

BUILD & IMPROVE

Break it, then make it better

Your agent works when you test it. What happens when someone else tries to break it?

An agent that only handles easy questions isn't ready. Try to break your own agent with awkward, vague and off-topic questions, then improve its instructions until it holds up.

1 Try it your way

Open the agent you built. If you haven't built one yet, start with Scenario 1.

[Try in Copilot →](#)

2 Need a cue?

Start with the outcome you need:

[Edge cases](#)[Off-topic](#)[Fixes](#)

3 Still stuck? Reveal a sample prompt

Have a go first — the sample prompt is on the next page when you want it.

[Reveal sample prompt ↓](#)

4 Did you get something useful?

- ☐ I tested awkward and vague questions
- ☐ I found at least one failure
- ☐ I changed the instructions to fix it
- ☐ The fix held up on a retest

5 Take it one step further

Ask Copilot: "Suggest five difficult questions a real user might ask this agent."

Need a little help?



[Watch a quick video](#)
(3 mins)



[See a worked example](#)



[Take the learning path on ASN](#)

🏆 Top tips

- ✓ Use your own work if you can.
- ✓ Be specific about the output you need.
- ✓ Iterate — your first prompt doesn't have to be perfect.
- ✓ Check the result — would you actually use it?

Your notes — what worked, what you'd change

BUILD & IMPROVE

Sample prompt

Break it, then make it better

“ Act as a sceptical user of this agent. Give me ten test questions: three vague ones, three that are off-topic, two that need information it doesn't have, and two designed to make it guess. Then suggest instruction changes to handle the failures.

Select the text above and copy it into Copilot.

Why this works

- ✓ It turns testing into a structured stress test, not random poking.
- ✓ It targets the failure types that matter most: vague, off-topic and guessing.
- ✓ It ends with fixes, so testing leads directly to improvement.

[Try it in Copilot →](#)[← Back to the challenge](#)

BUILD & IMPROVE

Worked example

Illustrative example — names and details are fictional

The situation

Your Finance Ops Onboarding Helper works well on the 20 standard questions. Before you share it with the whole Woodgrove team, you want to know where it falls over.

What Copilot returned

 Copilot

Here's what happened when the agent was stress-tested.

WHAT BROKE

- A vague question — 'how do I do expenses?' — got a long, unfocused answer.
- Asked about holiday policy, it guessed instead of saying it's not covered.

INSTRUCTION CHANGES

- Ask a clarifying question when a request is vague.
- Only answer from the attached documents; otherwise say so.

RETEST

- Vague questions now prompt a short clarifying question.
- Off-topic questions get a clear 'not covered' with who to ask.

What to check

- ✓ Failures are useful: each one tells you which instruction is missing.
- ✓ Retest after every change — fixes can break other answers.
- ✓ Ready to share? If it survives a sceptic, it's ready for the Agent-a-thon.

[← Back to the challenge](#)