Policy is not doing what I expect

An MCP server is blocked that I allowed

Check the precedence. A **per-tool** rule beats a per-server rule. If you allowed the server but blocked one of its tools, that tool stays blocked.

Also: when a single request names several MCP servers, **the most restrictive decision applies to the whole request**. The Gateway cannot partially block one API call. One blocked server in the set blocks it all.

A server I never approved is working

Approval only applies if **Require approval for new MCP servers** is on. Without it, the tenant default applies — and if that is Allow, everything works.

Remember that "reviewed" means *any explicit rule*. There is no separate approve action.

Nothing is being blocked at all

Check the detector actions on **Policy ? Risk detectors**. The defaults are mild by design: secrets warn, credentials and personal data report, lethal-trifecta warns. Report means recorded and not interrupted.

Filter Activity by **Risk events** to confirm the detectors are firing — if findings appear but nothing is blocked, the detectors are working and the actions are set to observe.

Too much is being blocked

Usually the classifier sensitivity is one step too high. Move **High** to **Medium**. The clearest cases are caught at every setting; High mostly adds the ambiguous ones.

Check **Blocked conversations** at the bottom of Policy — you can unblock individually.

A policy change has not taken effect on a developer's machine

Local MCP decisions are cached for as long as **Policy ? Local MCP decision cache** allows. Set it to 0 while you are actively tuning, then put it back.

Revision #6
Created 2026-08-02 10:20:20 UTC by Sentilai Docs
Updated 2026-08-04 08:40:00 UTC by Sentilai Docs