Can AI Tell Why A User Got Stuck?

How Dovetail and Perplexity Interpret Usability Observations

Research Plan





By Theresa Wilkinson

Study Objectives

This study will evaluate how AI tools interpret a usability observation when they know what a participant did but have yet to hear why. It will examine whether the tools distinguish observed behavior from possible explanations, recognize when they need more information, and suggest a follow-up question that could help clarify what happened.

The study will then give each tool the participant's explanation and examine whether it revisits its initial interpretation and recommends a change that addresses the difficulty the participant described.

Research Questions

This study addresses the following research questions:

- When given only an observation of a participant's behavior, how do the AI tools distinguish what was observed from possible explanations?
- When given only the task and an observation, do the AI tools recognize that they don't have enough information to explain why the participant got stuck? What follow-up question would they ask the participant to clarify what happened?
- After receiving the participant's explanation, how do the AI tools update their initial interpretation?
- Do the changes the AI tools recommend address the specific difficulty the participant described?

Tools and Test Setup

The study will compare two AI tools across three fictional usability scenarios, for a total of six test runs. For each tool, the tool name, the model name or version shown in the interface, and the test date will be noted.

A fresh chat will start for each scenario in each tool. Both tools will receive the same scenario text and prompts in the same order. Within each run, an observation will be noted, and the participant's explanation will be shared in the same chat so the tool can revisit its initial response.

Full responses from both stages will be saved for the side-by-side comparison.

Usability Scenarios

The study will use three fictional usability scenarios. Each has a task, an observation, and a participant explanation. In the first stage, both AI tools will receive only the task and observation. The participant's explanation will be shared in the second stage.

Scenario 1 — Reporting an injury

- **Task:** Complete an online form to report a wrist injury.
- **Observation:** The participant pauses at the required "Date of injury" field, leaves it blank, and moves to the next field.
- **Participant's explanation:** "My wrist started hurting gradually. I didn't know whether to enter the day it first hurt or the day it got bad enough that I reported it."

Scenario 2 — Finding examples of past projects

- **Task:** Find photos of kitchens a remodeling company has completed.
- **Observation:** The participant opens "Services," scrolls down the kitchen remodeling page, and returns to the home page. They end the task without opening "Portfolio."
- **Participant's explanation:** "I saw 'Portfolio,' but I thought that was where the company listed its credentials. I expected the kitchen photos to be under 'Services.'"

Scenario 3 — Saving a change

- **Task:** Update a phone number in an online account.
- **Observation:** The participant enters a new phone number, clicks "Save," waits a few seconds, and clicks "Save" again.
- **Participant's explanation:** "I clicked again because I didn't see anything telling me the new number had been saved."

Test Procedure and Prompts

The following procedure will be used for each of the three scenarios in Dovetail and Perplexity. Each scenario will begin in a fresh conversation or isolated workspace so that earlier scenarios and participant explanations do not influence the response. Both tools will receive identical text in the same order.

Step 1 — Provide the task and observation

Copy only the task and observation from the scenario. Do not include the participant's explanation. Add this prompt:

> Based on this task and observation, what can you conclude about what happened? Explain your reasoning.

Save the complete response before continuing. Record whether the tool independently acknowledges uncertainty, identifies a need for more information, or suggests a follow-up question.

Step 2 — Ask for a follow-up question

In the same conversation, enter:

> "What follow-up question would you ask the participant to better understand what happened?"

Save the complete response. Keep this prompted response separate from any question the tool suggested independently in Step 1.

Step 3 — Provide the participant's explanation

Copy the participant's explanation from the scenario into the same conversation. Add this prompt:

> "Given the participant's explanation, what can you now conclude about the difficulty they experienced? Explain whether this changes or confirms your earlier interpretation. What specific change would you recommend, and how would it address the difficulty described?"

Save the complete response.

Step 4 — Record and evaluate the results

Label the saved responses with the tool name, model or version if displayed, test date, scenario number, and step number. Evaluate the responses using the Evaluation Criteria.

Repeat the procedure for all three scenarios in both tools, producing six test runs in total.

Evaluation Criteria

Each AI response will be evaluated using the criteria below. Each criterion will receive a rating of Met, Partially met, or Not met, supported by a short excerpt from the response. Both tools will be evaluated against the same criteria across all three scenarios.

Stage 1 — Task and observation only

- Separates observation from interpretation. The tool accurately describes what the participant did and presents any possible reasons as hypotheses.
- Recognizes when more information is needed. The tool indicates on its own that the observation leaves the participant's reason uncertain. I will assess this using its first response, before asking for a follow-up question.
- Suggests a useful follow-up question. The question focuses on the moment the participant got stuck, gives them room to explain in their own words, and avoids suggesting an answer. I will record whether the tool offered the question on its own or after being prompted.

Stage 2 — After receiving the participant's explanation

- Updates its interpretation using the new evidence. The tool accurately reflects what the participant said and revises or confirms its initial interpretation accordingly. If the tool did not draw a conclusion in Stage 1, its next response will be checked to see whether it correctly identifies the difficulty explained by the participant.
- Recommends a change that addresses the difficulty. The tool proposes a specific change and explains how it would help with the issue the participant described.

Analysis

The analysis will compare the two tools side by side for each scenario using the ratings and excerpts recorded under Evaluation Criteria. In Stage 1, the comparison will focus on how each tool handled uncertainty and whether it suggested a useful follow-up question on its own or after prompting. In Stage 2, the comparison will focus on how each tool used the participant's explanation to update its interpretation and recommend a change.

The analysis will summarize patterns across the three scenarios for each tool. Counts and short excerpts will show where the responses differed and how the participants' explanations affected the recommendations.

Limitations

This is a small exploratory study with three fictional scenarios and one run per scenario in each tool. The findings apply to the prompts, scenarios, and model versions tested. The same tool could respond differently in another run.

Each scenario provides a clear participant explanation. In actual usability sessions, participants may give partial or conflicting explanations, or several factors may contribute to the same difficulty. The study evaluates whether the AI's recommendations address the difficulty described. Determining whether those changes improve usability would require testing them with users.