

The Evidence-Gated AI Checklist

Eight tests for AI used in expert analysis. Tick one box for each test.

Test	Tick one
1 Baselines Has the AI been compared with a qualified expert and with a simpler method, on the same tasks? What a Yes looks like: Results on cases with known answers, side by side with an expert and a simpler approach.	☐ No ☐ Partly ☐ Yes
2 Verification separate from generation Is the work checked by something other than the system that produced it? What a Yes looks like: An independent check, by a different method or a person, that confirms required content is present and correct.	☐ No ☐ Partly ☐ Yes
3 Executed, not generated, numbers Does every number that matters come from a calculation that was actually run? What a Yes looks like: Each figure links to the calculation that produced it, and rerunning it gives the same value.	☐ No ☐ Partly ☐ Yes
4 Specification uncertainty Has anyone tested whether other reasonable choices of data, sample, definitions or model change the answer? What a Yes looks like: Results shown across the plausible alternatives, with the choices that drive the answer named.	☐ No ☐ Partly ☐ Yes
5 Causal-claim checks Where the output says one thing caused another, does it state the comparison and the assumptions behind the claim? What a Yes looks like: Each causal claim names what was compared with what, and the assumption that makes the comparison fair.	☐ No ☐ Partly ☐ Yes
6 Provenance Can every claim and figure be traced to a real, dated source that someone opened? What a Yes looks like: Every source link works, was checked by a person or an independent process, and carries the date it was accessed.	☐ No ☐ Partly ☐ Yes
7 Human gates for assumptions and values Do named people decide the assumptions, value judgments and final recommendation? What a Yes looks like: Defined sign-off points where a named person approves assumptions and the recommendation, with a record of what they approved.	☐ No ☐ Partly ☐ Yes
8 Reproducibility Could someone else rerun the analysis and get the same answer? What a Yes looks like: Inputs, versions, settings and steps are kept, and a rerun reproduces the result.	☐ No ☐ Partly ☐ Yes

How to score

Yes 2, Partly 1, No 0. A No on executed numbers, causal claims or human gates means the output should not drive a decision yet, whatever the total.

Your score: _____ of 16

How to use the score

- 14 to 16, with no red flags.** Evidence-gated. The output can support decisions with routine expert review.
- 9 to 13.** Usable with caution. Treat the output as a strong draft and close the gaps before a decision relies on it.
- 8 or below, or any red flag.** Not yet defensible. Use the output as a starting point for expert work, not as the evidence.