SANCTUMSHIELD · AGENTIC GOVERNANCE FOR THE OTHER 99%

AI Agent Acceptable Use Policy — One-Pager Template

Audience: mid-market organizations (50–2,000 employees) with a 1–3 person security/IT team and no platform-engineering group. **Last updated:** November 9, 2026

> **What this is.** A one-page, fill-in governance record you complete **per AI agent or agent class** (e.g., ChatGPT Team, Microsoft Copilot, a Salesforce agent). It is a *distillation* of the full SanctumShield AI Acceptable Use Policy **§7 (AI Agent and Agentic Workflow Policy)** — designed so a non-technical owner can complete it in under 30 minutes without buying any software. The full AUP remains the governing document; this page is the per-agent operational record §7.2 (Agent Identity Registration) asks you to keep.
>
> **The core idea (Cunningham):** *the agent's job description IS the policy.* Everything the agent may touch is written down; everything else is prohibited by omission. Any action outside the job description is a reportable incident — detected from the **first off-spec tool call**, not after the fact.

How to use this template — in plain English

What this is: a one-page record you fill in for each AI tool or agent your organization uses, so everyone knows what it is allowed to do and who is responsible for it. If you can describe a job in a sentence, you can fill this in.

Why the wording is precise (and why that helps you): the fields and terms are exact on purpose — they line up with real AI-governance standards (like the CSA Agentic Trust Framework). That precision is what makes a completed page hold up in front of an auditor, a board, or an insurer. You do **not** need to be a security expert to use it: just answer each field honestly, in your own words.

How to fill it in (about 30 minutes):

1. Make one copy for each AI tool or agent (ChatGPT Team, Microsoft Copilot, a Salesforce agent, and so on).
2. Write its **job in one sentence** — that sentence is the rule. Anything outside it is off-limits.
3. List exactly what it may touch (which data, which systems). Everything you don't list is prohibited.
4. Say how it signs in — its **own account** or a **shared password/key** (a shared key is a weakness worth noting).
5. Pick an **autonomy level** from the back page. New tools start at **Intern** (look-only).
6. Name **one accountable person** — a name, not a team.
7. Write **3–5 "never do this" lines**, then sign it and set a review date.

What this is NOT (the honest limits):

- It is **not legal advice**. It helps you *do* and *show* good governance — it does not declare you compliant or "safe."
- It does **not guarantee** any legal or insurance outcome. The named person stays accountable no matter what.
- There is **no software to install**, and SanctumShield does **not** run or control your agents. This is a record *you* keep.

FRONT OF PAGE — the record (one per agent)

Field	What to write	Your entry
Agent name & vendor	What it is and who makes it.	`_____`
Agent identity (non-human)	How is this agent identified and authenticated to your systems — *separate from the person who owns it? * Does it have its **own unique identity/credential**, or a **shared API key / login**? Mature form: a machine identity (SPIFFE/SVID, Microsoft Entra Agent ID) or the identity your IdP / AI-	`☐ Own unique identity: _____ ☐ Shared key/login (gap → plan: _____)`

Field	What to write	Your entry
	governance platform assigns per agent. **A shared key is a gap** — note it and a plan to move to a unique identity.	
Job description (persona)	**One sentence:** what was this deployed to do? *This sentence is the policy — write it as if it's the only rule.*	`_____`
Tool envelope	Exactly what it may touch — which data, which systems, which actions. **Everything not listed here is prohibited.** If it answers from company documents for a filing, a contract or a consequential decision, each answer cites the document, version and page or section.	`_____`
Connections (MCP gateway or connectors)	The gateway or connectors that give it reach into business systems, and who owns that connection. One gateway can reach many systems, so list the systems behind it.	`Gateway: _____ Systems: _____ Owner: _____`
Other agents	Can it start subagents or hand work to other agents? If so, where those agent-to-agent exchanges are logged.	`☐ No ☐ Yes, logged in: _____`
Changes it may make	The systems it may write to; whether it runs on a schedule; whether it may keep running past its usage quota; and the named person who approves the changes it proposes.	`Writes: _____ Schedule: _____ Approver: _____`
Spend owner and cap	Who owns what it spends, the cap, and confirmation that the platform's terms for agents are reflected in your AUP.	`Owner: _____ Cap: _____ ☐ Platform terms in AUP`
Who can stop it	The named person who can stop or pause it mid-task, the date that stop	`Name: _____ Last proven: __/__/____ ☐ Changed since`

Field	What to write	Your entry
	was last exercised in a test or for real (or "never"), and whether the agent has changed since.	
Autonomy level	Intern / Junior / Senior / Principal (see back of page). New agents start at **Intern**.	`☐ Intern ☐ Junior ☐ Senior ☐ Principal`
Named human owner	One accountable **person** — a name, not a team, not a distribution list.	`_____`
Prohibited actions (3–5 red lines)	Explicit "never" list. Common examples: no outbound email; no credential/secret access; no bulk export or delete; no production writes; no spend without approval.	`1. _____ 2. _____ 3. _____`
Off-spec = incident	Confirm: any action outside the job description above is reported through the same channel as a security event.	`☐ Acknowledged — reports to: _____`
Review date	Quarterly review by the named owner.	`Next review: ____ / ____ / _____`

Owner sign-off: `Name _____ Signature _____ Date _____`

BACK OF PAGE — the autonomy ladder (definitions)

Four levels of *earned* autonomy. Autonomy is **granted, not assumed** — every new agent starts as an **Intern** and is promoted only after it demonstrates it can be trusted at the next level (the curriculum's Autonomy Ladder module covers the promotion gates). Higher autonomy requires stronger controls and higher approval authority.

*Framework basis: the four-tier earned-autonomy model in the CSA **Agentic Trust Framework** (Feb 2026) and the autonomy levels in CSA's "Levels of Autonomy for Agentic

AI"; persona/off-spec detection from Cunningham, **Agentic Zero Trust** v3.0 (May 2026). See Sources.*

Level	What it may do	Human involvement	Who approves this level
Intern	**Read-only.** Access data, analyze, draft, summarize, recommend — but **cannot change any external system.** Worst case is a bad suggestion.	Human does every action the agent proposes.	Business owner
Junior	**Recommend specific actions** with reasoning; a human clicks "approve" before anything executes.	Explicit human approval on every action.	Owner + IT/security sign-off
Senior	**Execute within defined guardrails** and notify a human of what it did and why (real-time).	Oversight after the fact; humans spot-check.	Formal review (owner + security + a manager)
Principal	**Broad autonomy** within its envelope; continuous validation instead of per-action approval.	Continuous monitoring; humans intervene by exception.	Executive authorization + documented risk acceptance

Rule of thumb: if you can't name the person who would be paged when this agent does something wrong, it is not ready to be anything above **Intern**.

WORKED EXAMPLE (so anyone can fill it in) — "ChatGPT Team, marketing use"

Field	Example entry
Agent name & vendor	ChatGPT Team (OpenAI)

Field	Example entry
Agent identity (non-human)	☑ Shared Team workspace login via SSO — no per-agent machine identity. *Acceptable for a read-only Intern; revisit before it's given any tool/system access.*
Job description (persona)	*Drafts and edits marketing copy from material our team provides; it does not touch customer data or company systems.*
Tool envelope	Text the marketing team pastes in; web browsing for public research. **No** file connectors, **no** CRM, **no** email, **no** code execution.
Connections	None. No gateway or connectors are enabled.
Other agents	☑ No
Changes it may make	None; it drafts only. No schedule. Approver for any future change: Dana Reyes.
Spend owner and cap	Dana Reyes; the marketing team's seat count; ☑ platform terms reflected in our AUP.
Who can stop it	Dana Reyes removes workspace access; last proven 10 / 01 / 2026 in the quarterly access review; ☐ no change since.
Autonomy level	☑ Intern (read/draft only — a person publishes everything)
Named human owner	Dana Reyes, Marketing Manager
Prohibited actions	1. No pasting customer PII or non-public financials. 2. No connecting company drives/CRM. 3. No sending anything externally.
Off-spec = incident	☑ Acknowledged — reports to: IT helpdesk / security@ourco
Review date	Quarterly — next: 01 / 15 / 2027

*Second example to try on your own: Microsoft Copilot (Junior — proposes actions in email/docs, human approves) and a Salesforce agent (map its exact object/field permissions

into the tool envelope).*

How this connects to the rest of SanctumShield

- **Full policy:** this one-pager operationalizes AUP **§7**; keep the full AUP as the governing document.
- **Agent identity:** the non-human identity field maps to AUP **§7.2 (Agent Identity Registration)** and to Cunningham's Recommendation #2 (unique, attestation-backed identity; migrate shared keys). It's the same question a vendor must answer in the **Vendor Trust Questionnaire (question 1)** — asked here of your *own* agents.
- **Registry:** one completed page per agent = your **agent registry** (the curriculum's first module).
- **Board report:** the autonomy levels you assign here feed the autonomy-distribution chart in the **Quarterly Agent Governance Report**.
- **Buying:** when a vendor *is* the agent, screen it with the **AI Vendor Trust Questionnaire** before it earns a page here.
- **Glossary:** Agent Identity, Subagent Spawning, Scheduled Agent Task, Proven Stop and MCP Gateway define the terms used in the new rows.
- **Academy:** module M13 (Agent Identity, Access & Delegation) and M16 (Effective Human Oversight) teach this page; see the CISO track.

Guardrails honored

- No "certification" or accreditation language — this is an internal governance record, not an attestation of compliance.
- No guarantee of legal/insurance outcome. The named owner remains accountable regardless of vendor or platform controls.

Sources

- Cunningham, C. *Agentic Zero Trust* (Research Paper v3.0, May 2026), published by Cequence — https://www.cequence.ai/agentic-zero-trust/

- Cloud Security Alliance, *The Agentic Trust Framework: Zero Trust Governance for AI Agents* (Feb 2, 2026) — https://cloudsecurityalliance.org/blog/2026/02/02/the-agentic-trust-framework-zero-trust-governance-for-ai-agents
- Cloud Security Alliance, *Levels of Autonomy for Agentic AI* (Jan 28, 2026) — https://cloudsecurityalliance.org/blog/2026/01/28/levels-of-autonomy

---

© PIGENAI LLC · SanctumShield · sanctumshield.com · Last updated November 9, 2026 · This is an educational template, not legal advice.