Response Analysis: Example

Written by Febogamedev

Consider a hypothetical scenario. A designer asks a language model how to improve students' onboarding on an educational platform.
The model responds:
"Adding an automatic video tutorial to the home page makes the interface more intuitive and reduces student drop-off."

The response is clear and plausible, but contains at least two distinct claims:

- The automatic tutorial makes the interface more intuitive.
- The automatic tutorial reduces drop-off.

The first claim is primarily evaluative, while the second introduces a causal relationship.
The four techniques allow these claims to be analyzed from different, complementary perspectives.

The process can be summarized as follows:
Claim → domain knowledge → inter-response consistency → retrospective verification of criteria → cross-checking → verified rewriting

Domain knowledge
Domain knowledge first helps us recognize suspicious or insufficiently defined elements.
An experienced interface designer might observe that:

- "Intuitive" is a vague evaluation.
- The effect may depend on the type of user.
- An automatic tutorial may help a beginner but annoy an experienced user.
- Autoplay may create problems with accessibility, attention or data usage.
- Drop-off may depend on many variables other than the tutorial.
- A simple correlation does not demonstrate a causal relationship.

Domain knowledge does not demonstrate that the response is false.
However, it allows us to unpack it and identify what needs to be checked.

The general claim is thus turned into more precise questions:

- "What does 'intuitive' mean in this case?"
- "For which students is it supposed to work?"
- "How is drop-off measured?"
- "Is the tutorial mandatory, optional or skippable?"
- "What other variables might influence user behavior?"

Domain knowledge therefore serves an initial diagnostic purpose:
it flags critical points and suggests which criteria, variables and sources will be needed for verification.

Inter-response consistency

The second technique involves checking whether the model maintains its position when the same issue is addressed through different formulations or operations.

Consistency under paraphrasing
The request can be repeated in separate sessions:

- “Does automatically showing a video tutorial on first access improve the platform’s initial usability for students?”
- “Does automatically starting a video tutorial when a student first accesses the platform make it easier to use initially?”.
- “For a student accessing the platform for the first time, does an automatically started video tutorial make it easier to use?”.

The model might respond relatively consistently that the tutorial facilitates initial learning.

Cross-aspect consistency
The same concept is then subjected to different questions:

- “Is the automatic tutorial also useful for experienced users?”
- “What accessibility problems might it introduce?”
- “Could a mandatory video increase drop-off instead of reducing it?”
- “How do you distinguish the tutorial’s effect from that of the interface’s overall quality?”

At this point, the model might acknowledge that:

- Autoplay may feel intrusive.
- Experienced users may perceive it as an obstacle.
- A long video may increase drop-off.
- The effect depends on whether it can be skipped.
- Drop-off cannot be causally attributed to the tutorial without data.

The initial response therefore turns out to be partially consistent.
Its core remains plausible, but the absolute formulation is not maintained when the claim is subjected to more specific questions.

Inter-response consistency does not yet establish which version is true.
However, it reveals that the claim is unstable and depends on conditions initially omitted.

Domain knowledge flags the problem; inter-response consistency shows how much the problem affects the stability of the response.

Retrospective verification of criteria
At this point, the original claim is isolated and the model is asked to make its criteria, conditions, assumptions and limitations explicit.
A request could be:

- “On what criteria do you determine that the automatic tutorial makes the interface more intuitive?”
- “What conditions must be present for it to reduce drop-off?”
- “Which users are you considering, and in which cases might the relationship not hold?”

The model might identify the following metrics:

- Percentage of users who complete the first task.
- Time needed to start using the platform.
- Number of errors made.
- Requests for assistance.
- Drop-off rate during first access.
- The user's ability to find the main functions independently.

It might also make certain assumptions explicit:

- Users are predominantly beginners.
- The video is short.
- Audio does not start automatically.
- The tutorial can be skipped.
- The content is accessible.
- The connection is fast enough.
- The video shows operations that are actually necessary.

Finally, some limitations might emerge:

- Experienced users might find it repetitive.
- A long video might slow down access.
- Some users might prefer text instructions.
- Autoplay might create problems with assistive technologies.
- Drop-off might depend on other aspects of the platform.

The initial formulation can therefore be turned into a more precise hypothesis:
"A short, accessible and easily skippable tutorial can facilitate less experienced students' first use of the platform and might reduce initial drop-off."

This formulation is easier to check, but has not yet been verified.
The criteria and conditions were generated by the same model and might represent a subsequent rationalization.
They must therefore be treated as working hypotheses.

Retrospective verification does not prove the claim: it brings to light or suggests criteria, conditions and evidence to check.

Cross-check
Cross-checking introduces elements independent of the model.
In our example, verification could use four types of evidence:

- Accessibility documentation and guidelines.
- Platform analytics data.
- Usability tests.
- A comparative experiment between different versions of the interface.

<u>Source checking</u>
Guidelines and documentation allow us to check whether automatically starting the video introduces technical or accessibility problems.

However, finding a blog that says autoplay is negative or positive is not enough.
We need to assess:

- Who produced the document.
- Which criteria it uses.
- Whether it is relevant to the case.
- Whether it is up to date.
- Whether other independent sources reach compatible conclusions.

Data analysis

Two versions of the platform can be compared:

- Version A: automatically started video tutorial;
- Version B: optional tutorial that the user can activate.

In a hypothetical scenario, the data might show:

Measure	**Autoplay**	**Optional tutorial**
First-task completion	78%	84%
Average time to get started	95 seconds	80 seconds
Initial drop-off	18%	12%
Voluntary tutorial use	---	43%

The results do not support the initial claims.
The tutorial may be useful, but activating it automatically does not seem to produce the best result.
Qualitative observations from testing might show that:

- Some beginners appreciate the video.
- Experienced users close it immediately.
- Some students prefer text instructions.
- Automatic playback interrupts the flow of those who already know the platform.
- The optional tutorial is used mainly by users who actually need it.

Checking independence

The evidence must not all depend on the same origin.
Usage statistics, observational tests and technical guidelines represent different forms of checking:

- The data show what users do.
- The tests help us understand why they do it.
- The guidelines allow technical and accessibility aspects to be evaluated.

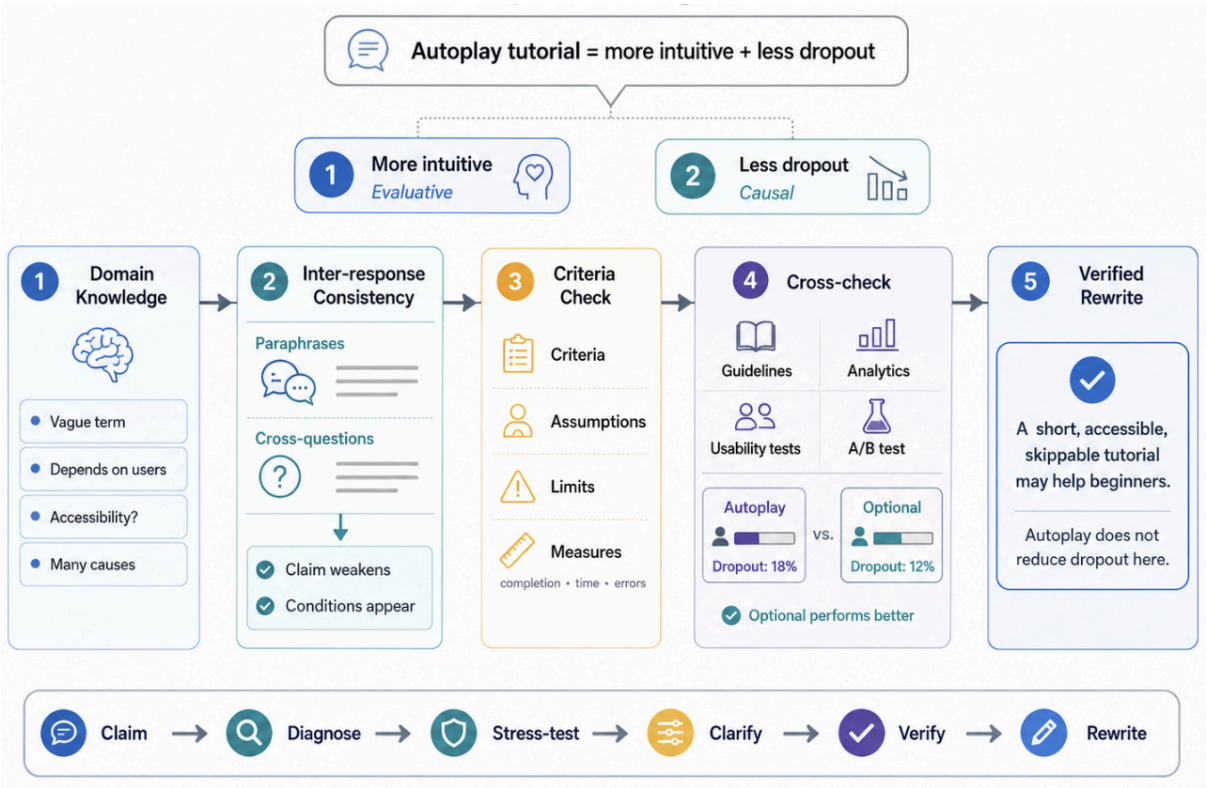
Convergence between different methods is more meaningful than the repetition of the same opinion across multiple pages.

Checking outcome

The two initial claims can now be classified separately.

“The tutorial makes the interface more intuitive”
Outcome: qualified.
The tutorial seems to help some users, particularly beginners, but the effect depends on its format, duration and whether users can choose to use it.

“The automatic tutorial reduces drop-off”
Outcome: contradicted in the scenario examined.
The hypothetical data show lower drop-off in the version with an optional tutorial. There is therefore no adequate support for the initial causal relationship.



The original response was:
“Adding an automatic video tutorial to the home page makes the interface more intuitive and reduces student drop-off.”

After applying the four techniques, it could be rewritten as follows:

“In the tests considered, a short, optional tutorial primarily helped less experienced students understand the platform’s initial functions. Automatic playback, however, did not reduce drop-off and slowed down some users. The tutorial’s effectiveness therefore depends on the audience, duration, accessibility and ability to skip it.”

The new formulation:
- Distinguishes between different users.
- Removes the absolute statement.
- Separates the tutorial’s usefulness from the effect of autoplay.
- Specifies the conditions.

- Attributes the conclusion to the observed data.
- Maintains a degree of certainty proportionate to the evidence.

The techniques do not perform the same task and are not interchangeable.

- Domain knowledge provides guidance: It allows us to recognize that terms such as “intuitive” and “reduces drop-off” require criteria and measurements. It also suggests which variables and problems to consider.
- Inter-response consistency applies pressure: It shows that the initial formulation does not remain stable when different users, conditions and counterexamples are introduced.
- Retrospective verification makes the claim operational: It turns vague concepts into measurable criteria, such as task completion time, errors, completion and drop-off. It also brings assumptions and limitations to light.
- Cross-checking compares with external evidence: It checks whether criteria, assumptions and conclusions are supported by documents, data, tests and independent sources.

The combined process does not necessarily end with cross-checking.
External results may make it necessary to return to earlier steps.
For example, discovering that beginners and experienced users react differently may lead us to:

- Modify the definition of “user”.
- Distinguish new criteria.
- Formulate further consistency questions.
- Run a new test.
- Rewrite the claim again.

The procedure therefore takes an iterative form:
analysis → checking → evidence → revision → further checking

In conclusion, the four techniques build different levels of confidence.

- Domain knowledge and inter-response consistency help identify weaknesses.
- Retrospective verification turns these weaknesses into criteria, conditions and questions that can be checked.
- Finally, cross-checking compares the claim with external evidence and allows it to be rewritten in a form proportionate to the results.

The first three techniques prepare the verification question; cross-checking looks for the answer in the evidence.

None of the first three techniques alone demonstrates that the claim is true.
Indeed, cross-checking could be poorly targeted without the first three techniques: we would not know precisely which part of the response to verify, which sources to look for or which data to collect.

Supplementary material

Exercises

Mobile website navigation
A designer asks a generative system:
“For a website viewed mainly on smartphones, is it better to use a hamburger menu or to display all the main items directly?”

The system responds:
“The hamburger menu is the best solution for mobile devices because it takes up less space, makes the interface cleaner and allows users to find the website’s sections more easily.”

Critically analyze the response by applying the techniques studied in the chapter on response analysis.
Decide independently:

- Which claims deserve to be isolated.
- Which aspects can be questioned through domain knowledge.
- Which consistency tests to perform on the model.
- Which criteria, assumptions, conditions or limitations to make explicit.
- Which external evidence would be appropriate for checking the most important claims.

At the end of the analysis, assign an outcome to the main claims and rewrite the original response in a form you consider adequately supported by the checks performed.

Briefly document the path followed and justify your choices.

Promotional video
A company wants to improve a promotional video published on social media and asks:
“Can adding background music make the video more effective?”

The system responds:
“Adding rhythmic music increases viewer engagement and consequently raises average viewing time. For a promotional video, it is therefore always preferable to use an energetic soundtrack.”

Subject the response to a full analysis.

You do not necessarily have to verify every sentence: determine which parts are truly relevant to judging the overall reliability of the response.
Use any of the following:

- Domain knowledge.
- Inter-response consistency.
- Retrospective verification of criteria.
- Cross-checking.

If you consider a technique unhelpful for a particular claim, you may choose not to use it, provided you justify your choice.

Finally, you must produce:

- The claims you decided to check.
- A summary of the anomalies or uncertainties that emerged.
- The evidence you considered relevant.
- Your final judgment on the reliability of the response.
- A new version of the response, with a degree of certainty proportionate to what you actually managed to verify.

Posting on social networks

A communications manager asks:
“What is the best time to publish a post on social media?”

The system responds:
“The best time is between 6:00 p.m. and 8:00 p.m., because during that period most people have finished work or school and are more active on social media. Posting during this interval typically increases engagement by around 30%.”

Assess the reliability of the response as if you were going to use it to make an actual communication decision.
You are free to determine:

- Which claims to start with.
- Which elements require more rigorous verification.
- Which variations of the question to use to test the stability of the response.
- Which assumptions might be hidden in the conclusion.
- What “best time” means in concrete terms.
- Which sources, data or tests might provide useful evidence.

You do not need to reach a clear-cut conclusion. If the evidence is insufficient, conflicting or context-dependent, this can be a valid result of the analysis.

Conclude by classifying the main claims as supported, requiring qualification, contradicted, unresolved or not verifiable in their current form, then formulate an alternative response.

Quality of an audiovisual product

A student asks:
“To make a professional short film, is it better to record in 4K rather than Full HD?”

The system responds:
“Yes. 4K produces images with four times the detail of Full HD and guarantees a more professional result. It also allows greater freedom in editing, so recording in Full HD is now generally discouraged for a quality production.”

Imagine having to decide whether to include this response in a handout for audiovisual production students.
Analyze it independently using the method developed in the previous handouts.
Do not limit yourself to looking for factual errors: also assess:

- The precision of the concepts used.
- The relationships constructed between data and conclusions.
- The conditions that might change the validity of the statements.
- Any generalizations.
- The type of evidence needed to support a design recommendation.

If new problems emerge during the analysis, you may formulate further questions, perform new comparisons or modify the verification path.
Finally, submit a short report showing how the initial response changed during the analysis, and produce the version you consider suitable for the handout.

Editorial notes

In producing these materials, I have used generative artificial intelligence tools to support the writing process, particularly to improve the form of the text, reorganize content, refine wording, and speed up certain editorial tasks.

As I work independently on the production of these materials, I try to automate any activities that can reasonably be automated, so that I can devote more time to research, design, and content development.

However, artificial intelligence does not determine the content of this work: the choice of topics, structure, ideas, interpretations, examples, and teaching approach are developed by me. AI is therefore used primarily as a tool to support production and formal revision, while authorship and responsibility for the design and content remain mine.

Unless otherwise indicated, this material is distributed under the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International license (CC BY-NC-SA 4.0).

You are therefore free to share, redistribute, adapt, and create derivative works based on this material, provided that appropriate credit is given, the material is not used for commercial purposes, and any modified or derivative versions are distributed under the same license.

