THE C.A.I.S.O. AGENT

Using the C.A.I.S.O. Agent

A step-by-step guide to deploying and working with your Chief AI Security Officer agent, powered automatically by Blaick.ai.



Inside: what you get · what to prepare · the 7-step process · your daily routine · where humans stay in charge · checklist, FAQ and glossary

1 What you are getting

An always-on Chief AI Security Officer for your organization.

The C.A.I.S.O. Agent finds the AI systems you run, watches them around the clock, responds to threats, and reports to your team. It uses Blaick.ai as its security engine and connects to it automatically, so there is no separate security product to set up.



See everything

A living inventory of chatbots, copilots, model APIs, internal agents and third-party AI services in use.

Stop threats

Detects and contains prompt injection, data leakage and compromised model supply chains.

Prove control

An auditable timeline of every detection and action, ready for leadership and auditors.

Stay in charge

You decide what the agent may do alone and what needs a human to approve.

Why it matters. Organizations adopt AI faster than they secure it. An agent that never sleeps closes the gap between the two without adding headcount.

2 Before you start

Five things to line up. A rough version of each is enough.

An owner

The person accountable for AI security: often the CISO, head of IT, or operations lead. They approve high-impact actions and receive the reports.

Operations access

Someone able to deploy software into your environment. A test or staging area is a great place to begin.

A list of your AI tools

Chatbots, copilots, model APIs, internal agents and third-party AI services your staff use. The agent will find what you forget, so don't worry about completeness.

Your policies

Existing security, privacy or acceptable-use rules. No policies yet? Baseline protections are applied by default.

Contacts for alerts

Who should be told, and how: email, chat, or on-call. Name a backup for when the owner is away.

Quick readiness questions

- Which AI tools hold customer or staff data?
- Who may connect a new AI service today?
- What would be the worst thing an AI tool could leak?
- Who is called at 2 a.m. if something happens?

3 The process (steps 1 to 4)

1

Request a deployment

Contact us with a short description of your environment. We confirm scope and help your operations team plan the rollout.

You provide: a few lines on your setup and your goals. Typical time: a short conversation.

2

Operations deploys the agent

Your ops team launches the agent into your environment, starting with a test or staging area if you prefer. Illustrative example:

```
caiso deploy --env production
```

You provide: the environment and an operations contact.

3

Blaick.ai attaches automatically

The agent provisions Blaick.ai as its security engine. There is no separate configuration, licence or integration step for you to perform.

You provide: nothing. This step is hands-off.

4

The agent maps your AI

It discovers AI systems, data flows and risks, and builds an inventory. Review it and correct anything it missed or mislabelled.

You provide: a quick review. Mark each item approved, restricted or unknown.



3 The process (steps 5 to 7)

5

Apply your policies

Baseline protections are applied by default. Add your own rules: what data may reach which models, who can use which tools, and what must always be blocked.

Examples: "customer records never go to public chatbots"; "only the finance team may use the forecasting model".

6

Set your approvals and alerts

Decide which actions the agent may take on its own (for example, blocking a suspicious request) and which need a human to approve (for example, shutting down a production system). Add your alert contacts.

Start cautious and loosen as trust grows.

7

Continuous protection

The agent detects, responds and reports 24/7, escalating the big decisions to your team.

From here on, your effort is a few minutes a day.

What a typical first month looks like

Week	Focus
1	Deploy to test, review the discovered inventory
2	Apply policies, set approvals and alert contacts
3	Move to production, watch the daily summary
4	First monthly report review; tune policies

Tips for a smooth rollout

- **Start in staging.** Prove the setup, then promote to production.
- **Be strict at first.** Require approval for more actions early on.
- **Tell your staff.** A note that AI use is monitored avoids surprises.

4 Day-to-day use

A light routine keeps you in control.

When	What you do
Daily	Skim the activity summary; act on any items marked for approval.
Weekly	Review new AI tools discovered and decide whether to approve or restrict them.
Monthly	Read the security report; review incidents, trends and policy changes.
When something changes	Adding a new model, vendor or AI feature? Tell the agent so it can assess and cover it.
After an incident	Review the agent's timeline of what was detected and done, then tighten policy.

Example daily summary (illustrative)

✔ 12 AI tools monitored, no active threats
✔ 1 prompt-injection attempt blocked on support chatbot
! New AI service detected in Marketing — awaiting your approval
! Policy suggestion: restrict file uploads to public models
✔ Audit trail updated

> **Tip.** Ten minutes with the daily summary and a weekly look at new tools is enough for most teams. The agent handles the rest.

5 What the agent does for you



Inventories

AI systems and data flows, kept up to date as your staff adopt new tools.

Detects

Prompt injection, data leakage and compromised model supply chains.

Contains

Acts quickly to limit damage and keeps an auditable record of everything.

Enforces

Applies your policies consistently, across every tool, every hour.

Reports

Plain-language updates for leadership, boards and auditors.

Learns

Suggests policy improvements after incidents and as your AI use grows.

Without it, AI risks tend to be found after the damage. With it, they are found and handled as they appear.

6 Where humans stay in charge

> The agent handles routine detection and response on its own. High-impact decisions are always escalated to your team for approval, and you can adjust that boundary at any time.

The agent does on its own	It asks a human first
Blocks a suspicious request	Shuts down a production system
Logs and categorizes events	Revokes a team's access to a tool
Flags a newly found AI tool	Approves a new tool for sensitive data
Prepares reports and timelines	Changes a core policy

Examples are illustrative. You set the real boundary during step 6.

Escalation path

1 · Detect

The agent spots something unusual and contains it if your rules allow.

2 · Notify

Your named contacts are alerted by email, chat or on-call.

3 · Decide

The owner approves, declines or adjusts the proposed action.

4 · Learn

The timeline is reviewed and policy tightened if needed.

7 Getting started checklist

☐ Name an owner and an operations contact

☐ Gather a rough list of AI tools and existing policies

☐ Request a deployment

☐ Deploy to a test environment, then production

☐ Review the discovered inventory

☐ Set approvals and alert contacts

☐ Schedule a monthly report review

Common questions

Do I need to buy or set up Blaick.ai separately?
No. The agent connects to it automatically.

Can I start small?
Yes. Begin in a test area, then expand.

Will it act without asking me?
Only on routine matters you have allowed. Big decisions come to you.

What if it finds a tool I didn't know about?
It appears in your weekly review to approve or restrict.

Glossary

Prompt injection	Hidden instructions that trick an AI into ignoring its rules.
Data leakage	Sensitive information escaping through an AI tool.
Model supply chain	The models and components your AI depends on, which can be tampered with.
Blaick.ai	The security engine the agent uses automatically.

The deployment command shown is illustrative. © 2026 The C.A.I.S.O. Agent · Security engine: Blaick.ai (blaick.ai)