

TRY A LOCAL CODING MODEL

Check capability, endpoint and output before trusting the setup.

Copilot CLI can discover supported models from a running Ollama instance through `/model`. Ollama and the model must already be installed, and the model must support tool calling and streaming.

Make a controlled first run

1. Check the current Copilot CLI requirements in the linked announcement. Install the runtime and a compatible model through their official instructions.
2. Start the local runtime, open `/model`, and review the discovered provider and endpoint before adding a model.
3. Use a small sample project. Ask for a single change with a test you can run yourself.
4. Record response time, memory use and whether the original test passes. Compare the same task with your usual model.

Copy this test

In this sample project, fix [small bug]. Keep the tests unchanged. Work only in [files]. List the changes, then give me the exact command to verify the behavior. If a tool call fails, report it instead of claiming success.

Check the data path

1. A local model does not automatically make every tool or connector local. Review network access and integrations separately.
2. Do not put private files into the test until you understand which services receive data.

ORIGINAL RESOURCE

[GitHub's local model discovery announcement](#)

<https://github.blog/changelog/2026-10-07-discover-local-models-in-github-copilot-cli/>

An independent companion by Huzaifa Shoukat. Check the original project for current setup and availability. Prepared 12 October 2026.