How a policy decision is made

Several things can have an opinion about a single request. This is the order they are consulted, so you can predict what will happen before you change anything.

For MCP tool calls

1. A **per-tool rule** — the most specific thing you can write.
2. A **per-server rule**.
3. The **pending action**, if approval is required for new servers and this one has no explicit rule.
4. The **tenant default**, `MCP default action`.

The first one that matches wins.

When one request names several servers

A single AI request can declare tools from more than one MCP server. The Gateway cannot partially block one API call, so **the most restrictive decision across all named servers applies to the whole request.** One blocked server blocks the request.

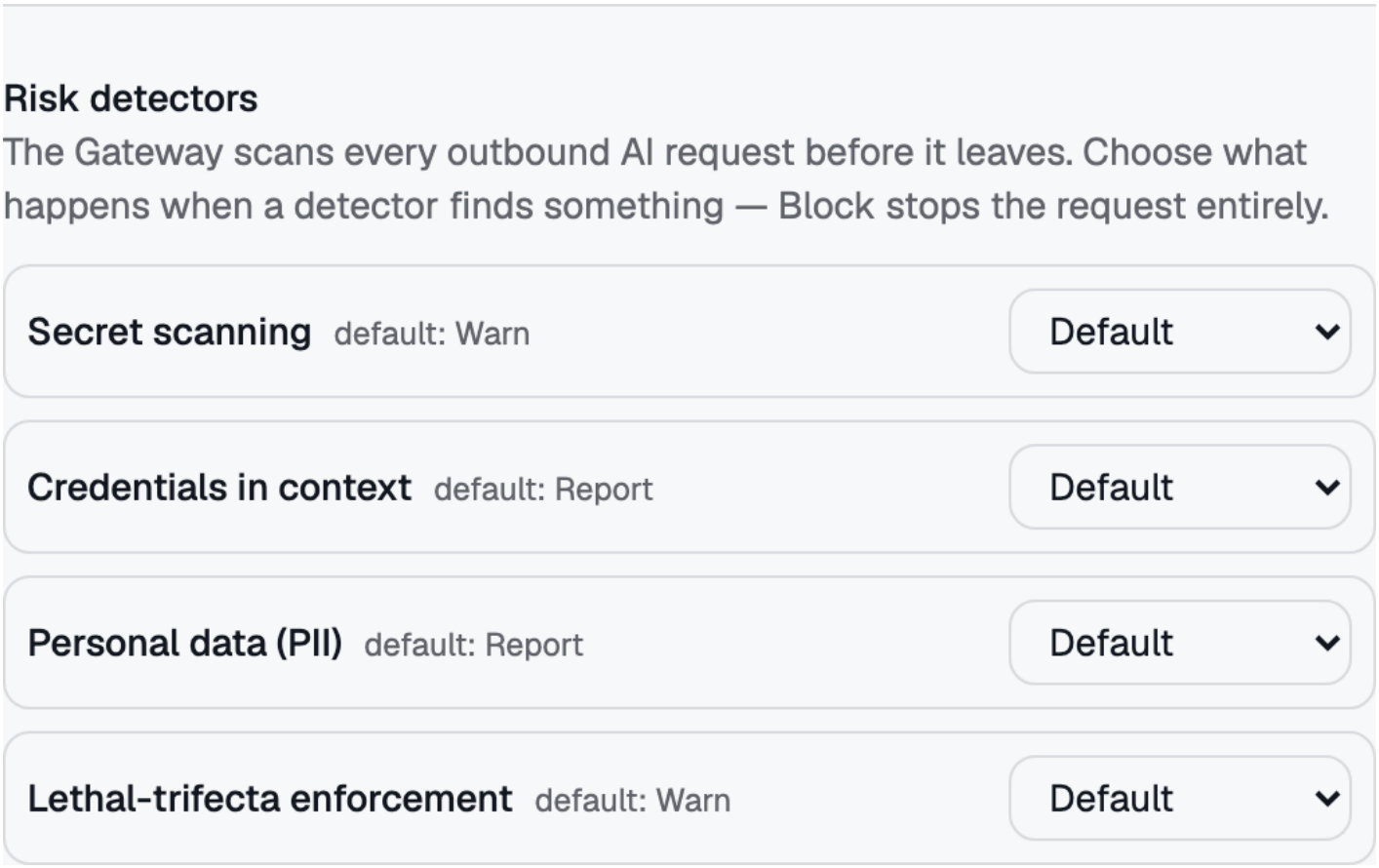
Risk detectors are separate

Detectors run on the content of the request, independently of MCP rules. Each has its own action, and the strongest action taken by any detector determines the outcome. A request can be allowed by MCP policy and blocked by secret scanning.

The actions

- **Allow** — nothing recorded beyond the normal audit row.
- **Report** — recorded as a finding, visible in Activity and Compliance. Nothing is interrupted.
- **Warn** — recorded more prominently and, where the path allows it, surfaced to the developer.
- **Block** — the request or tool call does not proceed.

Report is the setting you want while you are still learning what your team does. It gives you the same visibility as blocking with none of the disruption.



Risk detectors and their actions. The strongest action any detector takes decides the request.

Revision #10
Created 2026-08-02 10:20:00 UTC by Sentilai Docs
Updated 2026-08-05 06:29:26 UTC by Sentilai Docs