

Hallucination Categories

A reference card for the verification work

AI hallucinates in patterns. Once you know the patterns, you spot them faster.

This card names five categories that show up most often in medical writing. Keep it open while you verify. When something feels off in your draft, scan the list. Naming the type of error makes the fix obvious.

The five categories

1. Fabricated references

The AI invents a citation that does not exist. Author names, journal, year, even DOI can look completely plausible.

What it looks like: A reference list with a Smith et al. 2022 Lancet paper that was never published. Or a real author paired with a study they never wrote.

What to check: Every citation. Search the title in PubMed or Google Scholar. If you cannot find it, the AI made it up.

2. Misattributed claims

The source exists, but it does not say what the AI says it says. The AI borrowed the author's credibility to support a claim the author never made.

What it looks like: An AI draft cites a real ASCO abstract as the source for a survival figure, but the abstract reports a different endpoint or no survival data at all.

What to check: Open the actual source. Find the exact sentence that supports the claim. If you cannot find it, the AI is wrong about what the source says.

3. Statistical drift

The number is close to the real number but slightly off. 23% becomes 27%. A hazard ratio of 0.44 becomes 0.40. Median PFS of 16.0 becomes 16.4.

What it looks like: An AI draft says camizestrant cut the risk of progression by roughly 60%. The real figure from SERENA-6 is closer to 56%. Close enough to feel right, wrong enough to be wrong.

What to check: Every number against the source. Do not eyeball it. Read the digit.

4. Conflated findings

The AI merges results from two different studies, or from two different arms of the same study, into one statement that does not match either.

What it looks like: A draft says a trial showed improved overall survival and progression-free survival. The trial reported PFS only. The OS data came from a different study, or were still immature.

What to check: For any combined claim, separate it. Verify each half against its actual source. Watch especially for OS, PFS, ORR being treated as interchangeable.

5. Phantom subgroups or analyses

The AI invents a subgroup analysis, a secondary endpoint, or a post-hoc finding that was never reported.

What it looks like: A draft mentions benefit in postmenopausal patients with bone-only disease. The trial reported the overall result and one or two prespecified subgroups, but not that one.

What to check: Any subgroup, secondary endpoint, or exploratory finding. Confirm it appears in the source by name, not by inference.

Not on this list, but also real

These five are not all the ways AI gets things wrong. Three others worth knowing:

- Outdated information presented as current (a treatment that has been superseded, a guideline that has been updated)
- Drug name and dosing errors (wrong indication, wrong dose, wrong route)
- False precision (a confidence interval, a p-value, or a sample size invented to make a claim feel solid)

If anything in your draft trips one of these, treat it the same way: open the source, find the support, or fix the claim.

The fluent confidence problem

AI does not sound less sure when it is making something up. The voice is identical. A fabricated reference looks the same as a real one. A wrong number is delivered with the same flow as a right one.

This is the whole reason verification is non-optional. You cannot detect the categories above by tone. You can only detect them by opening the source.

Always verify

Six things, every time

1. Numbers (any number, every number, including dates and percentages)
2. References (author, title, journal, year, that the paper exists, that it says what you think it says)
3. Drug names (spelling, brand vs. generic, indication)
4. Dosing (amount, route, schedule)
5. Statistical claims (effect sizes, hazard ratios, confidence intervals, p-values)
6. Citations as a whole (does each claim trace back to a real, supporting source?)