
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

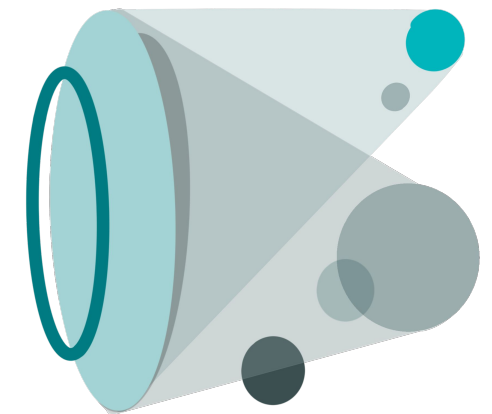
Introduction

01 Ask

02 Inspect

03 Answer

04 Audit



Root Cause Analysis

IN THIS SECTION

Use a fishbone diagram to work through different factors that can cause undesirable outcomes when your dataset is used.

INSTRUCTIONS

Systematically work through each factor in the fishbone diagram to either determine the cause or work your way up to an undesirable outcome stemming from the use of your dataset.

OUTCOMES

Use this worksheet to decide what needs to be documented so Data Card readers can navigate desirable and undesirable outcomes to use the dataset responsibly.

ACTIVITY LEVEL

Intermediate

Let's warm up

About this activity

A root cause analysis (RCA)

is a popular tool used to identify the root causes of problems that can then serve as a basis for corrective action, mitigations, or awareness.

Fishbone diagrams (or *Ishikawa* diagrams) are causal diagrams that help identify potential factors that cause an overall effect. They are popularly used in quality control processes.

Generative Use

How: Systematically work through each factor in the fishbone diagram to note key decisions, assumptions, criteria, tasks and factuials. Consider how these interact with each other to introduce an undesirable outcome when using the dataset.

Outcome: Decide what will need to be documented to help Data Card readers use the dataset responsibly and navigate these risks and outcomes.

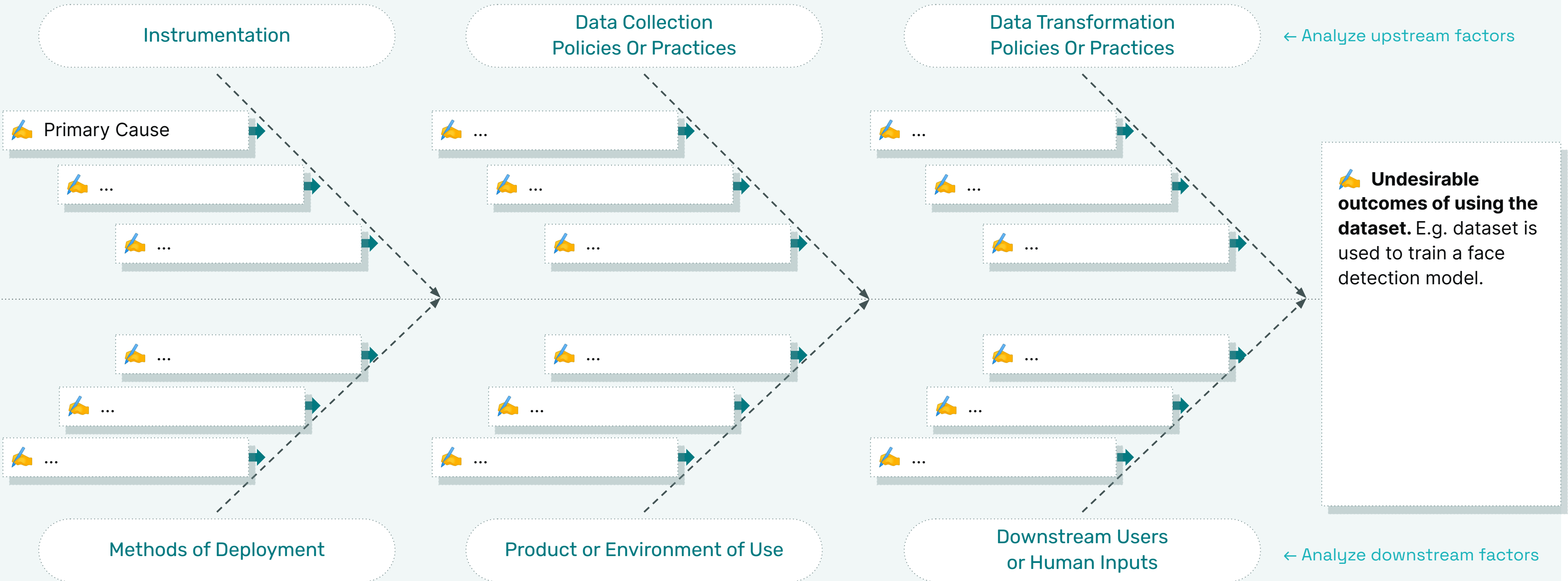
Deductive Use

How: Start with a known risk or undesirable outcome. Systematically review each factors to identify contributing factors and possible causes. This could be in the form of decisions about data structuring and design, prior research, assumptions made, or tasks executed.

Outcome: Decide what will need to be documented to help Data Card readers use the dataset responsibly and navigate these risks and outcomes.



1. Review all factors 2. Identify concerns and causes 3. Determine undesirable outcomes



1. Identify risk or undesirable outcomes

2. Review all factors


3. Brainstorm causes or concerns

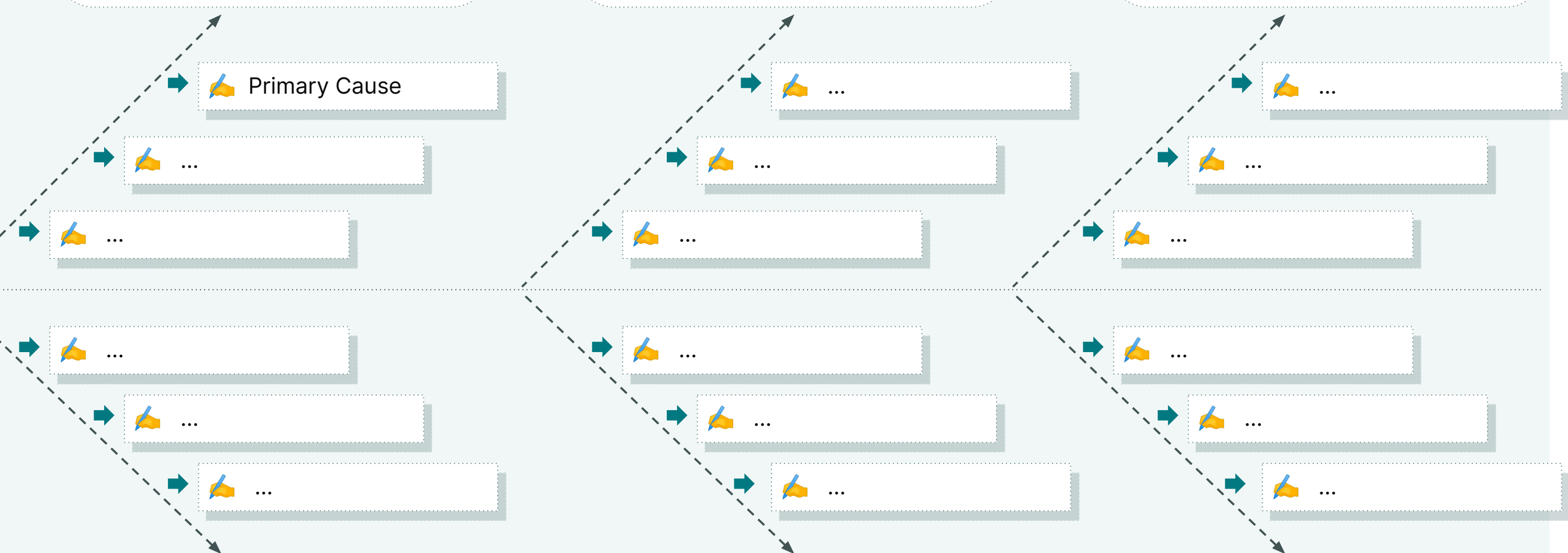
Analyze upstream factors →

Instrumentation

Data Collection Policies Or Practices

Data Transformation Policies Or Practices

 **Risk or Undesirable outcomes of using the dataset.**
E.g. dataset is used to train a face detection model.



Analyze downstream factors →

Methods of Deployment

Product or Environment of Use

Downstream Users or Human Inputs



Final checks

YOU SHOULD NOW HAVE

–

- ✔ Identified a set of undesirable outcomes
- ✔ Identified which factors contribute undesirable outcomes and how
- ✔ Important factors and undesirable outcomes that Data Card readers should be aware of



#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](#) [↗].