

What makes a good AI skill: the 23 elements every skill needs

Use this when you build or review an AI skill for Claude, ChatGPT, Gemini, Copilot or any other AI tool. Most skills that fail are missing the process and the testing, not more instructions.

Start with what the skill is for

- ☐ 1 **One clear job.** The skill does one task or process well, not everything at once.
- ☐ 2 **Says what it does and when to use it.** A short description the AI reads to decide when to pick the skill. A vague one means the skill is used at the wrong moment, or never.
- ☐ 3 **A named owner.** One person keeps the skill current, handles feedback and approves changes.

Then the process inside it

- ☐ 4 **An agreed process.** The people who do the work agree on the steps and on what a good result is. Do not automate a process that is still disputed.
- ☐ 5 **Defined steps.** The work broken into clear steps, in order.
- ☐ 6 **A data source for each step.** Which system, document or person each step takes its information from.
- ☐ 7 **The decision points.** The ifs and conditions, including the rules nobody ever wrote down.
- ☐ 8 **A check after each step.** How the AI confirms a step was done correctly before it moves on, and what to do if it was not.
- ☐ 9 **A quality check on the result.** The criteria the final output must meet before it is handed over.
- ☐ 10 **Examples of good output.** Concrete examples, or a template, of what a good result looks like.

Write only what the AI doesn't know

- ☐ 11 **Only what the AI doesn't already know.** Your company's knowledge, not general explanations. More detail does not make a better skill.
- ☐ 12 **Strict where mistakes are costly.** Exact instructions for fragile or high-stakes steps, and room for judgement where many approaches work.
- ☐ 13 **One term for one thing.** The same word for the same concept throughout, so the AI never has to guess.
- ☐ 14 **Nothing that goes out of date.** No dates or temporary rules buried in the instructions. Old ways of working are kept apart and clearly marked.

Test it before anyone else uses it

- ☐ 15 **Test cases with expected results.** At least three real tasks with the result you expect, written before the skill is finished.
- ☐ 16 **A before-and-after comparison.** How well the AI does the task without the skill, so you know the skill actually helps.
- ☐ 17 **Automatic scoring of the output.** An AI judge scores each result against your quality criteria, so testing does not depend on someone reading every output.
- ☐ 18 **Tested in every AI tool you use it in.** A skill can behave differently in Claude, ChatGPT or Copilot, and even between models from the same vendor.
- ☐ 19 **Tested by others on real work before sharing.** Colleagues use it on real tasks. Working twice for its author is not enough.

Keep it good after it ships

- ☐ 20 **Re-tested when the AI changes, and on a schedule.** New AI models change behaviour. A skill approved last month can quietly get worse.
- ☐ 21 **A feedback route to the owner.** People who use the skill report problems and suggest improvements where the owner sees them.
- ☐ 22 **Every change versioned and reviewed.** Each improvement becomes a new version, reviewed before it reaches people, with the previous one kept.
- ☐ 23 **A check that it is actually used.** Whether people and AI tools use the skill, or it just sits installed.