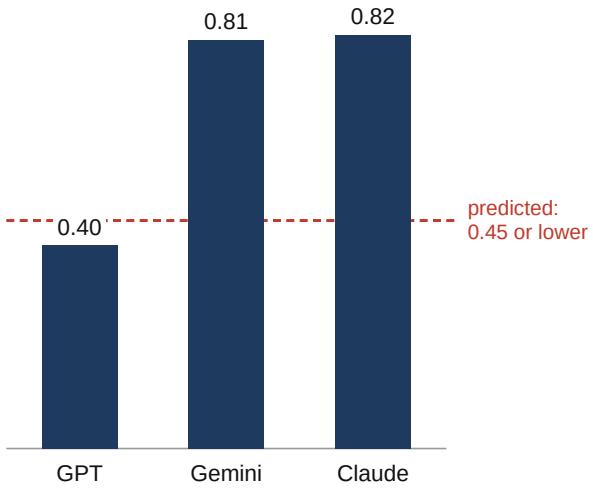
Belief Stability in AI Assistants

GPT-5.6, Gemini 3.8 Flash and Claude Sonnet 5 recommended software for 20 business situations, 15 times each. We asked plainly, asked for one pick, or asked for the pick in a follow-up.

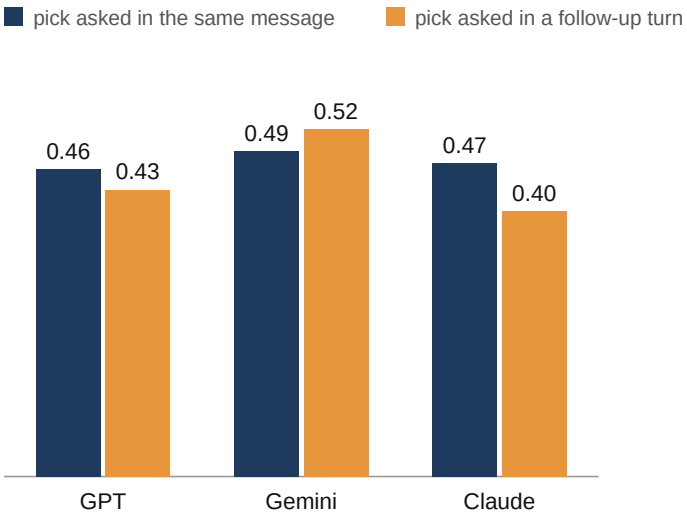
1. Asked once, are picks stable?

Average of 15 askings. GPT under the predicted 0.45; Gemini and Claude above it, from steady no-picks.



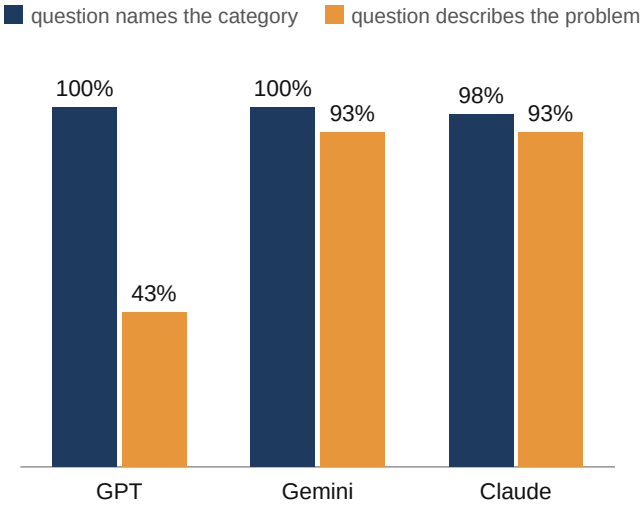
2. Does a follow-up turn help?

Single answers. No change past 0.10; a small shift for Gemini or Claude can't be ruled out.



3. Does phrasing matter?

Mostly for GPT. Gemini and Claude still name their leader 93% of the time.



Leader: each assistant's most-picked brand (exploratory).

Takeaway

Asked for one pick, each assistant narrows to about three brands and leans toward one.

Counting brand mentions in single answers mostly measures wording and chance.

Belief Stability Study, preregistered at osf.io/sn23j. Registration public since Oct 26, 2026; preprint forthcoming. Dhruv Walia, Nov 2026.