The Signal Brief · Decision quality · July 23, 2026

Subject: The AI said "hold." What did it assume?

Preheader: Three checks before a promising number turns into a bigger commitment.

Three Checks Before You Trust an AI Recommendation

A campaign's working estimate beat its target. Six conversations with Claude recommended holding or deferring. The useful question was not just what it recommended, but what each answer had quietly assumed.

Don Long · Founder, SummitMark · Helping people and organizations become capable with AI

An open notebook, a few coins, a pencil, and a blank calendar suggest a decision that is not ready to scale.

A promising result still needs a deadline, a baseline, and a reason to wait.

> *A number can beat its target without earning the right to be scaled.*

The working number was $3.74: $11.21 spent to promote a newsletter, divided by three subscribers thought likely to have come from the campaign. The target was $5. The campaign was still running, and the attribution was uncertain.

Don Long asked Claude whether to raise the next budget from $20 to $30. In three fresh conversations, the answer was hold. Then he asked each conversation what it had assumed. The recommendation stayed the same. Its confidence softened.

That is a familiar moment in business. A pilot clears its target, a software trial saves a few hours, or a new process survives its first week. A small result starts making the case for a much larger commitment. Should you scale it?

Before you trust the recommendation, check three things that are easy to leave out: your deadline, your baseline, and the cost of waiting.

What the first three answers knew

Long compared six fresh Claude conversations. Three received a concise account: the $3.74 working estimate, uncertain attribution, and the $5 target. They did not receive the fact that the campaign was unfinished or the other campaign details. Each recommended holding the budget, with expressed confidence between 65 and 70 percent.

Those answers focused on the small sample, the $8.79 still unspent against the original $20, and uncertain attribution. But the short prompt already called attribution uncertain and required a choice among raise, hold, reduce, or defer plus a confidence score. Some caution was part of the question itself. These answers were not an unprompted read of a bare number.

Long then asked each conversation to separate the facts he had supplied from assumptions it had filled in, and to name what evidence would reduce uncertainty. All three still recommended holding. Two confidence scores fell to about 60 percent. The third stayed around 60-65 percent, while describing the choice between holding and raising as nearly a coin flip.

The audit surfaced questions that could matter: Why had the spending stopped? How were subscribers attributed? Would the result repeat? That is the model's later explanation of its earlier answer, though - not a verified record of what actually drove that answer.

The "complete" record still had a blank

Three other fresh conversations received more context: the unfinished campaign, LinkedIn's engagement objective, no direct subscription attribution, one personal-network subscriber among the four new subscribers, and unknown repeatability.

Two recommended holding and one recommended deferring. All gave 70 percent confidence. Their split turned on a fact neither version of the prompt included: whether the campaign would finish before Long had to choose the next week's budget. The defer answer assumed it would finish in time; a hold answer assumed the budget had to be set regardless.

That does not establish that the missing deadline caused the different recommendations. Attribution and underspending were prominent concerns elsewhere. But it shows how a detailed account can still omit the fact that changes what the decision means. The deadline was in a calendar, not in the campaign data.

Three checks before you scale

The practical test travels beyond newsletter promotion - to a sales pilot, AI rollout, vendor trial, or hiring plan.

1. State the deadline. When must this decision actually be made? If there is time to collect more evidence, "defer" may be a real option. If the budget or commitment must be set today, the choice is different.

2. Name the baseline. What would have happened without the pilot? Three subscribers may look promising against a $5 target, but the comparison is incomplete if you do not know how many would have subscribed anyway.

3. Price the wait. What do you lose by holding for one more cycle? Waiting may cost momentum or a timely opportunity. It may also avoid scaling a result that is still noisy. Put that trade-off into the question.

Skip any one and the recommendation may answer a slightly different question from the one you meant to ask, even when the rest of the record looks complete.

What six conversations cannot settle

Before this comparison, Long tried looser wording; two of three responses said to raise the budget. Several parts of the prompt changed at once. With three runs per version, the difference could reflect ordinary variation, prompt framing, or both. It is a correlation worth noticing, not evidence that wording caused the shift.

A skeptic could also argue that mentioning an underspent budget and forcing a choice among four categories naturally invites caution. The follow-up audit could generate plausible caveats without diagnosing the first answer. Six runs cannot rule that out. The confidence percentages were Claude's descriptions, not measured probabilities.

So this is a narrow, one-model, one-decision experiment. It does not show that AI always chases attractive numbers - or that the hold recommendation was wrong. It shows something more modest: a recommendation can remain sensible while its explanation introduces assumptions the user never supplied. Asking what is missing gives you another answer to inspect, not a window into hidden reasoning.

Before the next green metric becomes a bigger commitment, ask the audit question, then replace the most consequential assumption with a fact.

What is the $3.74 in your current plan - and which missing fact could turn "scale" into "wait"? Reply with the number and the fact you would check.

M.R.

Source and scope: Adapted from Don Long's supplied article, "3 Checks Before You Trust Any AI Recommendation," dated July 23, 2026. The campaign details, six conversations, recommendations, and confidence scores are reported by Long in the supplied text; the underlying prompts and model outputs are not included for independent review. This issue does not treat expressed confidence as calibrated probability or infer causation from the small comparison.