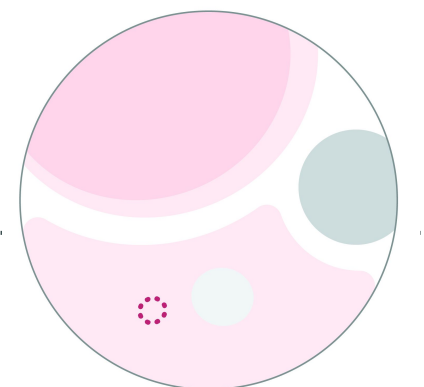
Microscopic

Questions about unobservable aspects of dataset(s), such as decisions, processes, impact.

For example: How was **S/PII handled** in this dataset?

If we only use microscopes, we can easily get lost in the details, and lose sight of the bigger picture.

Case in point:

TELESCOPIC
(global attributes)

(What is the type of Data in the Dataset?)
Primary Data Type
Tabular, Image, Audio, Speech, Time-series, Network, Graph, etc.

PERISCOPIC
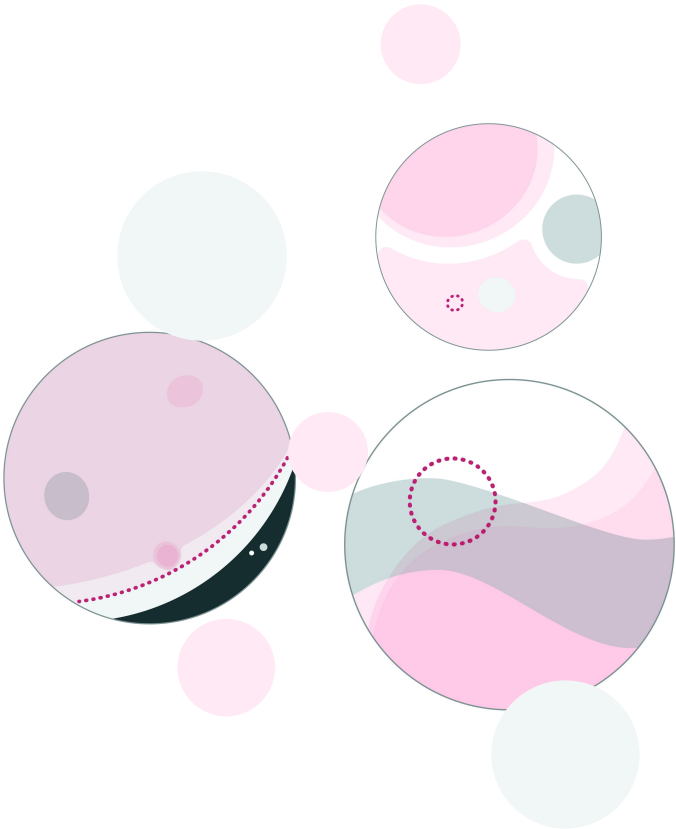(local observations, demands evidence)

(What is the breakdown of the data in the Dataset?)

Total Instances	478k+
Total Classes	6k+
Total Labels	1.27m+
Algorithmically Generated Labels	1.11m+
User Contributed Labels	505k+
Human Verified Labels	All labels verified
...	...

MICROSCOPIC
(unobservable explanations, demands rationale)

(What does the data in the Dataset describe?)
Nature Of Content Labeled images of objects (household goods, commercial products), vehicles, plants, animals and people (faces blurred).

(How was data included or excluded in the Dataset?)
Excluded Data All EXIF data including location has been removed.

(How was PII handled in the Dataset?)
Privacy PII associated with human subjects removed where possible.
To preserve demographic context, features are aggregated using ...

	TELESCOPIC	PERISCOPIC	MICROSCOPIC
Description	Characteristics	Observations	Explanations
Input Type	Yes/No or multiple choice responses	Key-Value pairs or short descriptions	Detailed, long-form descriptions
Format	Tags, Chips	Text, Tables, and Visualizations w/ Links	Paragraphs, Tables, Visualizations, Links
Content	Universal attributes of datasets	Unique, observable characteristics of the dataset	Rationales, decisions, and policies that shape datasets
Utility	Find information quickly; Indexical – search or scan Keywords, descriptors	Facilitate quick assessments Tactical – what statements	Provide context and links to additional information Rationale and process – why and how statements
Value	Provide an overview and help habituate agents in the Data Card. Make Data Cards easy to index in repositories	Often reproducible from the dataset, prevent re-work, and add more context for agents when stacked	Unobservable attributes and considerations important to agents, but accessible only to producers

As a **data scientist**, I want to know
about the structure of the dataset, so I ask…

… what is the data format?
… are there any media in the dataset?

`telescopic`

… how many features are in the dataset?
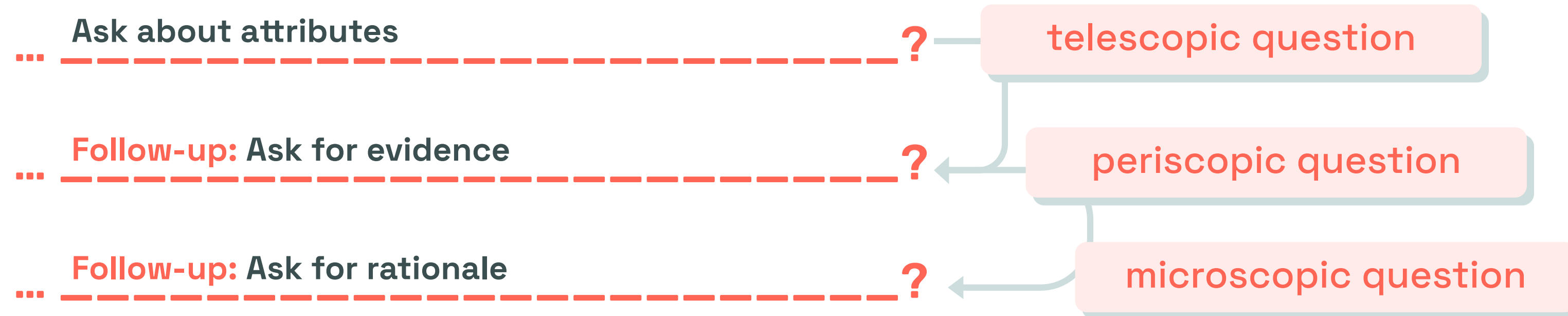… how many features are engineered?

`periscopic`

… which features are strongly correlated
and why? Was this intentional?

`microscopic`

Try it yourself

As a [your perspective], `perspective`
I want to know about the source(s) of the dataset, `lenses`
so I ask…

... **Ask about attributes** _____? — `telescopic question`

... **Follow-up: Ask for evidence** _____? — `periscopic question`

... **Follow-up: Ask for rationale** _____? — `microscopic question`

Give structure to your scopes

✍️ As a(n) [perspective], I want to know [lens]	Telescopic Question: ✍️ Expected Answers:	Periscopic Question: ✍️ Expected Answers:	Microscopic Question: ✍️ Expected Answers:
✍️ As a(n) [perspective], I want to know [lens]	Telescopic Question: ✍️ Expected Answers:	Periscopic Question: ✍️ Expected Answers:	Microscopic Question: ✍️ Expected Answers:
✍️ As a(n) [perspective], I want to know [lens]	Telescopic Question: ✍️ Expected Answers:	Periscopic Question: ✍️ Expected Answers:	Microscopic Question: ✍️ Expected Answers:

Checklist

✔ Familiar with the differences between scope types

✔ Able to break lenses down into sets of questions

✔ Have agreed upon your scopes and the expected answers

✔ Ready for the scopes brainstorming section

#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 ↗