**The Data Cards Playbook**

MODULE ANSWER | Foundation

# Avoid Vague or Non-answers

Write content that matches reader expectations and answers questions.

Mismatched expectations are a common difficulty that Data Card readers face. This happens when content doesn't exactly line up with the title or when an answer doesn't make sense in the context of the dataset or its location in the Data Card. Another mismatch in expectations occurs when a reader encounters a vague response.

Answers can be vague in all sorts of ways. For example, when a yes/no response doesn't reflect the broader context of the question. Or when readers see links to ambiguous documents that need special permissions to access or special knowledge to understand. Or when quantitative results don't explain how numbers came to be, what they represent, or why they are important.

Readers are more likely to turn away when they encounter vague or irrelevant responses. Conversely, readers are more likely to engage in fruitful discussion if they have been able to learn something significant from your Data Card. This is especially important if your readers already know your dataset. If they have attempted to use a dataset before unsuccessfully, they are more likely to implicitly distrust it. Or they may have built up a certain understanding from interactions with similar or older versions of datasets. When faced with a vague, ambiguous, or restrictive answer, readers are going to fill in the blanks with assumptions – or we're going to end up reinforcing incorrect beliefs about datasets and their use.

**People + AI Research**

*Readers are more likely to engage in fruitful discussion if they learn something from your Data Card, that they find useful in their work.*

## Key Takeaways

- Pay close attention to answers that are vague, incomplete, or seem to answer a different question.

- Missing information can introduce assumptions or worse, can reduce overall trust and accountability.

# Actions For Your Team

1. **Reflect the question.** Readers should be able to infer the questions that you are trying to answer from your response. Introduce clear cues (titles, subtitles, headings) that reflect these questions.

2. **Provide context.** Even a marginal reduction in uncertainty can be profoundly valuable for a decision. Provide descriptions, anecdotal or quantitative evidence, describe underlying assumptions and conditions to add context.

3. **Be transparent about known unknowns.** The absence of information is still information that can be factored in decisions about completeness and use of the AI system.

4. **Use N/A with care.** N/A is a broad term that is easily misinterpreted. Instead, consider using meaningful alternatives to express unknowns provided in the uncertainty table.

5. **When in doubt, use intended outcomes.** Open-ended or speculative answers can often be vague. Frame these answers from the perspective of the intended applications and motivations of the dataset.

# Resource

## Expressing Unknowns: Alternatives to N/A

This table describes terms that are useful alternatives to simply typing in N/A within a field on a Data Card. Use these according to the guidance provided, to add clarity to answers that you can't speak to. The follow up actions are useful strategies to use this language responsibly.

| Term | General Guidance | Follow-up actions |
|---|---|---|
| *Which term to answer a question with.* | *When to use it correctly use it.* | *How to responsibly navigate using this term.* |
| **N/A** | If the entire field is inapplicable to the model or dataset. | Suggest applicable equivalents |
| **Inconclusive** | If analyses were performed but results were inconclusive. | List the analyses performed and the corresponding hypotheses. |
| **Insignificant** | If analyses were performed but results were insignificant. | List the analyses performed and state the null hypothesis. |
| **Unknown** | If the conditions demanded by questions are untested by publishers. | Suggest alternative tests that could be performed; or list those that have been performed. |
| **Unavailable** | If information exists but is unavailable to the producer. | Provide reasons for unavailability |
| **Proprietary** | If specific information exists but cannot be published. | Provide a high-level description instead that would be safe for publishing |
| **In Progress** | If information is not ready to include yet | Provide notes on progress "In progress as of (date), expected completion (date)" |
| **NTK / Confidential** | Information cannot be provided as-is due to PII or other reasons | Provide reasons for confidentiality. Give a description or use a hypothetical to stand-in so access is requested only if needed. |

## Considerations

→ Can readers support their decisions with both evidence and counter-evidence from the Data Card?

→ Is there sufficient context for a simple understanding of your Dataset by experts and non-experts alike?

→ Is the granularity of answers suitable to the reader's domain expertise or data fluency levels?

→ Work with your team to build up a vocabulary around uncertainty. Start with using the uncertainty table, and take note of any other terms and follow-up actions that are specific to your team or datasets.

→ Schedule half-yearly check-ins to see if your data and model cards are still accurate, fresh, or require any updates!

---