



STRATOS CYBER



# AI Agent Security Readiness Checklist

A practical SMB guide to defending against AI-enabled attacks  
and controlling the AI operating inside your business.

## THE CORE IDEA

**Treat autonomous AI as a privileged digital actor.**

If an AI system can use credentials, call APIs, execute commands, communicate externally, or modify business systems, govern the authority it has been given.

## CONTEXT

## Why this checklist exists

AI changes the direction of cyber risk. Attackers can use AI to accelerate reconnaissance, phishing, vulnerability analysis, exploitation, and post-compromise activity. At the same time, an AI agent operating inside an SMB may hold corporate credentials, reach cloud services, execute tools, or communicate with external systems. That creates a second exposure: the business can become the origin of actions it never intended to perform.

This checklist gives business and technical leaders a quick way to examine both sides of that problem before an agent receives meaningful autonomy. It is built around Stratos Cyber's 3P model: **Purpose, Preparedness, and Practicality**.

## RISK DIRECTION

## The two-sided threat model

## EXTERNAL

**AI outside → attacking in**

AI-enabled attackers can reduce the time and effort needed to probe, deceive, exploit, and process information across many targets.

## INTERNAL

**AI inside → acting out**

An internal agent with tools, credentials, and network access can reach beyond its intended role if it is manipulated, over-permissioned, or poorly constrained.

## PRE-DEPLOYMENT

## Before you deploy an agent, know three things

**Capability**

What can the agent do through the tools available to it?

**Access**

Which identities, systems, data, APIs, and external resources can it reach?

**Autonomy**

How far can it proceed before a human must approve the next action?

## THE CHECKLIST

# The Stratos Cyber 3P quick check

Mark each statement only when you can support the answer with a defined process, technical control, or accountable owner. **"We assume so" should be treated as an open item.**

## P1

### PURPOSE

Define why the agent exists and keep its authority tied to that purpose.

- |                                                                                                |                                                                                                                          |
|------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> We can state the business purpose of the agent in one clear sentence. | <input type="checkbox"/> An accountable business owner and technical owner are identified.                               |
| <input type="checkbox"/> The agent has an identifiable service or workload identity.           | <input type="checkbox"/> Its permissions are limited to the data, systems, and tools required for its task.              |
| <input type="checkbox"/> Activities the agent must not perform are explicitly defined.         | <input type="checkbox"/> We know which external systems, APIs, customers, suppliers, or Internet resources it can reach. |

## P2

### PREPAREDNESS

Assume unexpected behaviour will eventually occur and make sure control can be regained.

- |                                                                                                          |                                                                                                                      |
|----------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Credentials used by the agent can be revoked quickly.                           | <input type="checkbox"/> The agent can be stopped or isolated without redesigning the environment.                   |
| <input type="checkbox"/> API access and outbound connectivity can be restricted when required.           | <input type="checkbox"/> Important actions can be reconstructed from logs or other evidence.                         |
| <input type="checkbox"/> AI-agent incidents are included in incident response and escalation procedures. | <input type="checkbox"/> We can identify the systems and third parties the agent interacted with during an incident. |

## P3

### PRACTICALITY

Match the safeguards to the authority and real-world impact of the agent.

- |                                                                                                    |                                                                                                            |
|----------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/> Controls reflect what the agent can actually access, change, and trigger. | <input type="checkbox"/> High-impact actions require stronger authorization than low-risk tasks.           |
| <input type="checkbox"/> Privileged agents are not governed like ordinary chat assistants.         | <input type="checkbox"/> New tools, credentials, external access, or autonomy trigger a reassessment.      |
| <input type="checkbox"/> The controls are realistic for our people, budget, and operating model.   | <input type="checkbox"/> The business understands the likely impact if the agent makes the wrong decision. |

**LEADERSHIP SIGN-OFF**

## Five gating questions

Before granting meaningful autonomy, leadership should be able to answer each question clearly. An unclear answer is a reason to pause the deployment and close the gap.

**1 Why does this agent exist?**

Its purpose, owner, and operating boundaries should be explicit.

**2 What authority have we given it?**

Know its identities, credentials, tools, permissions, and ability to change systems.

**3 What can it reach outside the organization?**

Understand external APIs, SaaS platforms, cloud services, customer environments, suppliers, and Internet access.

**4 What can it do without human approval?**

Define the exact boundary between autonomous action and human authorization.

**5 Can we stop it and reconstruct what it did?**

Containment and traceability should exist before the agent is trusted with material authority.

**RED FLAGS**

## Signals that deserve immediate attention

- An agent uses a shared human account or inherits broad administrator privileges.
- Outbound Internet access is unrestricted even though the agent needs only a small set of destinations.
- Tool calls, commands, or material actions cannot be traced after the fact.
- There is no reliable way to revoke credentials, terminate the agent, or isolate it quickly.
- Nobody can explain which actions require human approval or who owns the consequences of the agents' decisions.

NEXT STEP

# A quick check is only the start

If several answers are unclear, your organization may have delegated AI authority without establishing equivalent control.

Stratos Cyber can help you:

## Inventory AI agents

Know every agent operating in your environment.

## Map capability–access–autonomy

See what each agent can do, reach, and decide.

## Review identities & external reach

Examine privileged identities and outbound exposure.

## Build a 3P improvement plan

Translate findings into practical, prioritized action.

**Protect your organization from malicious AI and never give the AI operating inside it more authority than the business is prepared to control.**



**STRATOS CYBER**

AI Security · Cyber Governance · Practical Security for SMBs

AI Agent Security Readiness Checklist · Client Visibility Edition · v1.0 — Screening and awareness aid, not a certification or legal opinion.