

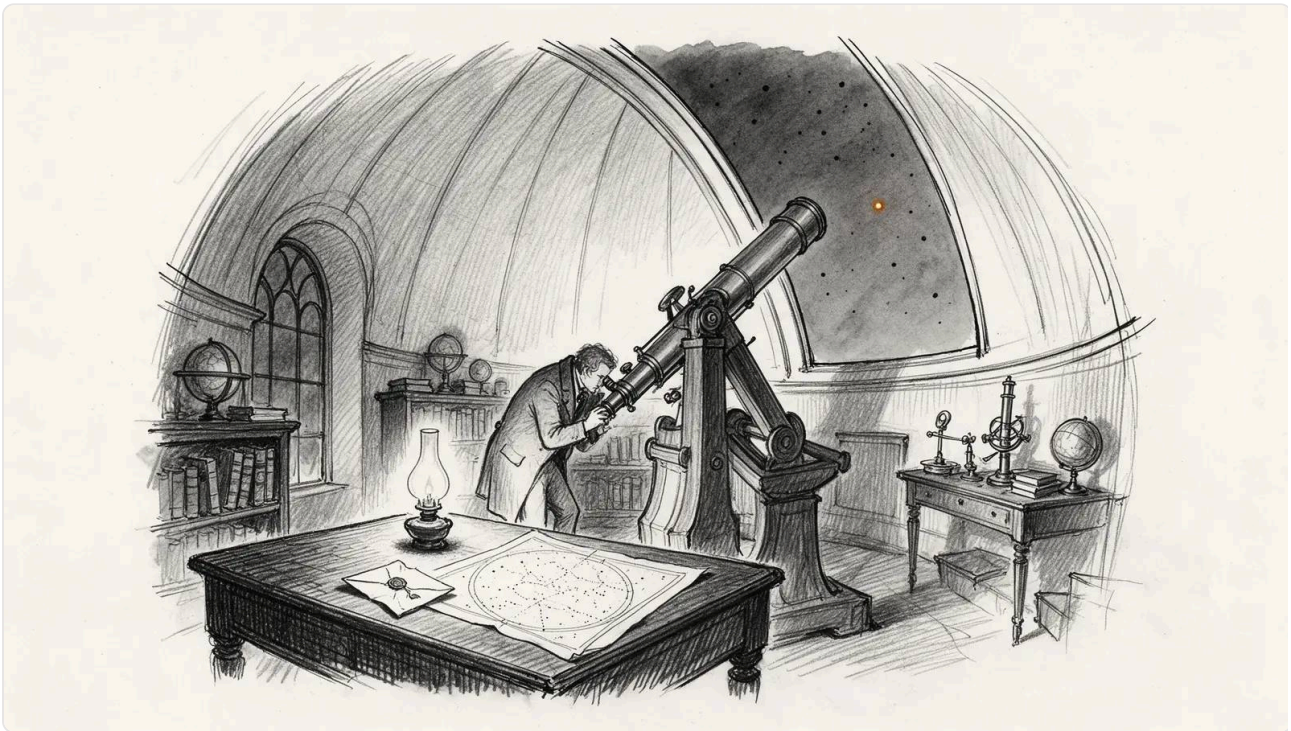
Two Planets, One Hallucination

In 1846 a mathematician predicted a planet nobody had seen, and it was there. Thirteen years later he did it again, and it was not. What that says about AI hallucinations, which ones we remove from a deck review, and why a wrong investor question still needs an answer.

pitchera.ai · 6 October 2026 · 12 min read · pitchera.ai/blog/two-planets-one-hallucination

On 23 September 1846 a letter reached Johann Galle at the Berlin Observatory. It came from Urbain Le Verrier, a French mathematician who had spent over a year on the orbit of Uranus. Uranus kept drifting off its predicted path, and Le Verrier had concluded that an unseen planet was pulling on it. The letter said where to point the telescope.

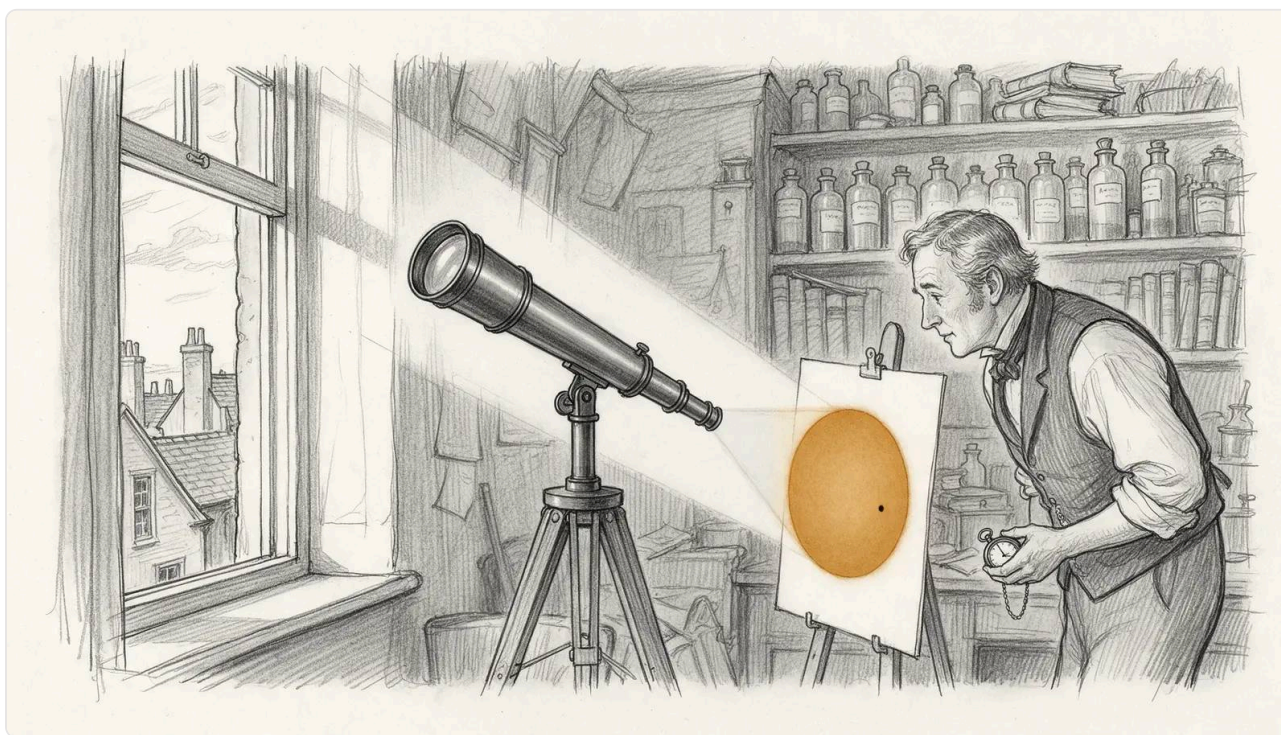
Galle looked that night. He found Neptune in under an hour, within one degree of the place in the letter. A colleague in Paris said Le Verrier had discovered a planet with the point of his pen.



Berlin, 23 September 1846. The letter said where to look, and something was there.

Thirteen years later he did it again. Mercury was misbehaving too: its orbit turned a little faster than the known planets could explain, by 38 seconds of arc per century, about a hundredth of a degree. Same man, same method, same conclusion. Something unseen had to be there, closer to the Sun than Mercury.

Within months a country doctor with a telescope, Edmond Lescarbault, wrote to say he had watched a small dark dot cross the Sun earlier that year. Le Verrier went to question him in person and came back convinced. The doctor got the Légion d'honneur. The planet got a name, Vulcan.



One dot, seen once, by one man. It was enough for a medal and a name.

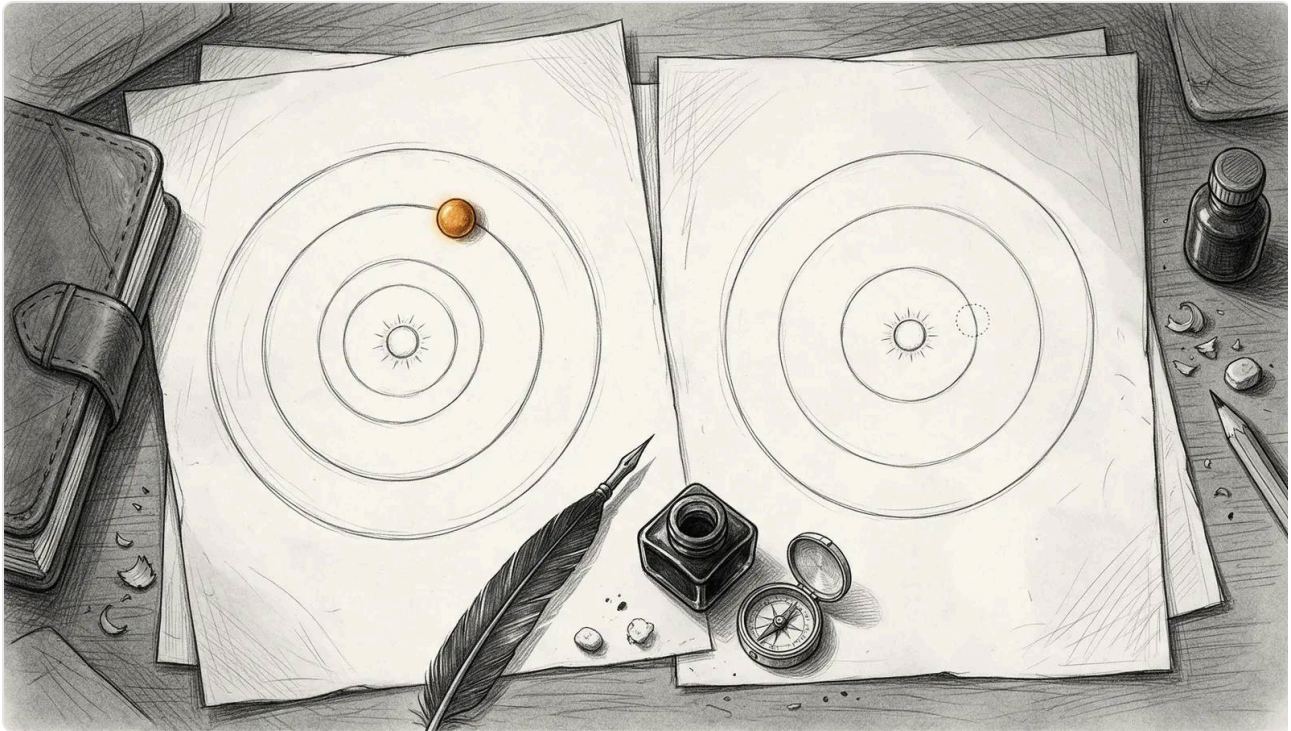
Nobody ever found it. Astronomers searched for decades, and a few experienced observers reported seeing it. Le Verrier died in 1877, on the anniversary of the night Galle found Neptune, still expecting Vulcan to turn up. In 1915 Einstein ran Mercury's orbit through his new theory of gravity, and the drift came out right with no extra planet.

On the day each letter went out, the two predictions were the same kind of thing: careful reasoning from real measurements to an object nobody had seen. One is now called a discovery. If a language model had produced the other, we would call it a hallucination. The reasoning did not get worse in between. Somebody looked, and the second time nothing was there.

What is a hallucination, then?

"The model made something up" is not much of a definition. Everything a language model writes is a prediction of what should come next. Most of the predictions hold, so nobody gives them a name.

A more useful one: a hallucination is an assumption that was checked and turned out to be false. Before the check it is only a claim. Neptune and Vulcan were both claims until someone pointed a telescope.



Same pen, same method. Only one of the two circles has a planet in it.

That definition has an awkward edge. To call something a hallucination you need something to check it against, and some sentences come with nothing to check.

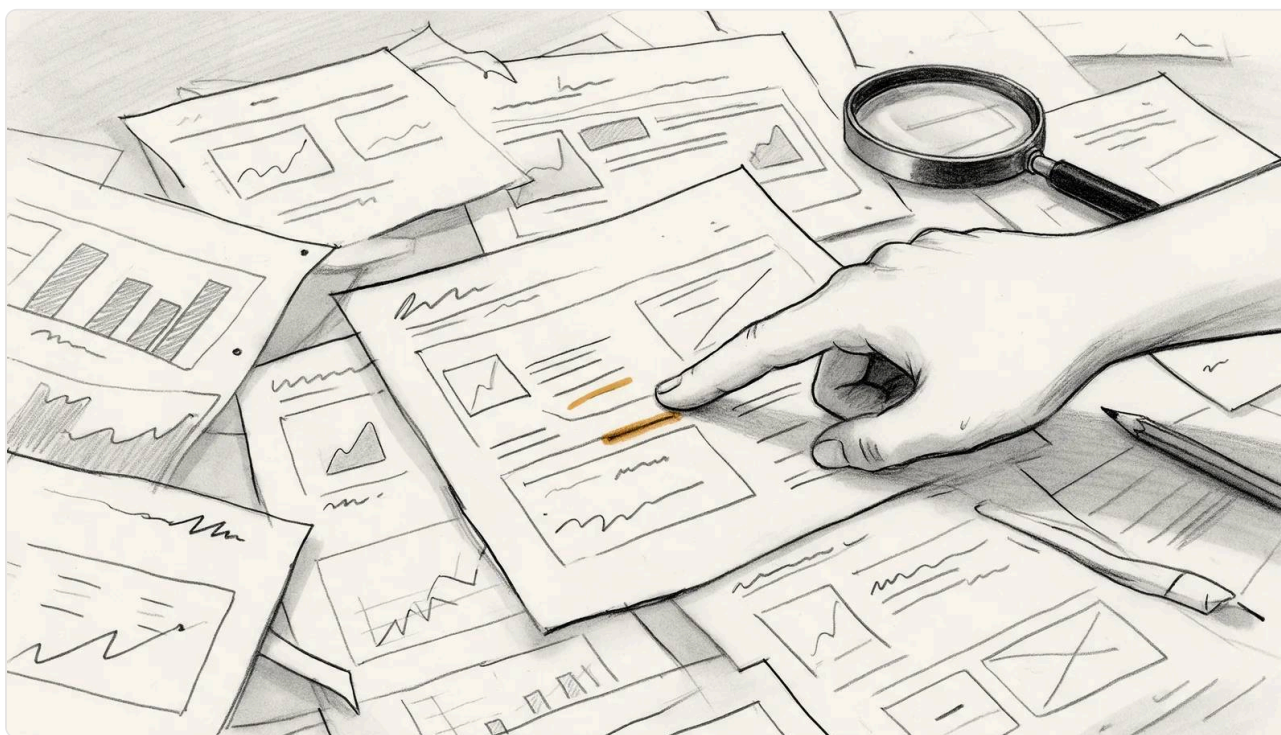
We meet that edge every day. pitchera reads a pitch deck the way an investor would and writes back what is there, what is missing and what you will be asked. It runs on a language model, so hallucinations are not an academic subject for us. Whether one is a problem depends on which of three kinds of sentence it turns up in.

1. Sentences about your deck

"The deck gives no price." "The team slide lists three founders." "Revenue is forecast at two million in year three."

These can be checked, and the check is cheap, because the deck is right there. If a review says your deck has no price and slide nine says 99 euros a month, that is a hallucination in the plain sense. It has no upside. One of those and you stop believing the other forty lines, and you would be right to.

So in a full report a second pass goes back over the analysis with the original slide text beside it. It looks for numbers that are not in the deck, people who are not on the team slide and quotes nobody wrote, and it rewrites what fails. This is Galle's job: point the telescope where the sentence says the thing is.



THE TEST

Can you put a finger on the slide?

2. Sentences about the world

"Hospital procurement takes nine to eighteen months." "Two established suppliers already sell this to hotel chains."

Neither is in your deck. The model brought them in from what it has read about the world. They can be checked, but not by anyone holding only your slides. Somebody has to look it up or pick up the phone.

Until then each one is a letter to Berlin. It may be Neptune. It may be Vulcan. From the inside they read exactly the same. Le Verrier was the best in the world at this, which is why so few people doubted the second planet.

In a full report these sentences carry a label. Every claim is tagged with where it came from: your deck, your own answers, public knowledge, or the model's own interpretation. A line tagged as interpretation is a lead. Check it before you repeat it to anyone, and above all before you repeat it to an investor.

Your own deck is full of them too. "The market is worth [4.2 billion](#)." "Three per cent of trials convert." "We will be in three countries by 2028." A pitch deck is mostly sentences of this second kind, written by someone who believes them. The investor meeting is where the telescope comes out.



THE TEST

Who could you phone to find out, and have you?

3. Questions

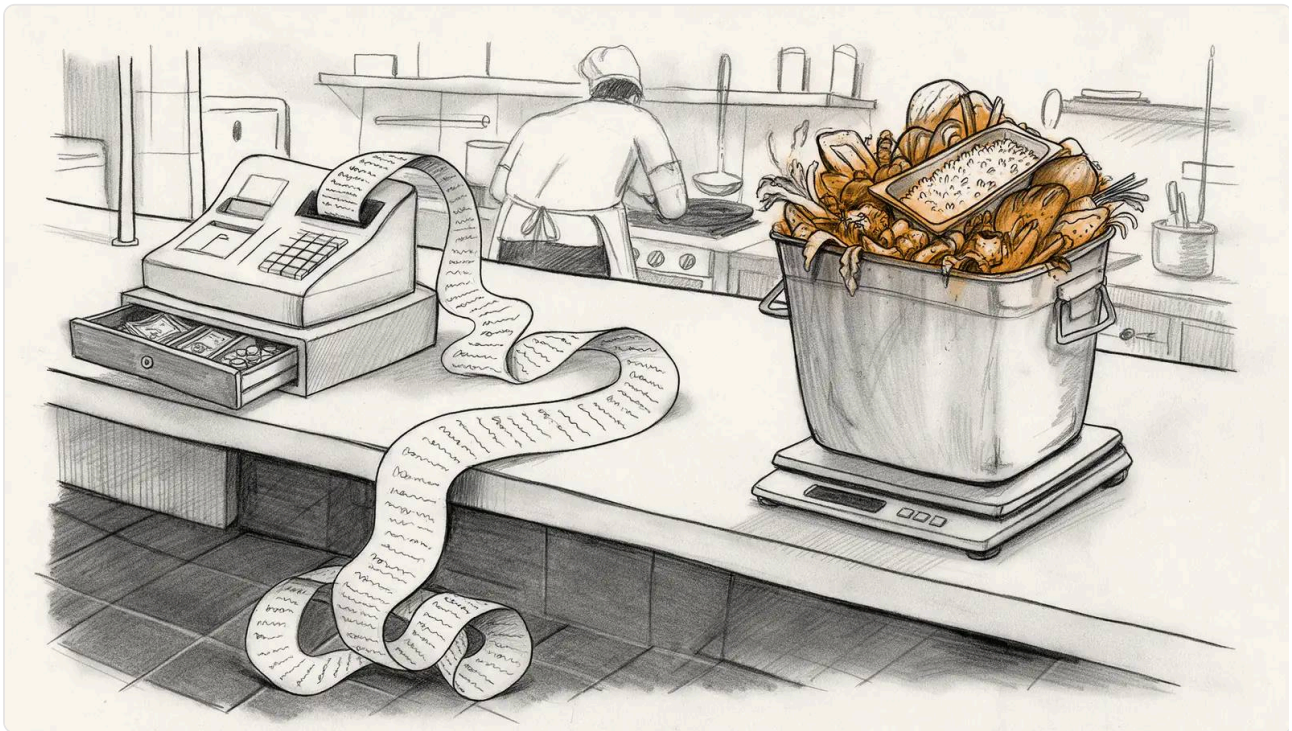
This is where the definition runs out.

Take Fenwick, the made-up team from [the last post](#). They build a smart scale for commercial kitchen bins: it weighs what gets thrown away and shows which dishes waste the most. One founder is an engineer, the other a former chef.

Suppose the review of their deck comes back with this question:

Restaurants already know what they sell from the cash register. Why would they need a scale as well?

The former chef reads it and laughs. The cash register knows what was sold. It has no idea what went in the bin: the trimmings, the overproduction, the tray of rice nobody ordered. Anyone who has worked a service knows that. The question rests on a wrong picture of how a kitchen works.

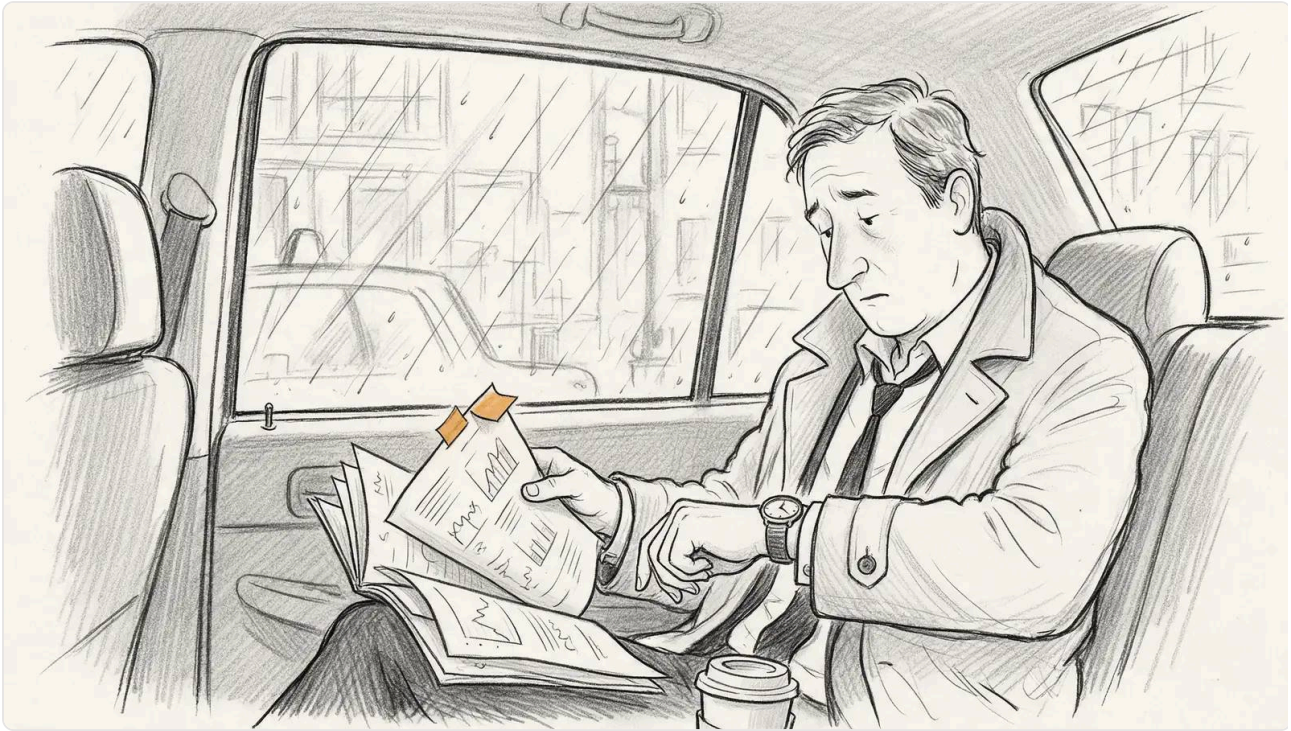


The cash register has a long memory for what was sold. The bin is where the rest went.

So did the model hallucinate?

Try to check it. A question is not true or false. The only claim in sight is "an investor might ask this", and you will never test that. You cannot interview every investor Fenwick will ever meet.

If you could, here is who you would find. Most of them have never worked in a kitchen. A generalist fund looks at payroll software on Monday, a medical device on Tuesday and a kitchen scale on Wednesday. One partner read the deck in a taxi. Another knows hospitality well and is having a bad day. A third knows the answer perfectly and wants to see what you do with a naive question.



Your deck, as it is often read.

The question nobody in your industry would ask is the one somebody outside it will. Most of the money sits outside it.

So this is the one kind of sentence we deliberately do not clean up. The questions in a review are written by a model playing an investor who has only your deck to go on. When that reader fills a gap with a wrong idea, the wrong idea is the finding, because the deck left room for it. Correcting the reader would hide the exact place where your deck let them guess.

THE TEST

Can you answer it in one sentence, without sighing?

What to do with a question that looks wrong

Write the one-sentence answer. "The cash register knows what was sold. We weigh what was thrown away, and that is where the margin goes." If your answer needs a paragraph, the question was less silly than it looked.

Find the gap it came through. Something in the deck let a careful stranger think cash register data was enough. Usually one line closes it, on the slide where the doubt starts.

Rehearse your face. Founders rarely lose investors on the content of a naive question. They lose them in the half second of "how can you not know this". The investor is watching how you will treat a customer who asks the same thing, and plenty of

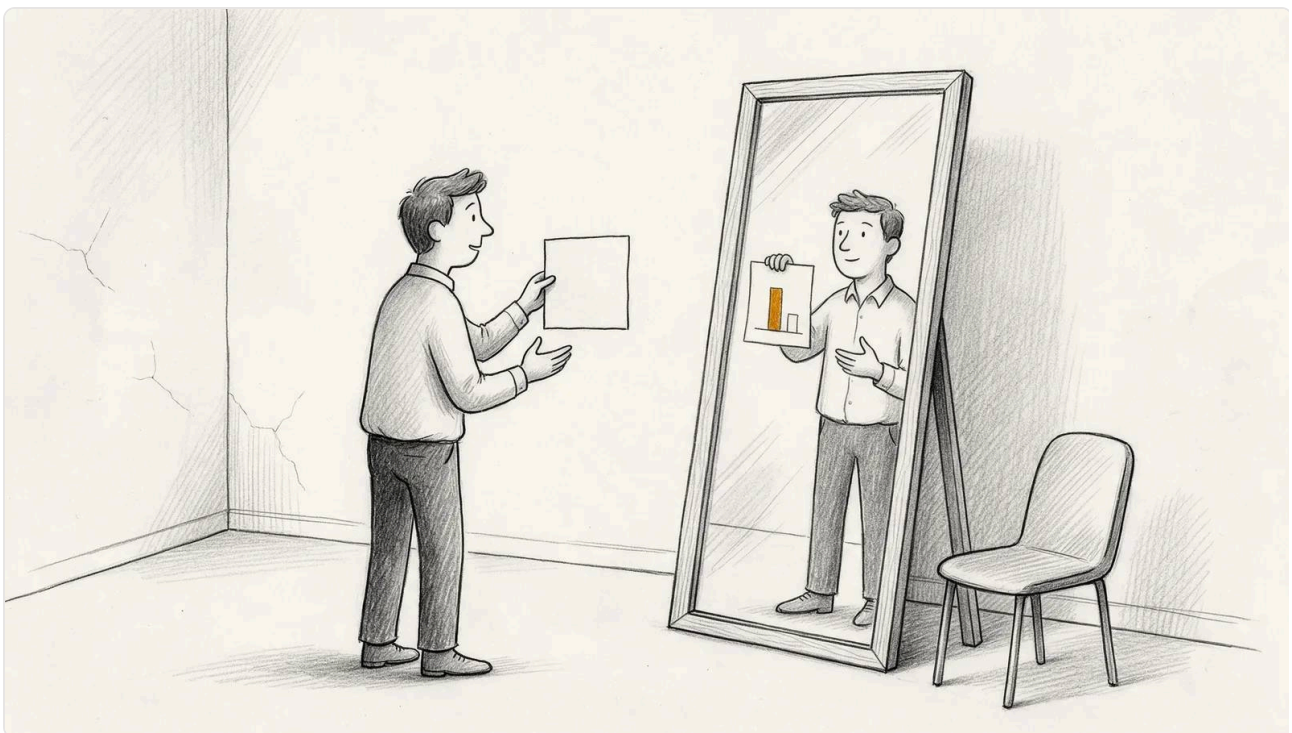
customers will.

Then check whether it is wrong at all. Now and then the outsider is right and the expert was the one assuming. "Why has the cash register vendor not built this already?" sounds naive until you notice you never looked. Vulcan was not an amateur's mistake.

We tried zero

While building and calibrating pitchera we did the obvious thing first and set the tolerance for hallucination to zero. Nothing went into a review unless it could be traced to a slide. Deterministic engine, full steam ahead.

What came back was a faithful account of the deck and nothing else. It told you, accurately and at some length, what your own slides say. No question you had not already thought of, no doubt from anyone outside your industry. A mirror is never wrong either, and it has never asked anyone a question.



The zero-tolerance version. Agreeable, accurate, and no help at all.

So zero was the wrong target, because the value was never in repeating the deck. What we ended up with is two machines that work under different rules.

The first is deterministic, and it owns the arithmetic and the rules. The overall score is worked out in code from a fixed set of weights, stage adjustments and caps, so the model does not get to pick it. A deck with no team slide cannot score above a set ceiling, however well the rest reads. A financial figure that cannot be found in the text of the

slides is dropped. Then the loop from section one runs: write, compare with the slides, rewrite what fails. If a serious error cannot be repaired, an invented metric or a person given the wrong role, its sentence is cut, and the report is not allowed to call the deck VC-ready, whatever the score says.

The second machine is allowed to imagine, inside a fence. It writes the investor questions, the objections and the blind spots with only your deck to go on, and whatever it brings in from outside the deck is labelled as interpretation. When a section leans too far that way, part of it is held back. A competitor map the deck barely supports is not shown at all.

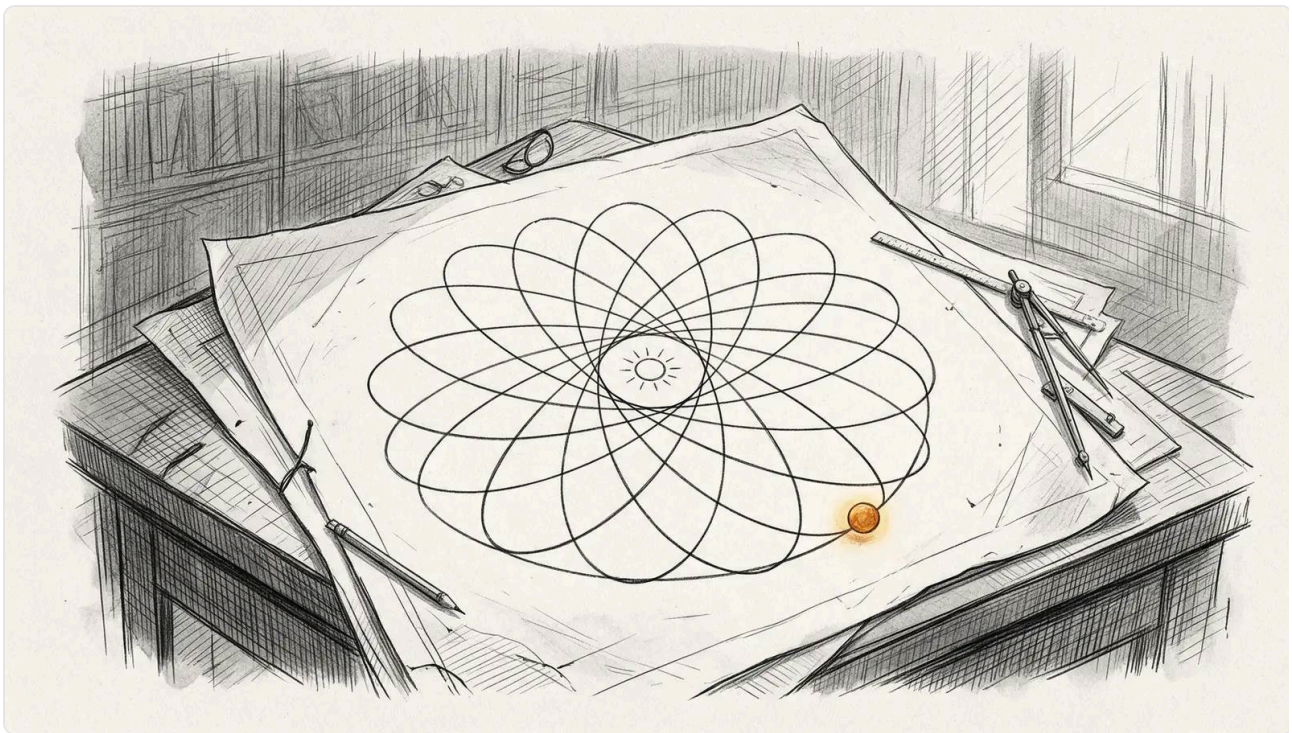
That is the same split as the three kinds of sentence above, learned the hard way. Where there is a slide to check against, the tolerance is zero. Where there is not, the model gets a licence to guess, and the guess wears a label.

So are hallucinations bad?

It depends on what there is to check them against.

- **About your deck:** bad, every time. We check them against the slides and rewrite the ones that fail.
- **About the world:** unproven. We label them, and you check before you repeat them.
- **Questions:** the word does not apply. The only thing to ask is whether you have an answer.

Vulcan sits in the second group, and it earned its place in the textbooks anyway. The planet was never there, but the gap it was invented to fill was real. Mercury's orbit does drift, by about 43 seconds of arc per century on modern figures, and for half a century that number sat in the tables waiting for an explanation. When Einstein's theory produced it without any help, it was one of the first signs that he was right.



Mercury's orbit turns a little with every lap. No missing planet required.

The odd question in a deck review works the same way. The reader's guess may be wrong. The gap it came through is still in your deck, and it is easier to close on a slide this week than across a table next month.

Where this comes from

Galle found Neptune on the night of 23 September 1846, within one degree of Le Verrier's position; the "point of his pen" line is attributed to François Arago. Le Verrier published the Mercury discrepancy in 1859, Lescarbault's letter and the visit followed in December that year, and the name Vulcan was in use by early 1860. Le Verrier's figure was 38 seconds of arc per century; the modern one is about 43.

The definition of a hallucination used here is our working one. In AI research the word is used more broadly, for output that its source does not support. Fenwick is invented, and the cash register question was written for this post, not taken from anyone's report. The checking pass is itself a model, so it catches most errors and not all of them, which is the reason the labels exist. How the labels work is on the [methodology page](#).

Try it on your own deck

The [free deck review](#) reads your deck the way an investor would and returns the three questions you are most likely to be asked. One of them may look wrong to you. Answer that one first. No payment, no call.



Which three questions would your deck draw? The free deck review runs the same analysis on your own deck and returns the three an investor is most likely to ask. No payment, no sales call.

Scan the code, or go to pitchera.ai/deck-review