SOC Enablement Pack: AI Foundry and Agent 365

Telemetry, ingestion, SIEM use cases, the kill switch, and the threat model

For: Head of SOC and the detection engineering team **Platform:** Microsoft Foundry, Agent 365, Microsoft Sentinel **Verified:** November 2026 **Version:** 1.0

1. The ask on the SOC

The organisation is enabling AI agents on Microsoft Foundry and Agent 365. An agent is a new monitored entity: a non-human, autonomous actor with its own identity that reads data and calls tools. The SOC is being asked to see these agents, detect misuse, and be able to contain one. This pack sets out what you need to ingest, the use cases to build, where the kill switch lives, and the threats to detect against.

Concretely, the expectation on the SOC for AI Agent 365 is five things:

1. **Ingest the telemetry** from the Foundry platform, the agent identities, and the supporting Azure resources into the Sentinel workspace (section 2 and 3).
2. **Maintain an agent inventory view**, know which agents exist, who owns them, and what they can reach.
3. **Build and run the detections** in the use-case catalogue (section 5), tuned to agent behaviour, not just infrastructure.
4. **Operate the kill switch**, run the containment playbooks and the manual runbook when a detection fires (section 6).
5. **Report coverage**, show which agents are monitored, which use cases are live, and where the gaps are.

> **THE ONE-LINE ASK**
>
> Treat each agent as a monitored, containable identity. The SOC's job is to make every production agent visible in Sentinel, detectable against the use cases here, and stoppable through a known set of controls.

2. What the SOC must enable, their side

Nothing arrives by default. Diagnostic logging is off until switched on, per resource. The setup is a landing workspace plus routing from each source into it.

Step	What to do
Landing workspace	Use one Sentinel-enabled Log Analytics workspace as the agent-security system of record. Route all agent-relevant logs to it so correlation works in one place.
Foundry diagnostic settings	On every Foundry / Cognitive Services account, add a diagnostic setting sending **Audit, RequestResponse, Trace, AzureOpenAIRequestUsage, ManagedNetworkEvent** and AllMetrics to the workspace. One setting per account; multiple accounts each need their own.
Agent telemetry (OpenTelemetry)	Connect the project's Application Insights (which carries the OpenTelemetry agent traces) to the same workspace, so agent reasoning, model and tool-call spans land in Sentinel. This is the behavioural layer and it is GA.
Entra diagnostic settings	Route **AuditLogs, SignInLogs, ServicePrincipal and agent sign-ins, and the agent risk logs** to the same workspace. Without AuditLogs the agent identity trail is invisible to Sentinel.
Supporting resources	Diagnostic settings on the data and infrastructure the agents use: Azure AI Search, Storage (Blob), Cosmos DB or SQL, Key Vault, NSG flow logs, Azure Firewall, and the subscription Activity Log.
Preview connectors (optional)	Install the Agent 365 connector, the Agent Identities connector, the Microsoft Copilot connector, and onboard Defender XDR, for richer agent inventory and detection. Additive; not required for the core use cases.
Workspace hygiene	Decide retention and tier per table (Analytics tier for behavioural tables), set RBAC so the SOC can read agent tables, and confirm ingestion within about 15 minutes of enabling each source.

3. What telemetry we can get, the inventory

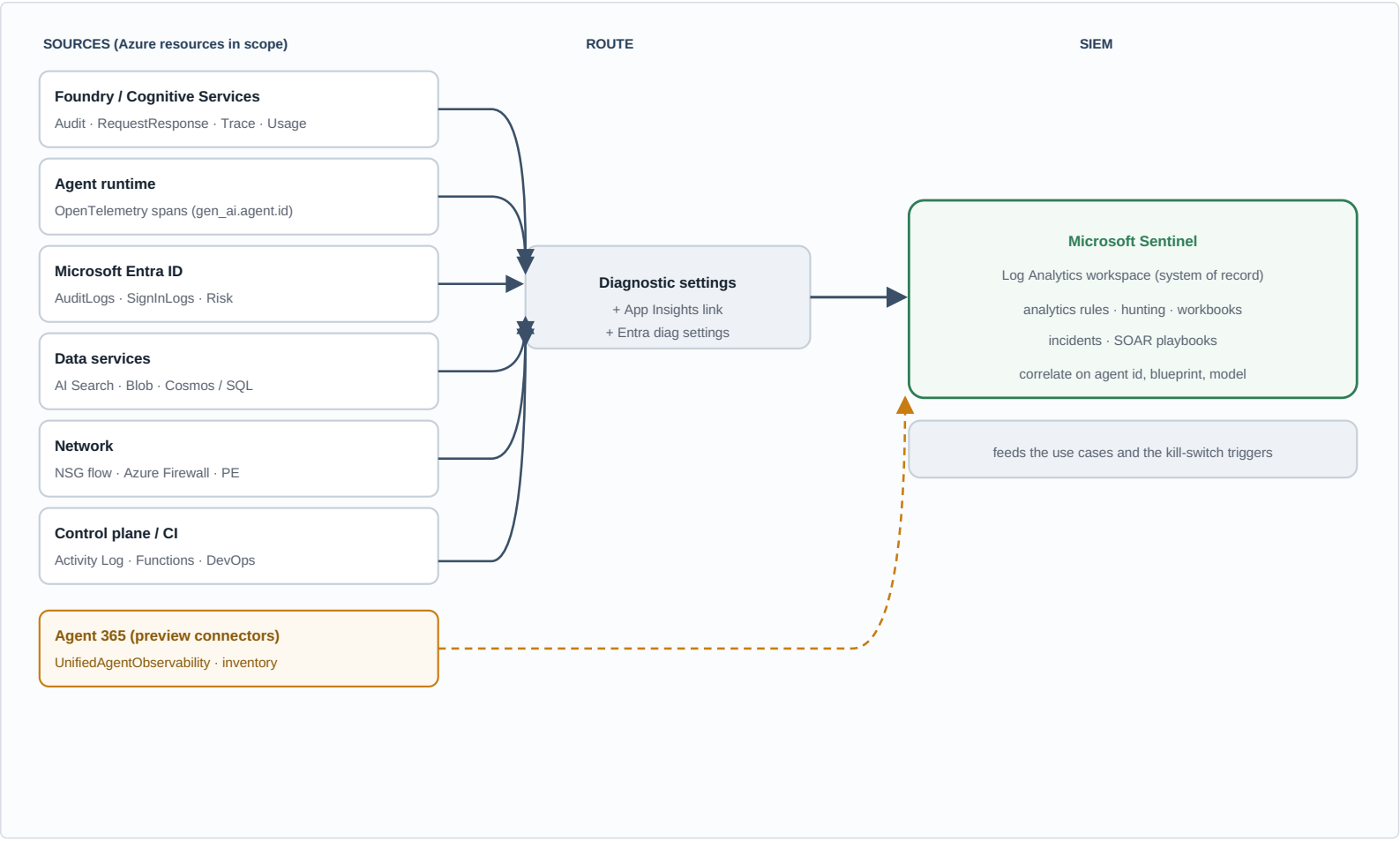
This is the full picture of what is available to ingest today, the table it lands in, and what it tells you.

Source	Table / category	What it gives the SOC	Status
Foundry inference	`AzureDiagnostics` (RequestResponse)	Every model request and response, status codes, latency. The record of what the agent asked the model and got back.	GA
Foundry admin	`AzureDiagnostics` (Audit)	Key access (ListKeys), deployments, configuration and access-control changes. The platform-tamper trail.	GA
Foundry trace / usage	`AzureDiagnostics` (Trace, AzureOpenAIRequestUsage, ManagedNetworkEvent)	Inference trace detail, token usage (spend signal), managed-network events.	GA
Agent behaviour (OTel)	`AppTraces`, `AppDependencies`, `AppRequests`	Agent reasoning, tool invocations, multi-agent calls, errors, with `gen_ai.agent.id`. The behavioural heart of agent detection.	GA
Agent identity audit	`AuditLogs`	Agent identity and blueprint create, disable, ownership and role changes. Who changed what about an agent.	GA
Agent sign-ins	`SignInLogs`, service principal / agent sign-in tables	When and from where an agent authenticated, and to what. The identity activity trail.	GA
Agent risk	`RiskyAgents`, `AgentRiskEvents`	Entra ID Protection risk detections for agents (unfamiliar access, sign-in spikes, failed access).	GA
Data services	`AzureDiagnostics`, `StorageBlobLogs`, Search / Cosmos / SQL logs	Data the agent reached: search queries, blob access, database reads. Data-access and leak signals.	GA
Network	NSG flow logs, Azure Firewall (`AZFWApplicationRule`, `AZFWNetworkRule`)	Agent-subnet egress, allowed and denied destinations, private-endpoint traffic. Egress and exfiltration signals.	GA
Control plane	`AzureActivity`	Who deployed, changed or deleted Foundry, agent or network resources. Change and tamper context.	GA
Agent 365 activity	`UnifiedAgentObservability`	Unified agent activity and behaviour via the Agent 365 connector.	Preview
Agent inventory	`EntraAgentIdentities`, blueprint and agent-user tables	Agent and blueprint inventory, ownership and governance context via the Agent Identities connector.	Preview

Copilot activity	`CopilotActivity`	Copilot agent activity via the Microsoft Copilot connector.	Preview
Agent threat detection	Defender XDR tables	Defender detections on agents via the Agent 365 integration.	Preview

CORRELATION KEY

Tie it together on the agent identity (`gen_ai.agent.id` and the Entra agent object id), plus blueprint, model and deployment, session, tool call, and the user or workload the agent acted for. If these do not join across tables, the use cases below weaken.



Log flow: every Azure resource in scope routes into one Sentinel-enabled workspace through diagnostic settings, the App Insights link and Entra diagnostic settings. The preview connectors add agent inventory and unified activity on top.

4. SIEM use cases

What to detect, grouped by theme. Each use case names the primary data source, what it catches, the STRIDE category, and the response. Response levels: `Alert` SOC triage, `Contain` playbook with approval, `Kill` auto-contain for high-tier agents.

A. Agent identity and access

#	Use case	Data source	Detects	STRIDE	Response
A1	New or unknown agent identity created	AuditLogs, inventory	Shadow agents, agents created outside the sanctioned path	Elevation	Alert
A2	Agent identity with no owner	AuditLogs, inventory	Orphaned agents that cannot be governed or stopped	Repudiation	Alert
A3	Agent sign-in from unfamiliar location or at anomalous volume	SignInLogs, RiskyAgents	Compromised or misused agent credential	Spoofing	Contain
A4	Agent granted broad or privileged roles	AuditLogs, AzureActivity	Privilege creep, excessive standing access	Elevation	Alert
A5	Agent accessing a resource outside its remit	SignInLogs, data logs	Confused deputy, lateral reach	Elevation	Contain
A6	Repeated failed access by an agent	SignInLogs, RiskyAgents	Probing, misconfiguration, or an agent being steered	Spoofing	Alert

B. Agent behaviour and misuse

#	Use case	Data source	Detects	STRIDE	Response
B1	Tool call outside the agent's allowed set	AppTraces / AppDependencies	Indirect prompt injection steering the agent to new tools	Tampering	Contain
B2	Write or high-impact action without approval	AppTraces, AzureActivity	Agent taking actions it should have gated	Elevation	Kill
B3	Prompt-injection indicators in agent input or behaviour	AppTraces, RequestResponse	Planted instructions in content the agent read	Tampering	Contain
B4	System-prompt or instruction extraction attempt	RequestResponse, AppTraces	Attempts to reveal the agent's instructions	Info disclosure	Alert
B5	Runaway loop or abnormal action volume	AppTraces, AzureOpenAIRequestUsage	Loops, denial of wallet, denial of service	Denial of service	Kill
B6	Token or spend spike for an agent	AzureOpenAIRequestUsage	Cost abuse, runaway or hijacked agent	Denial of service	Contain
B7	Intent drift toward a destructive or out-of-policy action	AppTraces (sequence)	The agent heading toward something it should not do	Tampering	Kill

C. Data and egress

#	Use case	Data source	Detects	STRIDE	Response
C1	Agent egress to a new or unapproved destination	Azure Firewall, NSG flow	Exfiltration, command and control	Info disclosure	Contain
C2	Sensitive data in agent output or retrieval	RequestResponse, DLP / Purview	Data leak through the agent	Info disclosure	Contain
C3	Agent reaching another user's or tenant's data	Search / data logs	Cross-user or cross-tenant access through retrieval	Info disclosure	Kill
C4	Large or unusual data pull by an agent	StorageBlobLogs, data logs	Bulk collection, staging for exfiltration	Info disclosure	Contain
C5	Egress to an allowed destination used as a carrier	Azure Firewall, AppDependencies	Exfiltration through a permitted endpoint	Info disclosure	Alert

D. Model and platform integrity

#	Use case	Data source	Detects	STRIDE	Response
D1	Guardrail or content-safety configuration changed or disabled	Audit, AzureActivity	Guardrail tampering from inside	Tampering	Contain
D2	Unapproved model deployed or data-retention feature enabled	Audit, AzureActivity	Shadow model, training-on-data turned on	Tampering	Alert
D3	Key listed or created on a Foundry resource	Audit (ListKeys)	Credential harvesting, standing-key creation	Spoofing	Contain
D4	Public network access enabled on a Foundry or data resource	AzureActivity, ManagedNetworkEvent	Exposure of a resource that should be private	Elevation	Contain
D5	Pipeline identity changing agent configuration at deploy time	AzureActivity, DevOps audit	Pipeline identity abuse	Elevation	Alert

E. Detection coverage and health

#	Use case	Data source	Detects	STRIDE	Response
E1	Agent telemetry stopped flowing	Ingestion health, heartbeat	Blinding of the SOC, diagnostic setting removed	Repudiation	Alert
E2	Diagnostic setting disabled or deleted	AzureActivity	Deliberate removal of logging	Tampering	Contain
E3	A production agent with no monitoring coverage	Inventory vs telemetry	Unmonitored agents in production	Repudiation	Alert

5. The kill switch, the logical buttons

"The kill switch" is not one button. It is a set of logical controls across five planes. A response may press one or several. The SOC runbook and the SOAR playbooks orchestrate them.

Identity plane	Runtime plane	Network plane	Data plane	Platform plane
Disable agent identity new tokens blocked	**Delete agent, force** ends active sessions	**Cut egress (firewall)** deny agent subnet	**Rotate / revoke keys** kill standing creds	**Disable project** whole project off
Disable blueprint all of that type	**Delete agent version** one version	**NSG deny rule** isolate the subnet	**Revoke data connection** unlink source	**Disable guardrail bypass** re-assert safety
Disable agent user separate account	**Delete conversation** one live session	**Remove private endpoint** cut data path	**Purview DLP block** stop the data leaving	**Disable public access** re-privatise
Revoke sessions close the token gap	**Pause deployment** stop the model serving			
CA block (tenant) all agent auth				
Remove role assign. block new agents				

Red buttons = the core kill for one agent: disable identity + delete runtime + cut egress.

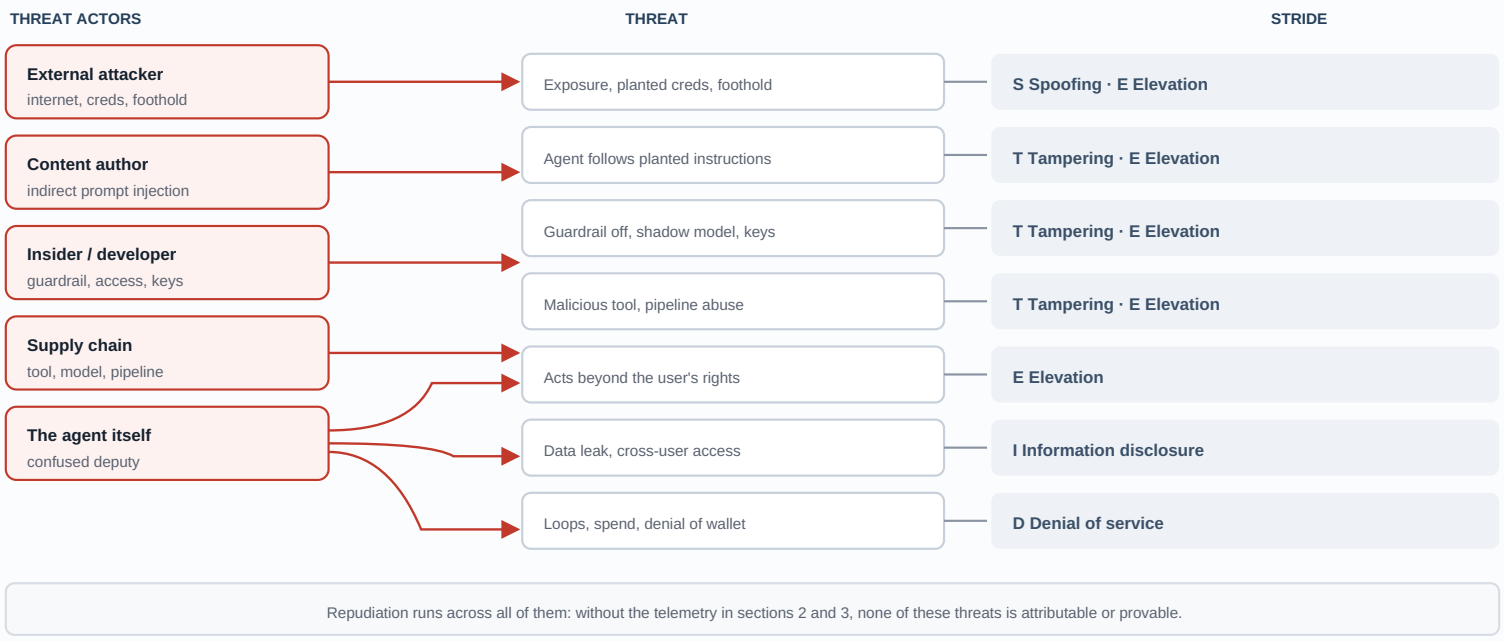
The rest are graded or situational. A SOAR playbook presses the right combination for the trigger and the agent's tier.

The kill switch is a panel of logical buttons across five planes. The three red buttons are the core single-agent kill; the others scale the response up or sideways.

6. Threat actors and STRIDE

Who attacks an agent, what they do, and how it maps to STRIDE. Use this to sanity-check that the use cases cover the threats.

Threat actor	How they act against the agent	STRIDE emphasis
External attacker	Scans for exposed resources, uses planted or leaked credentials, seeks a network foothold to reach private endpoints.	Spoofing, Elevation
Content author (indirect injection)	Never signs in. Plants instructions in a web page, document or tool response the agent later reads, turning the agent into their proxy.	Tampering, Elevation
Malicious or careless insider / developer	Disables a guardrail, widens access, enables public access or keys, deploys an unapproved model, or creates shadow agents.	Tampering, Elevation
Compromised supply chain	A malicious or vulnerable tool, model or pipeline identity used to reach internal systems or alter the agent at deploy time.	Tampering, Elevation
The agent itself (confused deputy)	Not hostile, but manipulated into acting with its own privileged identity, doing what the user never could.	Elevation, Info disclosure



Threat actors to the threats they create, mapped to STRIDE. Repudiation sits underneath all of them: it is the case for the logging itself.

7. SOC readiness checklist

- One Sentinel-enabled workspace is the agent-security system of record, and all sources route to it.
- Foundry, Entra, data, network and control-plane diagnostic settings are enabled and confirmed flowing.
- Agent OpenTelemetry traces (App Insights) land in the workspace and carry the agent id.
- The use cases in section 4 are built as analytics rules, with response levels set per agent tier.
- The kill-switch buttons in section 5 are wired into playbooks, and the manual runbook is held and rehearsed.
- An agent inventory view exists, and every production agent maps to telemetry and an owner.
- Ingestion-health detections (E1 to E3) are live, so the SOC is told if it goes blind.
- Coverage is reported: which agents, which use cases, which gaps.

SOC enablement pack, version 1.0, November 2026. Table names, log categories and preview status verified against Microsoft documentation at the date shown; preview features and schemas change, reconfirm before build. Use cases are a starting catalogue to implement and tune, not finished analytics rules.