

Lens Brainstorm

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





01 ASK • ACTIVITY

#datacardsplaybook

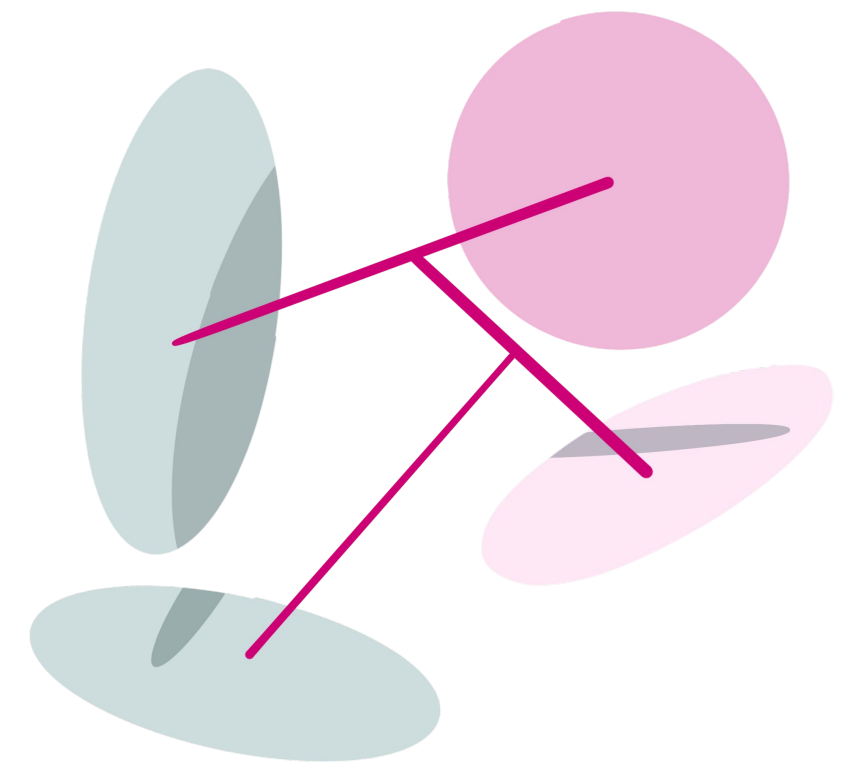
Lens Brainstorm

Prerequisite checklist

BEFORE YOU BEGIN

—

- Read through OFTEEn Framework
- Read through Agent Information Journeys
- Decide on your agents using Priority Matrix and/or Align on Agents
- Update results from aforementioned in the Activity Tracker for future reference



#datacardsplaybook

Lens - Origins

As a: <perspective>

I want to know

ORIGINS: A short, succinct statement that conveys a single idea about the early stages of a dataset's lifecycle when decisions to create a dataset are made.

Lens - Factuals

As a: <perspective>

I want to know

FACTUALS: A short, succinct statement that conveys a single idea about the process of collecting, labeling, rating examples, and corresponding outputs.

Lens - Transformations

As a: <perspective>

I want to know

TRANSFORMATIONS: A short, succinct statement that conveys a single idea about processes that make raw data usable, such as filtering, validating, parsing, formatting, and cleaning.

Lens - Experience

As a: <perspective>

I want to know

EXPERIENCE: A short, succinct statement that conveys a single idea about the use of the dataset in experimental, production, or research practice.

Lens - n=1

As a: <perspective>

I want to know

n = 1: A short, succinct statement that conveys a single idea about actual samples of data - in distribution, out-of-distribution, etc.

Lens - Origins

As a: ML data scientist

I want to know

If the publishers thought about fairness when deciding to make the dataset.

ORIGINS: A short, succinct statement that conveys a single idea about the early stages of a dataset's lifecycle when decisions to create a dataset are made.

Lens - Factuals

As a: ML data scientist

I want to know

If the dataset is large enough for my specific task.

FACTUALS: A short, succinct statement that conveys a single idea about the process of collecting, labeling, rating examples, and corresponding outputs.

Lens - Transformations

As a: ML data scientist

I want to know

If and how the data was aggregated.

TRANSFORMATIONS: A short, succinct statement that conveys a single idea about processes that make raw data usable, such as filtering, validating, parsing, formatting, and cleaning.

Lens - Experience

As a: ML data scientist

I want to know

If the dataset can be used for ensemble learning tasks.

EXPERIENCE: A short, succinct statement that conveys a single idea about the use of the dataset in experimental, production, or research practice.

Lens - n=1

As a: ML data scientist

I want to know

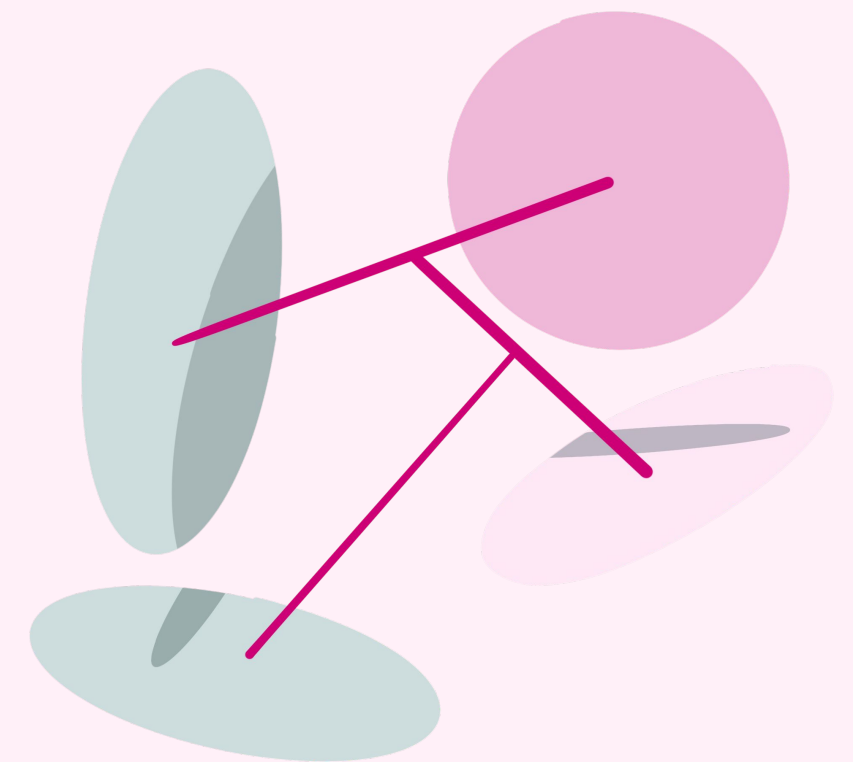
What types of examples are uncharacteristic of this dataset.

n = 1: A short, succinct statement that conveys a single idea about actual samples of data - in distribution, out-of-distribution, etc.

Lens Brainstorm

Create and prioritize Agent Information Journeys for your Data Card that can be tracked over time.

- Create perspective-lens pairs that capture your top agents' information needs
- Brainstorm lenses that are succinct and convey a single idea
- Use OFTEen prompts in Lens Cards to establish well-balanced foundations for your Data Card template



#datacardsplaybook

Brainstorm and prioritize lenses for your agents

Step 1: Review your agents

Step 2: Individually, fill out as many Lens Cards as possible for your agents

Step 3: Prioritize your lens cards and share with the group

Step 4: Cluster and vote on your most important lenses

Step 5: Transfer results to the Lens Brainstorm tab in your tracker

These will form the backbone of your Data Card.

Subsequent activities will expand these AIJs to include in your Data Card.

As a(n) [agent], I want to know [origins]	As a(n) [agent], I want to know [factuals]	As a(n) [agent], I want to know [transformations]	As a(n) [agent], I want to know [experience]	As a(n) [agent], I want to know [n=1]

Fill out as many Lens Cards as possible. Go over all OFTEⁿ stages, pushing past the obvious ideas.

A lens is a short succinct statement which conveys a single idea about:

ORIGINS: the early stages of a dataset's lifecycle when decisions to create a dataset are made.

FACTUALS: the process of collecting, labeling, rating examples, and corresponding outputs.

TRANSFORMATIONS: Processes that make raw data usable, such as filtering, validating, parsing, formatting, and cleaning.

EXPERIENCE: The use of the dataset in experimental, production, or research practice

n = 1: Actual samples of data - in distribution, out-of-distribution, etc.

Pro-tips for Lens Brainstorm



Pick one

Pick one OFTEEn category, even if they overlap; that's ok



Keep it grounded

Think about agents as groups of people; your lenses are about real people



Strive for high level

A sign of a good lens = can it be converted into a title?



One is wonderful

Each lens should communicate a single idea



Always be refining

Perfect, consolidate, and revise your lenses based on clusters and context

Next steps

OUTCOMES

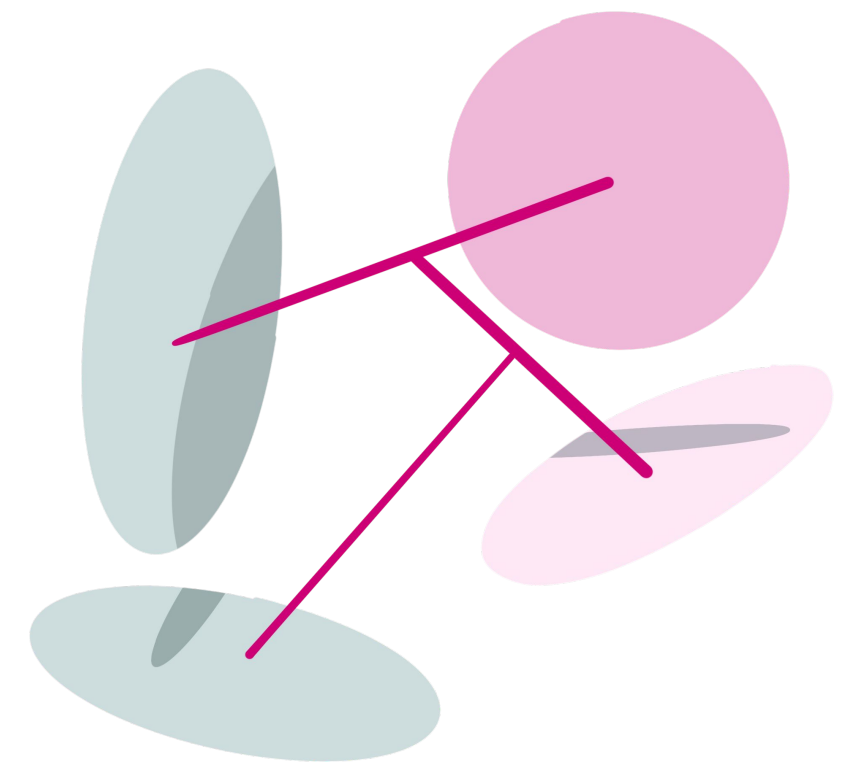
–

Define and decide on the most important information journeys of your Data Card or documentation.

FOLLOW UP

–

- Update your Activity Tracker
- Next: Scopes Brainstorm
- Later: Templatize



#datacardsplaybook



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 ↗](#).