Secret scanning

Stops credentials from leaving your organization inside a prompt. **Default: warn.** This is the detector most organizations move to **block** first, and the one where doing so is least controversial.

What it looks for

Recognizable, high-confidence credential formats:

- AWS access key IDs
- GitHub tokens and personal access tokens
- PEM private keys
- Anthropic API keys (`sk-ant-…`)
- OpenAI API keys (`sk-…`, `sk-proj-…`)
- Slack tokens
- Google API keys
- JSON Web Tokens

These are matched on shape, so the false-positive rate is low — but an example key in documentation your developer pasted will trip it, which is the right outcome anyway.

What is recorded

The kind and the count. Never the value. A finding says "one AWS access key id was present"; it does not store the key, an offset, or a preview. Building a security product that quietly collects a database of everybody's leaked credentials would be indefensible, so the detector is built so that it cannot.

The sensitivity dial does not apply

Detected secrets are handled according to **this detector's** action regardless of where the risk classifier sensitivity is set. The dial governs the semantic classifier, not pattern matching.

Setting it

Policy ? Risk detectors ? Secret scanning.

Revision #6
Created 2026-08-02 10:20:00 UTC by Sentilai Docs
Updated 2026-08-04 08:39:43 UTC by Sentilai Docs