



YOUR PERSONAL AI BENCHMARK: REFERENCE CARD

REFERENCE CARD · FRONTIER LAB · OCTOBER 1, 2026

Nufar Gaspar · Frontier Lab · October 2026

Your own real requests, given to two to four AI options side by side, names hidden. An option is a model, a setting inside one product, or an AI agent tool.

THE WORK MIX: COVER WHAT YOU REALLY DO

1. **What kind?** Writing, analysis, research, advice, planning, making things.
2. **How often?** Daily habits and rare big moments.
3. **What's at stake?** Quick asks and costly mistakes.
4. **Starting from?** A blank page or your own material.
5. **Which life?** Work and personal.
6. **Wish list?** Something AI has never done well for you.

Pick work where you can spot a weak answer fast. Add one matter-of-taste task and one you can check. Write what good looks like first. Made-up material only.

THE TASTE TEST

1. **Choose** two to four options.
2. **Run** every request in each, in a fresh chat.
3. **Hide** the names.
4. **Read** all, then **pick** the one you would use, or score each from 1 to 5 (few 5s). One line on why: clear or close?
5. **Reveal** and count.
6. **Decide.**

FOUR WAYS TO HIDE THE NAMES

A friend who keeps the key. A chat in a tool you are not testing. A spreadsheet sorted by a random number. The blind-shuffle script, also for images and websites.

THE DECISION

Finding your lineup: your AI lineup and default. **Testing something new:** **Switch** (new default), **Split** (joins your lineup for some tasks), or **Stay** (test again later).

1. **Must-haves:** allowed at work, privacy terms, on your plan.
2. **The answers:** who won, clear or close.
3. **Tie-breakers:** cost, speed, features you enjoy, plan limits, habits.

A clear win on tasks that matter moves you. A close call goes to the tie-breakers.

FOUR HABITS THAT KEEP IT FAIR

1. Same request, same settings, fresh chat.
2. Hide the names before you pick.
3. Notice the cost of getting to an answer you accept.
4. Test on day one; trust online reactions after a week.

EVERY TIME SOMETHING NEW COMES OUT

Rerun the same requests, unchanged. Hide the names, pick, decide. Refresh the benchmark only when your work changes, a new kind of AI ability arrives, or a task stops separating the options.

Optional: the scored version adds a scoring guide and a strict AI judge; your scores decide. The automated version runs it all.

Prompts 1 to 7 are in the lab.