

AI Visibility Test Log

One AI answer is one roll of the dice. Run the same prompts, several times, on several platforms, and write the conditions down. Your first full run is your baseline: the "before" every later change is measured against.

From lesson 2.3 · The full course and the GEO audit: reyse.co.uk/geocourse · October 2026 edition · General education, not legal advice

LESSON 2.3 · REYSE.CO.UK/GEOCOURSE/TESTING-AND-BASELINE

Testing Fairly, and Running Your Baseline

DO THIS FIRST Run your standard prompts on two platforms, three times each, in private windows, and log every answer before you change anything.

CHECK

Run your baseline · Do this before you change anything, or you'll never know what the "before" was.

- ☐ Take your standard set from lesson 2.2 (12 to 20 prompts).
- ☐ Pick two or three platforms to start: ChatGPT with search on, Google AI Mode, and Copilot or Perplexity.
- ☐ Run each prompt three to five times per platform, each in a fresh private window, spread over two to four weeks.
- ☐ Log every answer with the columns below.
- ☐ Write the headline as "named in X of N runs", never a position.

Test from three places · "Near me" answers depend on where the AI thinks you are.

- ☐ Run five local prompts from the office, from a phone on mobile data, and from a colleague's home on the other side of your patch.
- ☐ Compare "estate agents in [town]" with "estate agents near me" on the same assistant.
- ☐ Log the location every time.

Logged in versus logged out

- ☐ Run one prompt five times in a private window, then five times logged in.
- ☐ If the tool lets you switch memory or custom instructions off, note the setting.
- ☐ A big difference means your own history is colouring what you see.

Recommendation and comparison prompts · The most valuable prompts, and the least stable.

- ☐ "Who is the best estate agent in [town]?" "Which letting agent should I use in [town]?"
- ☐ "[Your agency] vs [your main rival]" and "What do customers say about [your agency]?"
- ☐ Note the sources each answer cites. They're your map for module 7.

FIX

Use these log columns

- ☐ Date, platform, logged in or out, search on or off, location, run number.
- ☐ Prompt ID, audience, exact prompt.
- ☐ Named (yes/no), cited with a link (yes/no, and the address), other agencies named, sources shown.
- ☐ Accurate (yes, no, partly), errors about you, screenshot file name, notes.

Size it honestly · Small numbers jump around. These are rules of thumb, not standards.

- ☐ Light baseline: 12 prompts, 2 platforms, 3 runs each = 72 answers. Treat moves under about 10 percentage points as noise.
- ☐ Standard baseline: 20 prompts, 2 to 3 platforms, 3 to 5 runs each = roughly 120 to 300 answers.
- ☐ Zero mentions in 20 runs doesn't prove you're never named. Keep going before you conclude anything.

Freeze the set

- ☐ Don't change the wording of your core prompts for at least six months.
- ☐ Add new prompts as a separate list.
- ☐ Never reword a prompt to get a better answer.

Let AI tidy the log, not judge it

AI PROMPT Here's my AI prompt test log (no client data). Count how often my agency was named and cited, per platform and per audience, and list every website cited, with how many runs it appeared in. Don't guess any value that isn't in the log.

Then: Repeat the same set every month, with three runs per prompt. Compare totals, not single answers, and look at the pattern over three months before you claim anything moved.

Why "we're number 2 in ChatGPT" is meaningless

Order changes from run to run, so a position is a snapshot of one roll. "Named in 11 of 48 answers" survives a re-test; "second" doesn't.

Two months of 30 answers each can easily differ by 10 points through chance alone. Only compare months with similar numbers of answers, and only believe differences that repeat.

The log

One row per answer. Platform: ChatGPT, AI Mode, Copilot, Perplexity, Claude, Gemini. Note logged in or out, and where you were.

Date	Platform, in/out, place	Prompt ID	Run	Named?	Cited (link)?	Others named	Accurate?

Monthly totals

Compare totals, never single answers, and look for a pattern across three months before you claim anything moved.

Month	Answers logged (N)	Named in (X)	Cited in	Errors found	Notes

Rather have us do this for you? Our managed GEO service takes it off your hands: reyse.co.uk/products/geo