DEEP CYBER

AI Readiness Assessment Playbook

Find. Secure Critical. Understand. Improve

A short, four-step method for getting control of the AI your organisation runs.

THE METHOD, AT A GLANCE

1	2	3	4
INVENTORISE List every AI & agent	**SECURE URGENT** Fix the worst, now	**EVALUATE** Score your posture	**GOVERN & PLAN** Own and set a plan

YOU CANNOT GOVERN WHAT YOU CANNOT SEE

Most organisations have no real list of the AI they run - only a procurement spreadsheets. This playbook helps you **find it**, **secure the riskiest**, and **build a plan you can take to stakeholders**— starting this week.

Anchored in: ETSI EN 304 223 · UK AI Cyber Security Code of Practice · OWASP Top 10 for Agentic Applications · OWASP State of Agentic Security Governance - Adoption Tiers AT0–AT8

WHY NOW · THE EVIDENCE

The risk is already here

Shadow AI **already happening**. By 2026, nearly **40% of AI interactions involve sensitive corporate data** (Cyberhaven, 2026). One landmark incident and four independent datasets — measuring different things — all point the same way.

TLDR;

In 2023, Samsung engineers pasted source code and internal meeting notes into ChatGPT. By 2026, that behaviour had become normalised. Verizon now reports shadow AI as one of the leading causes of non-malicious data loss. Nearly two-thirds of activity on personal AI accounts is work-related, over a third of data entered into AI tools is sensitive corporate information, and almost half of employees admit using unapproved AI tools. Most strikingly, 60% say the security risk is worth it if it helps them get the job done faster. ***Shadow AI is not a future threat. It is what happens when productivity pressure outruns governance. Agentic AI is set to make it more prevalent and riskier.***

2023

The first major shadow-AI leak — Samsung engineers pasted source code and meeting notes into ChatGPT. Not an attack; just people trying to work faster.

Forbes, 2023

64.5%

of activity on personal & free AI accounts is work, not personal — invisible to company logging. Context stays in personal history after staff leave.

Harmonic Security, May 2026

3rd

most common cause of non-malicious data leakage is now shadow AI. 45% of staff use AI (up from 15%); 67% reach it via non-corporate accounts.

Verizon DBIR 2026

34.8%

of corporate data pasted into AI tools is sensitive — up from 27.4% a year earlier and just 10.7% the year before.

Cyberhaven

49% / 60%

of workers use AI in ways their employer hasn't approved — and 60% would accept the security risk if it helped them work faster (63% when no approved tool exists).

BlackFog, January 2026

WHY AGENTIC MAKES IT WORSE

Shadow AI used to mean a person pasting text into a chatbot. Agentic AI takes the person **out of the loop**: an agent reaches for tools, APIs and data at runtime through MCP (the Model Context Protocol) and decides what to trust on the fly. There are already **7,000+ publicly accessible MCP servers** and 150M+ SDK downloads — with **30+ CVEs** found in MCP servers and clients in a single 60-day window, and **9 of 11 MCP marketplaces** successfully poisoned in a red-team test. Every server an agent connects to widens its blast radius — a poisoned tool or cloned agent inherits the agent's identity, credentials and reach. Unmanaged agentic AI isn't one more shadow tool; it is **shadow AI that can act**.

Source: OX Security & Trend Micro, April 2026.

WHY NOW · REGULATION

The rules are catching up

Regulators are **already** asking the same question: what AI do you run, and is it safe? More are about to make it binding. AI has entered your organisation faster than your governance has — staff use it, teams build with it, vendors embed it — and almost none of it is on a list. That gap is now a compliance problem, not just a security one.

NCSC Cyber Assessment Framework *live today*	**Cyber Security & Resilience Bill** *tightening*	**EU AI Act inventory duties** *reaches you by contract*	**ETSI Conformity Assessment (TS 104 216)** *coming*
▪ The framework regulators and assurance reviews already measure you against. ▪ A.3 (assets) and A.4 (supply chain) cover what supports important business and government services — AI is in scope by definition. ▪ Auditors are already flagging unmanaged AI as a gap. ▪ **"We have no AI inventory" is a finding today**, not a future risk.	▪ Moving through Parliament now; duties land on a timeline you don't set. ▪ Widens scope to MSPs and data centres — more organisations caught than before. ▪ Fines up to **£17m or 4%** of turnover; 24-hour incident early-warning. ▪ Asset and supply-chain duties assume you can already list your systems, AI included.	▪ Bites if you operate in, or supply into, the EU — even though the UK hasn't adopted the Act. ▪ Pulls through to UK organisations by contract, via procurement and supply-chain clauses. ▪ Inventory and transparency duties stand; expect "list your AI" in tenders. ▪ Moratorium rejected, but high-risk deadlines deferred (Dec 2027 / Aug 2028) — **less urgency, same obligation.**	▪ Conformity assessment for EN 304 223 — **expected summer 2026.** ▪ Turns the standard into something you can be audited and certified against. ▪ Defines how to evaluate and validate AI against the 13 principles. ▪ **Score your posture now**, so you're assessment-ready — not scrambling.

What "ready" looks like

- A single AI register — sanctioned, platform-built, and shadow.
- Every high-risk system has an owner and a control.
- A posture score against ETSI EN 304 223 / the UK Code of Practice.
- A named accountable person and plan in sprints.
- Evidence you can hand to an auditor, stakeholders, or governance board.

HOW TO READ THIS PLAYBOOK

Four steps, in order: **Inventorise → Secure Critical → Evaluate → Govern & Plan**. Each step is short and practical. The detail — the shadow-AI scan, the sector examples, the standards mapping — sits in the appendices so the method stays on one page.

One model, three questions

Every AI system gets read three ways. Answer the three questions and the controls fall out.

THREATS — what can go wrong
OWASP Top 10 for Agentic Applications (ASI01–ASI10)

RISK TIER — how much it matters
Baseline set by OWASP Adoption Tier (AT0–AT8), raised by risk amplifiers

LIFECYCLE — where and how to mitigate it.
UK Code of Practice (P1–P13) · ETSI EN 304 223 phases, principles, and provisions

- **Threats** name the failure modes — goal hijack, tool misuse, identity abuse, supply-chain, memory poisoning, rogue agents. *An agent is still an LLM application: the OWASP Top 10 for LLM Applications and the API Security Top 10 still apply beneath it. See Step 3.*
- **Risk Tier** says how hard to apply it. A log-summary helper and an autonomous payments agent do not need the same rigour.
- **Lifecycle** says where and how to mitigate it — design, build, deploy, run, retire.

THE GOVERNING PRINCIPLE — LEAST AGENCY

Give an agent the smallest autonomy, fewest tools, and narrowest access that still does the job. Everything else is detail.

ADOPTION TIERS — HOW EXPOSURE CLIMBS

AT0	AT1	AT2	AT3	AT4	AT5	AT6	AT7	AT8
Shadow AI	Vendor assistant	Platform-integrated	Citizen-built	Code-executing	Custom in-house	Externally extended	Multi-agent	Federated
Inherent risk tier:								
Unknown	Low	Low	Elevated	Elevated	Elevated	High	High	Critical

Why it climbs → *More autonomy · more tools · bigger blast radius*

- **AT0 —** invisible. No oversight; unmanaged data exposure. **Treat as high until discovered.**
- **AT1–AT2 —** vendor-bounded. Low autonomy, no code execution, contained reach.
- **AT3–AT5 —** acts on your data; executes code; you own the boundaries. Autonomy and attack surface rising.
- **AT6–AT7 —** crosses trust boundaries — external tools (MCP), agent-to-agent. Reach and blast radius expand sharply.
- **AT8 —** spans organisations. Federated trust, widest blast radius, hardest to contain.

FIVE RISK AMPLIFIERS THAT RAISE A SYSTEM ABOVE ITS BASELINE TIER

The adoption tier sets the baseline; any one of these can raise a system's risk above its tier. A "low" tool that can take an irreversible action is not low. (*See Appendix A for more details*)

Exposure — how reachable it is (publicly available vs private), and whether it reads untrusted outside input.
Data sensitivity — how sensitive the data it can reach is.
Autonomy — how far it acts without a person approving.
Integration privilege — how much it is connected to, and with what identity and rights.
Physical-world effect — whether its actions reach beyond the screen.

Step 1 · Inventorise your assets

You cannot secure, score, or govern what you have not listed. Build one register capturing the same basic facts for every AI system and agent. If a fact surprises you, that is a finding that needs further investigation.

Capture for each system	Four places AI appears
▪ **Name & owner** — who is accountable. ▪ **Risk and adoption tier (AT0–AT8)** with risk amplifiers. ▪ **Services, Tools & MCP servers** it can call. ▪ **Credentials** it holds and their scope. ▪ **Data** it can read or write. ▪ **Irreversible actions** it can take alone. ▪ **Human-in-the-loop?** Where, and is it real.	▪ **Sanctioned** — what procurement bought. *Listed is not assured: a sanctioned tool never reviewed or assured is a Step 2 finding* ▪ **Platform built (Embedded & Citizen Developer)** — Copilot, Custom GPTs, Amazon Q, Copilot Studio, Power Automate, Agentforce. ▪ **Custom-built** — agents in your own code & CI/CD. Jira and similar systems, GitHub repositories. ▪ **Shadow** — personal accounts, OAuth grants, browser/IDE extensions. Hardest, and where the risk lives.

The Shadow AI iceberg

Most AI is below the iceberg water line.

Above it: sanctioned and platform-built tools your IT team can see. Below it: personal ChatGPT / Claude / Gemini use on work data, OAuth grants to AI apps, browser and IDE extensions, citizen-developer flows and local LLMs. A six-part scan brings it to the surface:

- 1 · Procurement & finance pull
- 2 · OAuth grant audit — highest signal
- 3 · Platform admin reports
- 4 · Code repository scan
- 5 · Network egress review
- 6 · Engineering survey

Full method in Appendix B. The OAuth audit and egress review give the most signal for the least effort.

WATCH THIS

The **OAuth grant audit** is the one most likely to surface something awkward — AI apps connected to mailboxes and drives that nobody approved. It needs admin consent and a defined scope, agreed in writing before the audit begins.

Be realistic about scope & time

A full inventory is not a one-week job. Do it in three layers, biggest signal first.

SPRINT · Week 1	DISCOVERY · Week 2-4	SUSTAIN · Recurring
Sanctioned + platform list. Gap statement. Scope for the next phase.	Shadow AI closed out. Full, defensible register. Remediation plan.	Regular re-scan. New-tool triage. Audit-ready evidence pack.

What week one realistically delivers	Map the register to what regulators ask
▪ **Yes:** sanctioned list, from procurement and platform admin reports — mostly coordination. ▪ **Yes:** citizen-dev and custom-agent list, with admin access and an engineer. ▪ **Not yet:** true shadow AI on personal accounts and extensions — needs egress, OAuth and DLP review over weeks.	▪ NCSC CAF **A.3** (assets) and **A.4** (supply chain). ▪ Cyber Security & Resilience Bill asset & supplier duties. ▪ EU AI Act inventory obligations (if in scope) ▪ **ETSI EN 304 223 / UK Code of Practice — P5 asset inventory** ▪ Output is an **evidence pack**, not just a report.

THE TWO QUESTIONS THAT MATTER MOST IN THE ENGINEERING SURVEY

1. What AI tools or agents are in your stack today? 2. What would you build next if security weren't slow? The second answer tells you where shadow AI is about to appear, and where a need is going unmet. Both point to Step 2: sanction the right tool, or fix the friction, and you remove the reason people work around you.

By the end of Step 1 you have a register and a gap statement. It shows which systems are high-risk and poorly controlled — the systems Step 2 deals with. If you cannot produce the list at all, that is your headline finding: escalate, and move to Step 3 to put effective control in place.

Step 2 · Secure the worst, now

Before any roadmap, deal with the systems that are high-risk and unprotected today. The register will show some that are both high tier and poorly controlled and ones lacked assurance. Triage with one simple grid.

TRIAGE GRID — RISK TIER × HOW WELL CONTROLLED

	Weakly controlled	Well controlled
High tier	**FIX FIRST** high tier · weak control	**WATCH** high tier · controlled
Low tier	**SCHEDULE** low tier · weak control	**ACCEPT** low tier · controlled

What "high-risk & unsecured" looks like

Agentic

- **Irreversible actions** — payments, deletes, sends — with no human check.
- **Broad credentials** held by an app or agent nobody owns.
- **Sensitive data** flowing to a personal-account AI tool.
- **Internet-exposed agent** with no logging of what it did.
- **MCP server or tool** added with no review.

Custom apps & environments

- **Unassured app** — sanctioned or custom, deployed with no security review and assurance.
- **Under-controlled environment** — a data-science lab or dev account publicly reachable, live to sensitive data, with no controls. As dangerous as any high-tier app; it just takes longer to remediate.

IMMEDIATE — CONTAIN THE DANGER THIS WEEK

Revoke unowned or over-scoped OAuth grants · put a human approval on irreversible actions · block sensitive data to personal-account tools (DLP) · **block or restrict access to exposed apps and environments and turn on enhanced monitoring and detection**. These reduce the danger now, even where the lasting fix runs longer.

Follow-up — the longer fix, in parallel

Once contained, the slower remediation runs **in parallel as prioritised activities** — not deferred to Step 3. The danger is already held by the immediate measures; this is the work that makes the fix last.

Finding	The lasting fix (in parallel)
Unassured app *sanctioned or custom*	**Assure it:** threat model → security review → and, if public-facing or high-risk, a red-team / penetration test. Until assured, treat it as weakly controlled and tier it accordingly.
Under-controlled environment *data-science lab, dev account*	**Bring it to production standard:** an environment live to sensitive data carries production controls — separation, least privilege, logging, access review (CoP P6) — then threat model and pentest. If it cannot meet them, remove the sensitive data.

Unmanaged usage *shadow tools, personal accounts*	**Move it into a managed home:** route model access through a corporate account so you control identity, logging and data — e.g. Amazon Bedrock, Azure AI Foundry, or provider API keys (OpenAI, Anthropic) issued centrally to govern coding agents and co-working tools. Where it cannot be managed, remove it.

THE LEVER — LEAST AGENCY

Cut tools the agent does not need, narrow credentials to the task, and cap autonomy — propose, don't act, until trusted. Reducing the tier is often faster than adding controls.

RULE OF THUMB

For shadow AI (AT0) there is no "secure it in place." The only safe options are **move it into a managed state** or **remove it**. You cannot govern what stays invisible.

WHAT MANAGED MEANS

Access through a corporate account, not a personal one; identity known; credentials issued and scoped centrally; actions logged; data access bounded by policy; a named owner; and a known place in the register at a known tier. The point is not friction — it is a governed path people don't need to work around.

Step 3 · Evaluate functions & posture

Steps 1–2 looked at the AI. Step 3 looks at you — your controls and maturity against ETSI EN 304 223 and the UK Code of Practice. Assess in two passes: the principles you hold **once across the organisation**, and the ones you check **on each system**, where the bar rises with the adoption tier.

CROSS-CUTTING — ASSESS ONCE, ACROSS THE ESTATE

Programme capabilities. One review covers the whole organisation.

- **P1 Awareness & training** — staff know the AI-specific risks; training is current and role-tailored.
- **P3 Threat & risk management** — a process that assesses threats and tracks risks, reviewed as systems and threats change.
- **P4 Human responsibility** — a named accountable owner, with oversight designed in where it is a control.
- **P13 Disposal** — a policy for retiring data and models safely.
-

> **SCALE THE THREAT ASSESSMENT TO THE TIER**
> A full threat model for a low-tier vendor assistant is overkill and won't get done. At low tiers, use a short risk questionnaire; at elevated and high tiers, a full threat model; at critical, continuous threat assessment. Match the effort to the risk, or the analysis gets skipped.

Bring the threats into the evaluation

The OWASP Top 10 for Agentic Applications (ASI01–ASI10) does not replace the Code or the standard — it is the threat catalogue that sits alongside them. The Code says what outcomes you must achieve across the lifecycle; the ASI Top 10 says what an agentic system must defend against. Used together, the ASI Top 10 is the implementation guide for the Code's principles. Which threats dominate is not uniform — it climbs with the adoption tier, so weight your evaluation to the tier each system sits at.

Why the Agentic Top 10 specifically. An agent is still an LLM application, so the underlying risks have not gone away. The Code's Implementation Guide already references the OWASP Top 10 for LLM Applications, the AI Exchange and the API Security Top 10 against its principles, and these still apply beneath any agentic system. We focus on the Agentic Top 10 because it is the layer the Guide does not yet fully cover — what changes when systems act rather than answer. Take the others into account too; this evaluation adds the agentic layer on top of them.

WHICH THREATS TO WEIGH, BY TIER *OWASP State of Agentic AI Security & Governance, Appendix 3 — prioritise, not a flat checklist*

Low · AT1–2	Elevated · AT3–5	High · AT6–7	Critical · AT8
ASI01 goal hijack and ASI06 memory poisoning dominate; ASI09 trust exploitation a persistent human-factor risk. Surface constrained by vendor controls.	Adds ASI02 tool misuse, ASI03 identity & privilege abuse, and — once code executes — ASI05 as the defining risk. Six to eight ASI entries active.	The inflection point. ASI04 supply chain, ASI07 inter-agent comms and ASI08 cascading failures activate; at multi-agent, ASI10 rogue agents becomes structural. Full surface engaged.	All ASI entries active at maximum severity, with systemic risk from cross-organisational trust.

> **Note — hygiene vs dominant threat.** Supply-chain hygiene (P7 — knowing and documenting your components) applies from the start and is assessed on every system. ASI04 supply chain becomes the **dominant** threat only at AT6, when an agent reaches for external tools and MCP servers it did not build and cannot fully verify. The control scales from AT1; the threat peaks at AT6.

Per-system — the bar rises with the tier

Check these on each system. What "good" looks like climbs across the four tier-bands. AT0 is out of scope — shadow AI must be absorbed into a managed tier before it can be assessed.

Principle	Low · AT1–2	Elevated · AT3–5	High · AT6–7	Critical · AT8
P2 Design for security *build-only*	n/a — vendor-designed	Partial — simplified secure-design guidelines (from AT3)	Full secure-design review	+ design assurance across the agent chain
P5 Assets	In the register, owned	+ dependencies & data flows mapped	+ tools / MCP servers inventoried	+ cross-org assets tracked
P6 Infrastructure	Vendor config reviewed; access scoped	Dedicated environments; least privilege	+ network isolation, secrets management	+ federated trust boundaries
P7 Supply chain	Vendor SLA reviewed	Components documented	SBOM; MCP / tool provenance verified	Continuous supply-chain assurance
P8 Document data/models/prompts	Vendor docs retained	System design & data sources documented	+ prompt / config change log	+ cross-org documentation exchange
P9 Test & evaluate	Vendor assurance accepted	Functional + basic security testing	Adversarial testing; red-team	Continuous, cross-org evaluation
P10 Comms — users & affected	Users told it's AI	+ prohibited uses stated	+ affected-entity notification	+ cross-org disclosure
P11 Updates & patches	Vendor patching relied on	Tracked patching of components	+ model-update revalidation	+ coordinated cross-org updates
P12 Monitor behaviour	Basic logging on	Logs analysed for anomalies	Behavioural baselines; alerting	Real-time, cross-org telemetry

Score each system against its tier's bar; the gap between where a system sits and what its tier demands is your finding. The pattern across the estate feeds the maturity ladder and the Step 4 plan.

RED — Absent	AMBER — Partial	GREEN — In place
No control, owner, or evidence. A finding for the governance board and CISO.	Exists for some systems, not consistent, weak evidence.	Consistent, owned, evidenced, and tested.

From scores to maturity level

Maturity is not a count of greens. It is whether your controls are consistent and properly supported, read from the scores in two parts.

- **The cross-cutting principles set your lowest possible level.** If P1, P3, P4 or P13 are red — no training, no threat-modelling process, no named owner — your maturity is L0 or L1 whatever your individual systems score, because there is nothing in place to keep the controls running.
- **Consistency across systems decides whether you can rise higher.** Green on some systems and red on others — often green for sanctioned tools, red for custom ones — is L1 to L2: the policy exists but is

not applied evenly. Consistent green at the tiers you run, owned and evidenced, is L3. Green that is held in place by monitoring and continuous checking, rather than a periodic review, is L4.

The RAG definitions already describe this: amber means a control exists but is not consistent; green means it is consistent, owned, evidenced and tested. Moving a principle from amber to green is the same as moving up a level.

Match your ambition to your maturity

Your governance maturity sets a ceiling on the risk tier you can safely run. Do not run above your level.

Level	Name	Highest tier you should run
L0	Initial	AT1–AT2 (assume AT0 shadow is present; prioritise discovery)
L1	Repeatable	AT1–AT4 (address shadow AI via discovery + acceptable-use policy)
L2	Defined	AT1–AT5 (shadow AI substantially eliminated via policy + DLP/monitoring)
L3	Managed	AT1–AT6 (externally extended agents with full lifecycle controls)
L4	Optimised	AT1–AT8 (telemetry-backed, continuous control of multi-agent & federated)

How to read it

- Find your maturity level from the Step 3 scores.
- That level sets the **highest tier you should run**.
- Running above that highest tier is a governance gap — flag it.
- At every level, **assume shadow AI (AT0) is present** and prioritise discovery.

Two ways to close a gap

- **Raise maturity** — add the missing controls, owners, evidence.
- **Lower the tier** — cut autonomy, tools, or access until it fits.
- For shadow AI the only move is **eliminate or absorb** into a managed tier.

Step 4 · Govern it and set the plan

Turn findings into ownership, rules, and a plan in sprints, the governance board and CISO can track.

Governance essentials	Map findings to the obligations
One named accountable owner for AI security (P4). **Acceptable-use policy** for AI — what's allowed, on what data. **The register has an owner** and is kept current. **Tier gates** — a system can't go above the tier its controls justify. **Monitoring** of agent behaviour (P12) and a way to raise alarms. **A path to add new tools safely** — so people don't go around you.	NCSC CAF A.3 / A.4 objectives. Cyber Security & Resilience Bill asset & supplier duties. ETSI EN 304 223 / UK Code of Practice principles. Any sector rule — FCA, NHS DSPT, SRA, CRA (Appendix C).

THE ACTION PLAN, IN THREE SPRINTS

Sprint 1	Sprint 2	Sprint 3
Register built. Fix-first items remediated. Owner named. AUP drafted.	Posture scored. Shadow AI closed out. Tier gates set. Monitoring switched on.	Evidence pack to governance board /CISO/ auditor. Regular re-scan scheduled. Roadmap agreed.

> **THE DISCIPLINE THAT KEEPS IT WORKING**
> Re-run the scan every quarter, triage new tools as they appear, and keep **least agency** as the default for anything new. Readiness is a rhythm, not a one-off report.

What to do on Monday

No budget or tooling needed to begin — a spreadsheet and a few good conversations will do. Start the inventory:

- Pull twelve months of AI spend.
- Run the OAuth grant audit.
- Get platform admin reports.
- Grep repositories for AI SDKs.
- Send the two-question survey to engineering.

By the end of the week, you will have a starting register and a gap statement — enough to see where the risk sits and to scope the sprints that follow

Appendix A · Risk amplifiers defined

The adoption tier sets a baseline. These five factors describe the context of a particular deployment, and any of them can raise its risk above that baseline. Assess each one when you tier a system.

Amplifier	What it means	Lower risk	Higher risk	The question to ask
Exposure	How reachable the system is, and whether it processes untrusted input.	Internal only; fixed, trusted inputs.	Internet-facing, or reads external content such as email, web pages or documents that an attacker could craft.	*Who or what can send input to this system, and can any of it come from outside our control?*
Data sensitivity	The nature of the data the system can read or write.	Public or low-value information.	Personal, special-category, financial, or commercially confidential data.	*What is the most sensitive data this system can reach, and what would it mean if that data leaked or was altered?*
Autonomy	How far the system acts without a person approving the action.	Proposes; a person decides and acts.	Acts on its own, including actions that are hard to reverse.	*Can this system complete a consequential action without a human approving it first?*
Integration privilege	What the system is connected to, and the rights it holds.	Read-only, narrowly scoped to one task.	Broad credentials, write access, or connections to many systems through tools or MCP servers.	*If this system were misused, how far could it reach using the access it already holds?*
Physical-world effect	Whether the system's actions affect things beyond the screen.	Output is informational; a person decides what to do with it.	Controls equipment, moves money, or triggers actions that affect the physical world or that cannot be undone.	*Can this system cause a real-world consequence directly, without a person in between?*

> A system is only as low-risk as its strongest amplifier allows. A vendor assistant (AT1) that can move money carries more risk than its tier suggests, and should be assessed and governed on that basis.

Appendix B · The six-component Shadow AI scan

The repeatable method behind Step 1. Read-only by default. Pull from systems you already run.

Component	What it catches	Source data
1 · Procurement & finance pull	Sanctioned spend. Line items with AI, Copilot, GPT, agent, Claude, Gemini, LLM, ML or vendor names over 12 months.	Finance system, AP ledger
2 · OAuth grant audit highest signal	Third-party AI apps connected to mail, drive, calendar. Broad scopes nobody approved. The classic shadow-AI pattern.	Google Workspace Admin SDK, Microsoft Graph / Entra
3 · Platform admin reports	Deployed flows, agents, copilots, custom GPTs on low-code & SaaS platforms.	Power Platform, Copilot Studio, Agentforce, Atlassian, Slack, Teams
4 · Code repository scan	AI SDK imports (LangChain, LangGraph, AutoGen, Bedrock/OpenAI Agents), MCP client libraries, embedded API keys.	GitHub, GitLab, Bitbucket, Azure DevOps
5 · Network egress review	Traffic to known AI inference endpoints, correlated to users & devices.	Zscaler, Netskope, firewall, DNS, SIEM
6 · Engineering survey	What's in the stack, and what people would build next if security weren't slow.	Two-question form to teams

Known AI inference endpoints to filter for

`api.openai.com` · `api.anthropic.com` · `generativelanguage.googleapis.com` · `bedrock-runtime.*.amazonaws.com` · `*.cognitiveservices.azure.com`

Add major Chinese providers where relevant to your risk profile.

Scan design principles

- **Read-only** — pull from APIs, ingest logs, produce reports.
- **Federated** — use the IdP, MDM, EDR, SIEM you already have; no new endpoint agents.
- **Evidence-pack first** — outputs map to CAF objectives, not just a dashboard.

> **AUTHORISATION**
>
> Components 2 and 5 require admin consent and a written scope. Pulling OAuth grants or egress logs from a tenant must be covered in the engagement letter / Statement of Work before you start.

Appendix C · Examples

Six illustrative systems, each read through the method: its adoption tier, the principles to focus on, and the sector rule that bites hardest. These are focusing heuristics — where the risk concentrates — not a complete threat mapping. The full AT0–AT8 ladder is in Appendix D but it is for scoping purposes, you will need to use threat modelling to derive a full list of risks and controls.

For the threats that dominate at each tier, see Step 3 ("Which threats to weigh, by tier"); the "what to watch for" column below indicates the sector-specific risk in plain terms, with the relevant ASI categories tagged for reference. The sector rule is the obligation Step 4 maps findings against.

Use case	Tier	Focus principles	Sector rule	What to watch for
Healthcare — GP practice assistant *M365 Copilot pilot †*	**AT1** vendor-embedded	P1, P4, P10, P12	NHS DSPT / Caldicott; national + local acceptable-use policies	Inherits the user's access — can over-surface patient data across the M365 estate; staff over-trusting output *(ASI03, ASI09)*
SME Legal — case & drafting assistant *in legal software; legislation & case law*	**AT2** platform-integrated	P1, P3, P8, P9, P13	SRA obligations; client confidentiality	A wrong or fabricated citation reaching a filing unchecked *(ASI09, ASI06)*
Local government — service-request flow *citizen-developer, with connectors*	**AT3** citizen-built	P2, P3, P5, P6, P8	UK GDPR / DPA; public-law duties; NCSC CAF	Built outside IT; acts on citizen data through connectors with the maker's permissions *(ASI02, ASI03)*
HR — candidate-screening flow *citizen-developer, with connectors*	**AT3** citizen-built	P2, P3, P4, P8, P9	UK GDPR Art. 22 (automated decisions); equality duties	Screening decisions on candidates without meaningful human review *(ASI03, ASI09)*
FinTech — customer-facing decision agent	**AT5** custom in-house	P2, P3, P4, P9, P12	FCA Consumer Duty; may engage FCA enforcement powers	Goal manipulation steering a customer-facing decision; customers' right to contest it *(ASI01)*
Manufacturing — autonomous operations agent	**AT6** externally extended	P2, P6, P7, P8, P11, P12	EU Cyber Resilience Act (CRA)	Physical-world actions; tools and components it did not build *(ASI04, ASI07)*

† M365 Copilot in the NHS. National NHSmail guidance restricts M365 Copilot to administrative use; some trusts permit patient data under local acceptable-use policies, with clinicians accountable and AI not replacing professional judgement. The agentic risk is access amplification, not new access: Copilot inherits the user's existing permissions and surfaces sensitive patient data across the M365 estate faster and more widely than manual search, so loose role-based access or missing sensitivity labels turn into over-exposure. Governance is unsettled — national "admin-only" guidance sits alongside local policies permitting clinical-support use.

Sources: NHSmail Support, M365 Copilot FAQs §1.5 (reviewed Oct 2025), support.nhs.net/knowledge-base/copilot-faq; RDaSH NHS Foundation Trust, AI Policy / M365 Copilot Acceptable Use Policy v1.1 (approved Nov 2025), rdash.nhs.uk/policies/artificial-intelligence-ai-policy.

Tiers are illustrative of how the same method assigns different exposure to different systems; a system's tier can be raised by the risk amplifiers in Appendix A.

Appendix D · Quick Reference

ETSI EN 304 223 phases mapped to UK Code of Practice principles (P1–P13).

ETSI phase	UK Code of Practice principles
Secure design	P1 Raise awareness · P2 Design for security · P3 Evaluate threats & manage risk · P4 Enable human responsibility
Secure development	P5 Identify, track & protect assets · P6 Secure infrastructure · P7 Secure supply chain · P8 Document data, models & prompts · P9 Test & evaluate
Secure deployment	P9 Test & evaluate · P10 Communication with end-users & affected entities
Secure maintenance	P10 Communication & processes · P11 Updates, patches & mitigations · P12 Monitor behaviour
Secure end of life	P13 Proper data & model disposal

Adoption tiers and the OWASP Top 10 for Agentic Applications.

ADOPTION TIERS (AT0–AT8)

Tier	What it is
AT0	Shadow AI — unknown, ungoverned
AT1	Vendor-embedded assistant
AT2	Platform-integrated agent
AT3	Citizen-developer agent
AT4	Code-executing agent
AT5	Custom in-house agent
AT6	Externally extended (MCP, tools)
AT7	Multi-agent / orchestrated
AT8	Federated, cross-organisation

OWASP AGENTIC TOP 10 (ASI01–10)

ID	Risk
ASI01	Agent Goal Hijack
ASI02	Tool Misuse & Exploitation
ASI03	Identity & Privilege Abuse
ASI04	Agentic Supply Chain Vulnerabilities
ASI05	Unexpected Code Execution (RCE)
ASI06	Memory & Context Poisoning
ASI07	Insecure Inter-Agent Communication
ASI08	Cascading Failures
ASI09	Human-Agent Trust Exploitation
ASI10	Rogue Agents