
ENTERPRISE AI RISK
Agent identity is the new shadow IT
AI GOVERNANCE · ENTERPRISE RISK · KARIMSAIPLAYBOOK
ACCOUNTABILITY GAP
Nobody owns them. Nobody audits them.
And unlike a rogue SaaS account,
these agents are taking actions.
karimsaiplaybook.com
#AIGOVERNANCE

PRACTITIONER GUIDE

Agent Identity Risk

What Enterprise Teams Need to Ask Before Deploying Agentic AI

Shadow IT used to mean a finance team spinning up a SaaS tool nobody in IT knew about. Agentic AI has made that problem look quaint.

Now you have agents operating under user identities, service accounts, or delegated permissions that were set up once and never reviewed. Nobody owns them. Nobody audits them. And unlike a rogue Dropbox account, these agents are taking actions. Sending emails. Accessing files. Triggering workflows.

This guide is for enterprise architects, IT leaders, security teams, and governance practitioners who are either preparing to deploy agentic AI or who already have agents running and want to understand what they may have missed.

It does not assume a particular platform. The principles apply whether you are working with Microsoft Copilot agents, custom-built AI workers, or third-party agentic tools integrated into your environment.

PART ONE

The Questions You Need to Answer First

Before any agent is deployed, six questions should be answerable by someone in your organisation with authority and accountability. If they are not answerable, deployment should not proceed.

1. Identity: What or who is this agent acting as?

An agent needs an identity to operate. That identity determines what the agent can access, what it can send, and what audit trail it leaves. The question is not just technical. It is a governance question.

- Is the agent using a named user identity, a service account, or a delegated permission set?
- If it is using a named user identity, does that person know? Have they consented?
- Is the identity scoped to the minimum access required, or was it provisioned broadly for convenience?

- Who approved the identity configuration, and is that decision documented?

2. Authorisation: What is this agent permitted to do?

Permissions granted at setup tend to expand over time and rarely shrink. Agentic AI inherits this problem and amplifies it because the agent acts continuously, not occasionally.

- What actions can the agent take without human approval?
- Are those actions bounded by policy, or bounded only by what the underlying identity technically allows?
- Has anyone mapped the difference between what the agent is configured to do and what it could do with the permissions it holds?

3. Ownership: Who is accountable when something goes wrong?

Agents do not have accountability. People do. The question is which person owns the agent's behaviour in your organisation.

- Is there a named owner for each deployed agent?
- Does that owner have enough visibility to actually exercise accountability, or are they accountable in name only?
- What happens to agent ownership when the original owner leaves the organisation?

4. Audit: Can you reconstruct what the agent did?

An agent that cannot be audited is an agent that cannot be governed. This sounds obvious. It is frequently overlooked.

- Are agent actions logged at a sufficient level of detail to reconstruct a sequence of events?
- Are those logs stored separately from the agent's own operational environment?
- Who has access to the logs, and is that access itself governed?

5. Review: When does the identity and permission set get reassessed?

Most identity governance frameworks include access reviews for human accounts. Agents are rarely included.

- Is there a scheduled review cycle for agent identities and permissions?
- What triggers an out-of-cycle review? A change in the agent's scope? A security incident? An organisational restructure?

- Who conducts the review, and what authority do they have to modify or revoke agent access?

6. Termination: Can you switch this agent off cleanly?

Agents that are difficult to terminate are agents that will continue operating beyond their intended scope. Termination is not just a technical concern. It is an accountability concern.

- Is there a defined offboarding process for agents, equivalent to the process for human employees?
- Does agent termination include revocation of identity, permissions, and any stored credentials or tokens?
- Has that process been tested?

PART TWO

The Agent Identity Accountability Framework

The six questions above are diagnostic. This framework gives you a structure to move from diagnosis to governance. It is organised around four accountability layers.

Layer 1 — Identity Governance

Every agent must have a registered identity that is subject to the same governance controls as any other identity in your environment. This means:

1. Each agent is listed in your identity register with a clear designation that distinguishes it from human accounts.
2. The identity has a defined minimum permission scope, documented at the point of provisioning.
3. The provisioning decision is approved by someone with the authority to grant that level of access, not just the team deploying the agent.
4. The identity is included in your standard access review cycles.

> ***The failure mode here is not malice. It is convenience. Service accounts provisioned with broad access because narrowing them down takes time. That time debt becomes accountability debt.***

Layer 2 — Action Boundaries

Identity governance controls what the agent can access. Action boundaries control what the agent can do. These are not the same thing, and the gap between them is where risk accumulates.

5. Define which actions the agent can take autonomously and which require human approval before execution.
6. Document the decision logic: not just what the agent does, but under what conditions it decides to act.
7. Establish a change control process for expanding action boundaries. Any expansion should be treated as a new deployment decision, not a configuration tweak.
8. Test boundary conditions. What does the agent do when it encounters an ambiguous situation? Does it pause and escalate, or does it proceed with the broadest interpretation of its permissions?

Layer 3 — Human Accountability Assignment

Governance frameworks work when there is a person who owns the outcome. For each deployed agent, that ownership must be explicit, visible, and meaningful.

9. Assign a named owner with operational accountability for each agent. This person is not the vendor contact or the IT administrator. They are the person who answers for the agent's behaviour in a governance review.
10. Give the owner the tools and access to actually exercise accountability. An owner who cannot see what the agent is doing cannot own it in any meaningful sense.
11. Build agent ownership into your offboarding process for human employees. When a named owner leaves, ownership must transfer before the departure, not after.
12. Consider a secondary owner or deputy, particularly for agents that operate in high-risk or regulated environments.

Layer 4 — Continuous Monitoring and Review

Deployment is not the end of the governance process. It is the beginning of the operational governance process.

13. Establish a monitoring regime that surfaces anomalous agent behaviour. This does not require sophisticated tooling at the outset. It requires someone looking at the logs.
14. Define what anomalous looks like for each agent. Volume of actions outside normal range. Actions taken at unusual times. Access to resources outside the agent's typical scope.
15. Schedule formal reviews at a cadence appropriate to the agent's risk profile. High-risk agents warrant quarterly review. Lower-risk agents may be reviewed annually. The cadence should be documented and followed.

16. Create an incident response process specific to agent behaviour. If an agent takes an action that causes harm or violates policy, what is the first step? Who is notified? What is the containment approach?

CLOSING

The Accountability Gap Is Not a Technology Problem

Every control described in this guide is available to enterprise teams today. Identity governance tools exist. Audit logging is standard infrastructure. Access review processes are mature. The gap is not technical capability. It is organisational will.

Agentic AI is being deployed into environments where the accountability frameworks have not kept pace with the capability. That is not a criticism of the technology. It is an observation about how organisations adopt technology.

The organisations that will deploy agentic AI responsibly are not the ones with the most sophisticated AI tools. They are the ones that treated the governance question with the same seriousness as the capability question, before the first agent went live.

> ***If your organisation cannot answer who is this agent acting as, and who approved that, you have a shadow IT problem with better marketing.***

The questions and framework in this guide are a starting point. They are not exhaustive and they are not a substitute for legal, security, or compliance advice specific to your context. They are the minimum a responsible enterprise team should be able to answer.

Start there.

About

Karim writes about AI governance, enterprise AI strategy, and operating model accountability at karimsaiplaybook.com. His focus is the gap between what AI can do and what organisations are actually ready to govern.

karimsaiplaybook.com