SANCTUMSHIELD · AGENTIC GOVERNANCE FOR THE OTHER 99%

MSP AI Trust Questionnaire — 10 Questions to Ask Your Managed Service Provider

Last updated: November 9, 2026

mid-market organizations (50–2,000 employees) whose IT is run by a managed service provider — and providers who want to answer before they're asked.

> **Why this exists.** When your IT is outsourced, your provider's remote-management platform runs with the highest privileges on every machine you own — and those platforms are now shipping AI into alert triage and scripting. You didn't pick that AI, you can't see it, and you can't turn it off. But the compliance obligation is still **yours** — outsourcing IT does not transfer it. These ten questions are how you see past your provider, and how a good provider proves you don't need to.
>
> **How to use the guidance.** Under each question: ✅ **Good answer** (what a mature provider can show) and 🚩 **Watch out for** (the red flag). This is a decision aid for *your* decision; no score here approves a provider.

Scope — who this is for, and who's who

a security review, a renewal, onboarding a new provider, or answering a cyber-insurance renewal questionnaire. About 20 minutes.

— most of the confusion in these conversations comes from asking one party a question only another can answer:

- — you carry the regulatory and contractual obligations. Outsourcing IT does not move them upstream.
- — runs the remote-management tooling inside your environment and carries your service agreement.
- — the remote-management (RMM) or security-tooling software company that ships the platform and the AI inside it. Some answers originate here; where they do, the note points you to the paired questionnaire.

> **This is one half of a pair.** These ten are what *you* ask your provider. The other ten — the AI Vendor Trust Questionnaire — are what your provider should be asking their platform vendor (and what any buyer asks any AI vendor). **Ask your MSP these ten; then ask whether they've asked their vendor the other ten.**
>
> **Are you the provider?** Answer these before you're asked — the MSP AI Disclosure is the fill-in answer sheet that resolves all ten in advance and turns them into a trust asset.

How to use this questionnaire — in plain English

ten questions to ask the company that runs your IT, so you can tell whether AI is operating in your environment under control — or under nobody's — even if you are not a security expert.

each question and its "good answer" note is worded to demand something you can actually inspect — a policy, a record, an owner, a date — not a reassuring sentence. A provider that wants to look compliant can say "we monitor AI" and "we have human oversight" without producing a thing. The wording below is built to get past that.

1. Copy the ten questions into a plain email or your security-review doc (there is a ready-to-send version at the bottom).
2. Read each answer against the ✅ and 🚩 notes below it. A useful answer resolves to something you can inspect: the control, the artifact, the owner, and when it was last exercised.
3. Pay special attention to the three — questions (control what arrives · stop it fast · know where our data goes).
4. If you accept a weak answer, write down accepted it and why. That is your record.

- It informs *your* decision; no score here makes anyone "approved."
- It is or a guarantee. Use it alongside your normal procurement and legal review.

- Some of what a strong provider would show — a per-release bill of materials for AI components, a pre-agreed evidence artifact, an AI addendum to your agreement — is what you should *ask for*. Treat these as "what good looks like," not as things every provider has today.

---

1. What AI-driven features are currently active in our environment — and how do you learn when your platform vendor adds one?

- ✅ a current, specific list of AI-driven features live in your environment — across cloud, on-premises, and hybrid deployments alike — plus a described way the provider finds out when their platform vendor turns a new one on.
- 🚩 an inability to separate AI features from ordinary automation, "just the standard tooling," or no process for learning about new ones. AI features increasingly arrive inside existing modules with no separate toggle — "we'd know" without a mechanism is a guess.
- *Maps to: inventory prerequisite — you can't assign accountability for what nobody's listed; NIST AI RMF — Map.*

2. When your platform vendor ships a new AI feature, how much notice do you get — and what happens between that notice and it running on our endpoints? *(non-negotiable)*

- ✅ a written change-control policy for platform-vendor releases, a stated amount of notice, and a clear description of the gap between notice and it going live on your machines — backed by the last couple of real release notices.
- 🚩 "it just updates automatically" with no policy behind it. For cloud-delivered platforms, automatic arrival is common and honest — but a provider with *no* change control at all is telling you they can't govern what lands on your endpoints.
- *Maps to: OWASP Agentic Top 10 — ASI04 Agentic Supply Chain Vulnerabilities.*

3. Before an AI feature that can execute scripts or take action reaches our systems, what is your approval process — and where do we enter it?

- ✅ a documented approval step for AI features that can run code or take action, with a clear point where you (the client) enter it — and an example of the last time it was invoked.

- 🚩 "we'll mention it in the newsletter," or approval rights that sit entirely with the provider. Few operators concede per-client approval over their core platform — so treat *how they deflect* as the signal about where decision rights actually sit. Don't expect a "yes"; expect a straight description.
- *Maps to: change control / human-approval gates; CSA ATF earned-autonomy model.*

4. Can AI-driven actions be scoped or disabled for our organization specifically — or is it one setting across your whole client base?

- ✅ AI controls can be scoped or disabled for your organization (or a single endpoint), with administrative documentation showing each control's scope.
- 🚩 one global setting across the whole client base, or no clear answer. *Honest note:* many platforms can't do this yet — a straight "not yet, and here's our compensating control" beats a dodge. This question is upstream pressure on the platform vendors.
- *Maps to: CSA ATF segmentation / authorization.*

5. Can you produce a record of AI-initiated actions in our environment — what acted, on which asset, when, and under whose authority?

- ✅ the provider can produce, on request and on a monthly cadence, a record of AI-initiated actions — and can show you the , agreed in advance so you both know what's owed.
- 🚩 partial logs, records you can't get, "we'd have to check with the vendor," or an artifact that only appears after an incident. An example agreed up front is what makes it enforceable.
- *Maps to: CSA ATF monitoring; audit-evidence needs.*

6. Is a technician reviewing AI-generated scripts before they execute in our environment — or does generated code run unreviewed?

- ✅ a written policy that a technician reviews AI-generated scripts before they run in your environment, plus an actual approval record from the last 30 days.
- 🚩 generated code that runs unreviewed, or "the AI is very accurate" offered in place of a review step. Accuracy is not oversight.
- *Maps to: human oversight; CSA ATF earned-autonomy / human-approval gates.*

7. If the AI misfires, can you disable it for our organization or a single endpoint — without degrading the rest of the service? *(non-negotiable)*

- ✅ a documented "kill switch" that stops AI for your organization (or one endpoint), with a stated time-to-effect, a named authorized owner, and the date it was last tested.
- 🚩 "open a support ticket," a switch that only stops *future* invocations (not one already running), or one that can only be thrown platform-wide. A module-wide off-switch is a service outage, not a kill switch — and you need the off-switch to work *while the AI is misbehaving.*
- *Maps to: CSA ATF incident response / kill switches.*

8. When an AI feature processes our data, where does it go? *(non-negotiable)*

- ✅ a clear account of where your data goes when an AI feature processes it — ticket text, hostnames, file paths, session data — who receives it, under what terms, retained how long, and whether any of it is used to train a model, backed by a subprocessor list and data-processing terms.
- 🚩 no subprocessor list, "it stays in the platform" with no specifics, or no answer on training. For a regulated client, this is the actual exposure.
- *Maps to: data-protection obligations; supply-chain transparency.*

9. If your platform vendor or its AI subprocessor is breached, how many hours until we hear from you?

- ✅ a notification commitment stated — covering both the platform vendor and the AI subprocessor behind it — or a committed date to add one to your agreement.
- 🚩 "promptly," "as soon as practical," or no visibility into the AI subprocessor at all. Your own regulatory notification clock runs whether or not the vendor → provider → you relay is fast; "promptly" across three handoffs is a commitment to nothing.
- *Maps to: breach-notification flow-through; supply-chain transparency.*

10. Who at your company is the named owner accountable for AI behavior in client environments?

- ✅ a specific named person accountable for AI behavior in client environments.

- 🚩 "the team," "our platform vendor handles that," or no name. Trivially easy to answer if true, conspicuous if not — and it's the question the rest of this rests on.
- *Maps to: named-owner principle — mirrors the owner field on the Agent AUP One-Pager.*

Also worth asking

- Can we opt out of AI features and remain a customer in good standing — including any price or service-level difference? Get the answer in writing. A provider that can't answer has made the decision on your behalf.
- Does any AI in the ticket path infer priority or severity instead of taking what our user selected? If so, that inference becomes your SLA/SLR audit record — ask to see a ticket where it's recorded.
- How the AI behaves when it's denied access ("stop and log" vs. "retry another way") is product behavior set by your provider's *platform vendor*. That's question 4 on the paired AI Vendor Trust Questionnaire — ask your MSP whether they've put it to their vendor.
- Does your provider's platform reach our business systems through an MCP gateway or connectors? Ask for the list of systems behind each connection and who owns it.
- For any AI agent that can act on our systems, who can stop it mid-task, and when was that stop last exercised? Ask for the date and the name. A proven stop has both.
- Do technicians use personal AI accounts or personal devices with our data? Ask for the provider's written rule.

If your provider is a security provider (MSSP) — two more

When the provider is a security provider, two things change: the AI doesn't just draft scripts — it can isolate hosts, disable accounts, and suppress alerts (autonomous action against production) — and the MSSP often produces your compliance evidence, so the AI can end up grading its own homework.

- Which response actions can AI take versus recommend to an analyst — isolation, account disablement, alert suppression, detection tuning? Ask for the current

playbook inventory with the autonomy level marked for each.

- When AI suppresses or downgrades an alert, is that decision retained, reviewable, and present in the evidence you produce for our audits? Ask to see an example.

How to read the answers

Count the ✅s, but don't turn it into a certificate. Use this rule of thumb:

- — control over what arrives on your endpoints, the ability to stop it fast, and knowing where your data goes. A provider who fumbles these is running AI in your environment they can't fully account for.
- Weak answers aren't automatically disqualifying, but each one is a risk you're accepting on the provider's behalf. Write down who accepted it.

Email wrapper (copy-paste)

▼▼▼ COPY BELOW ▼▼▼ Subject: AI in our environment — 10 questions for our provider

Hi [provider],

As part of our security review, we're asking the same ten questions about AI running in our environment. Straight, specific answers — a policy, a record, an owner, a date — help us move quickly. Where a control isn't available yet, just say so and note the plan.

1. What AI-driven features are currently active in our environment — and how do you learn when your platform vendor adds one?
2. When your platform vendor ships a new AI feature, how much notice do you get, and what happens between that notice and it running on our endpoints?
3. Before an AI feature that can execute scripts reaches our systems, what is your approval process, and where do we enter it?
4. Can AI-driven actions be scoped or disabled for our organization specifically, or is it one setting across your whole client base?
5. Can you produce a record of AI-initiated actions in our environment — what acted, on which asset, when, and under whose authority?

6. Is a technician reviewing AI-generated scripts before they execute in our environment, or does generated code run unreviewed?
7. If the AI misfires, can you disable it for our organization or a single endpoint without degrading the rest of the service? What's the time-to-effect, who's authorized, and when was it last tested?
8. When an AI feature processes our data, who receives it, under what terms, retained how long, and is any of it used for training?
9. If your platform vendor or its AI subprocessor is breached, how many hours until we hear from you?
10. Who at your company is the named owner accountable for AI behavior in client environments?

Thank you, [name] ▲▲▲ COPY ENDS ▲▲▲

Guardrails honored

- A decision aid: no score here attests compliance or safety; it informs *your* risk decision.
- No guarantee of outcome. Framework references describe what to look for, not a warranty.
- The bill-of-materials, example-artifact, and agreement-addendum mechanisms are what to *ask for* — described as good practice, not as controls that every provider or platform ships today.

Sources

- OWASP GenAI Security Project, *OWASP Top 10 for Agentic Applications (2026)* (ASI01–ASI10) — https://genai.owasp.org/resource/owasp-top-10-for-agentic-applications-for-2026/
- Cloud Security Alliance, *The Agentic Trust Framework* (Feb 2, 2026) — https://cloudsecurityalliance.org/blog/2026/02/02/the-agentic-trust-framework-zero-trust-governance-for-ai-agents
- Cunningham, C. *Agentic Zero Trust* v3.0 (May 2026), Cequence — https://www.cequence.ai/agentic-zero-trust/
- NIST, *AI Risk Management Framework (AI RMF 1.0)* — https://www.nist.gov/itl/ai-risk-management-framework

ACKNOWLEDGMENT · Practitioner red-team review by Nate Moore, CISSP (linkedin.com/in/natemoore).

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