Synthia Agents - Security overview

/security - 2026-11-02

Control everything before you delegate

Every action is decided by your policy before it happens and recorded after. Approvals, boundaries, audit trail and budgets in one place.

Everything an agent needs. Nothing it shouldn't have.

The convenience of a familiar AI assistant, with what a company needs before it lets an agent near its systems.

A computer per coworker

Its own browser, workspace and browser profile, so it stays signed in between turns. Watch it live, and take the wheel at a login wall or 2FA prompt.

Approvals that escalate

Consequential actions wait for a person, escalate to a backup, and expire safely.

Your rules, decided first

CEL policies that fail closed, with version history, rollback and optional four-eyes approval.

An audit trail you can read

Search it, export it to CSV, stream it to your systems with signed webhooks.

Usage & budgets

Tokens and cost per coworker, with monthly caps.

Any model, per coworker

Choose the provider and model for each coworker.

Governed connectors

Google Drive, Notion, Composio and your own MCP servers - granted tool by tool, with health checks.

What it won't do

Each of these is a control in the code, not a promise in a prompt.

Type your passwords

A field or setting named as a password, passcode, two-factor or verification code is handed to you. You sign in through a private form the coworker never sees, and password fields are hidden from everything it reads back.

Pay for something

A pay, checkout, purchase or place-order button, or a card number, CVV or IBAN field, is handed to you. No rule you or your team saves can let the coworker do it.

Reach cloud metadata addresses

Whatever the network policy says, even allow all, the computer refuses link-local and cloud metadata addresses, where a machine's credentials are handed out.

What we defend against, and what we don't

The risks the workspace is built to resist, how it resists them, and what is out of scope.

Instructions hidden in pages and files

Text a coworker reads on the web or in a document is treated as information, not as orders. Attempts to slip in instructions are detected and flagged, and risky actions still wait for a person's approval.

Data leaving without permission

Each coworker's computer works under its own network policy. Outgoing messages and files are checked for secrets and personal data, and internal and cloud credential addresses are refused.

Stolen sign-ins and keys

People type passwords into a private form the coworker never sees. Stored secrets and page pictures are encrypted at rest, and sessions end when idle, on a sudden change of browser or network, or when you sign out everywhere.

Tampering with backups

Snapshots are encrypted, checksummed and listed in a signed backup list. A changed list is not trusted without your confirmation, and restores can require two-factor.

Quiet misuse from inside

Every administrator change and coworker action is written to a tamper-evident audit trail that can be exported and checked.

Out of scope

A signed-in person's own device or browser being taken over, a model or hosting provider acting against you, and attacks that need physical access to the servers.

Report a vulnerability

- This page as a PDF
- Security questionnaire answers
- Data processing addendum template
- How the product uses AI
- Where your data is kept