Personal data (PII)

Flags personal data appearing in prompts. **Default: report.**

What it detects

- Email addresses
- International phone numbers, where a country prefix is present
- IBANs, **validated by checksum** rather than merely shaped like one
- Payment card numbers, **validated with the Luhn check**

The checksum validation on the last two matters: it means a random 16-digit number in a test fixture does not become a card-number incident.

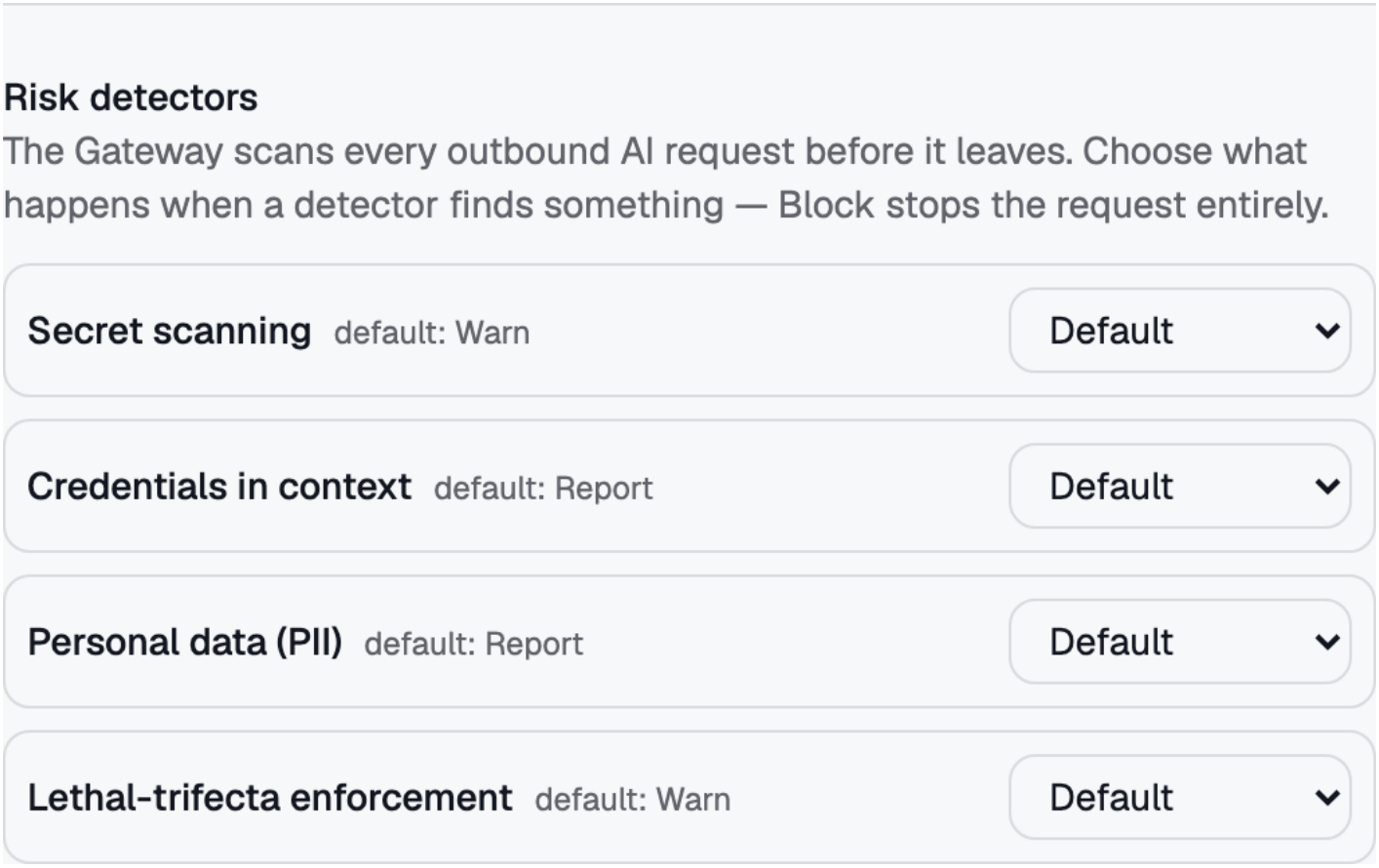
Why you would care

Two reasons, usually. Under GDPR, sending customer personal data to an AI provider is a processing activity you need to have thought about. And separately, most engineering teams simply do not intend to paste production data into a chat window — this is how you find out that it happens.

Tuning

Start at **report**. Look at where the findings cluster: if they come from one team debugging with real data, that is a conversation, not a policy setting. If they are spread evenly, consider **warn** so developers get feedback in the moment.

Block is aggressive for this detector — an email address in a prompt is often completely legitimate — so most organizations stop at warn.



Personal data (PII) is a detector row like any other — choose Warn to coach, Block to enforce.

Revision #11
Created 2026-08-02 10:20:01 UTC by Sentilai Docs
Updated 2026-08-06 04:31:35 UTC by Sentilai Docs