



MODULE ANSWER

Limitations, Benefits, and Uses

Activity

Level: Advanced

Duration: ⌚ ⌚ ⌚

Instructions

For this activity, invite individuals who are familiar with both the dataset and its use. Involve subject matter experts from the domain of the dataset, domain of applications, and other related fields, such as sociotechnical research, human-computer interaction, ethics, and policy.

This activity contains a list of seven attributes that you can use to articulate different benefits of your dataset, identify conditions that limit these benefits, and map each benefit-limitation pair to the intended uses of your dataset. Work through this table as a team to inform the content strategy of your Data Card.

Intrinsic & Contextual Attributes

A Data Card gives readers the necessary evidence and context clues so they understand the benefits and bounds of a dataset. In our work, we've found two kinds of "sources" of benefits and limitations. First, there exist intrinsic attributes that are a direct result of the data in a dataset and the transformations that may have been applied to them. The second kind are contextual attributes, which describe benefits and limitations that emerge due to using the dataset in a specific way, or in specific systems.

- **Attribute:** Use the attribute specified as a lens to identify the benefits of your dataset. In this worksheet, you'll find seven attributes. Intrinsic attributes include *People representation*, *Environments and Cultures*, and *Instrumentation and Probes*. Contextual attributes include *Purpose*, *Utility*, and *AI Cascades*. *Temporal Validity* is a unique attribute that is both intrinsic and contextual.
- **Benefit:** What are the benefits of the attribute in the dataset for the intended uses? What are the beneficial characteristics of your dataset that are related to this attribute?
- **Limitation:** What are the limitations of the attribute in the dataset for the intended uses? Under what quantitative, qualitative, or contextual conditions does the benefit cease to exist?
- **Use Cases:** Given the benefits and limitations, which use cases are suitable or unsuitable due to the attribute in the dataset?
- **Evidence:** What evidence, analysis, or information pertaining to this attribute should be provided in the Data Card so readers can make tactical decisions about using the dataset?
- **Actions:** If at all, what strategies can help dataset users overcome the limitations? Does this limitation affect the utility of the dataset, introduce failures or constraints, or does it mirror known errors in the dataset or model behaviors when trained or evaluated on the dataset?



People Representation

Type: Intrinsic Attribute

Description: Describes the effect of the dataset representing specific dimensions or attributes about some people, and not others.

Benefit	What are the benefits of the people represented in the dataset for the intended uses?
Limitation	What are the limitations of the people represented in the dataset for the intended uses?
Use Cases	Which of the intended use cases are suitable or unsuitable as a result of the people represented in the dataset?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?



Environments or Cultures

Type: Intrinsic Attribute

Description: Describes the effect of the dataset representing specific geographic and/or cultural environments.

Benefit	What are the benefits of the environments and cultures represented in the dataset for the intended uses?
Limitation	What are the limitations of the environments and cultures represented in the dataset in the dataset for the intended uses?
Use Cases	Which of the intended use cases are suitable or unsuitable as a result of the environments and cultures represented in the dataset?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?



Instrumentation and Probes

Type: Intrinsic Attribute

Description: Describes the effect of having collected the data using instruments that measure or probes that solicit information in very specific ways and fidelities.

Benefit	What are the benefits of the instrumentation and probes used to collect the data in the dataset for the intended uses?
Limitation	What are the limitations of the instrumentation and probes used to collect the data in the dataset for the intended uses?
Use Cases	Which of the intended use cases are suitable or unsuitable as a result of the instrumentation and probes used to collect the data in the dataset?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?



Temporal Validity

Type: Intrinsic and Contextual Attribute

Description: Describes the time, duration and circumstances in which the dataset is valid and accurate.

Benefit	What are the benefits of the time, duration and circumstances described by the dataset for the intended uses?
Limitation	What are the limitations of the time, duration and circumstances described by the dataset for the intended uses?
Use Cases	Which of the intended use cases are suitable or unsuitable as a result of the the time, duration and circumstances described by the dataset?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?

* Temporal validity is both an intrinsic attribute, for example, in a time-series dataset; and a contextual attribute, for example, a credit card fraud detection dataset that was released 10 years ago would no longer be valid because of changes in how people use credit cards.



Purpose

Type: Contextual Attribute

Description: Describes the effect of use cases that match or are closely related to or easily conflated with the intended use cases.

Benefit	What are the benefits of using the dataset for use cases that are closely related to the intended uses?
Limitation	What are the limitations of using the dataset for use cases that are easily conflated with the intended uses?
Use Cases	Based on benefits and limitations, which related or conflated use cases are suitable or unsuitable?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?



Utility

Type: Contextual Attribute

Description: Describes the effect of specific utilities in AI lifecycles, such as training, priming, testing or fine-tuning, etc.

Benefit	What are the benefits of using the dataset for model-specific utilities such as training, testing, validating, fine-tuning, or priming?
Limitation	What are the limitations of using the dataset for model-specific utilities such as training, testing, validating, fine-tuning, or priming?
Use Cases	Based on benefits and limitations, which model-specific utilities are suitable or unsuitable?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?



AI Cascades

Type: Contextual Attribute

Description: Describes the effect of known issues in AI lifecycles, such as model drifts, non-standard metrics, limitations of expertise, and organizational challenges.

Benefit	What are the benefits of using the dataset to address various AI cascades?
Limitation	What are the limitations of using the dataset to address various AI cascades?
Use Cases	Based on benefits and limitations, which AI cascades should readers be acutely aware of?
Evidence	What evidence, analysis, or information pertaining to this attribute can help readers can make tactical decisions about using the dataset?
Actions	If at all, what strategies can help dataset users overcome the limitations?

* AI Cascades is adapted from [Sambasivan, Nithya, et al. "“Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI.” *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021.](#)



Conclusion

- This activity can help you work through applications of your dataset in a systematic way. In this process, you might identify opportunities to improve your dataset.
- Use these worksheets in conjunction with The Data Cards Playbook’s transparency patterns to report on suitable use cases, unsuitable use cases, risks, limitations, and best practices in a Data Card.
- You do not need to include these worksheets in your Data Card.

The relationship between limitations and benefits	Datasets are bounded in their benefits. Readers will need help in understanding where benefits of datasets cease to exist so they can confidently decide what to use the dataset for (utility), and anticipate challenges when using it (usability).
---	--



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 ↗](#).

