OWASP ASI X EU AI ACT

Gap Assessment Workbook for Agentic AI teams

A practical five-page worksheet for mapping live agent workflows to the OWASP Agentic Security Initiative Top 10, the EU AI Act articles they trigger, the evidence auditors will ask for, and the remediation owners who must close the gaps before August 2, 2026.

RISK SET
10
OWASP ASI risks scored in one working session.

CORE ARTICLES
7+
Articles 9, 10, 12, 14, 15, 17, 26 and related triggers.

CADENCE
Quarterly
Re-run after material model, prompt, tool, or workflow changes.

HOW TO RUN THE WORKSHOP

1. Inventory the workflow. List every agent, tool, MCP server, memory store, and human approval surface involved in production execution.
2. Score ASI applicability. Mark each OWASP ASI risk as in-scope, out-of-scope, or unresolved for your specific workflow.
3. Name the evidence. If you cannot point to a log, report, approval record, or plan, treat the control as a gap rather than a claim.
4. Assign owners and dates. Turn unresolved risks into a tracked backlog with severity, accountable owner, and sign-off.

WHAT THIS WORKBOOK INCLUDES

- Scope sheet for legal role, Annex III classification, and workflow boundaries.
- System inventory for agents, tools, memory, and named human owners.
- ASI applicability matrix with EU AI Act article exposure per risk.
- Evidence review tied to the seven control areas most teams miss.
- Remediation backlog plus security, engineering, and compliance sign-off.

USE WHEN

You already know the theory and need the operating document.

This workbook is for teams past awareness. It is meant for internal review meetings, audit-prep workshops, and budget conversations where someone needs to answer: which agent risks apply, which EU AI Act obligations are triggered, and what proof exists today.

Supporting assets

Pair this PDF with the machine-readable crosswalk CSV and the controls checklist when you need reference detail after the workshop.

WORKBOOK · CHECKLIST · CSV APPENDIX

1. Define scope before scoring risk

Risk analysis fails when teams mix provider obligations, deployer obligations, and unclear workflow boundaries. Fill these fields first.

SYSTEM / WORKFLOW NAME

VERSION / RELEASE IN SERVICE

REVIEW DATE / FACILITATOR

LEGAL ROLE

ANNEX III CLASSIFICATION

NAMED OVERSIGHT OWNER

SYSTEM INVENTORY WORKSHEET

WORKFLOW OR DECISION	AGENT / ORCHESTRATOR	TOOLS, MCP, EXTERNAL SYSTEMS	MEMORY OR RAG	HUMAN OWNER	AFFECTED USERS / NOTES

BOUNDARY CHECKS

- Name the production path. Avoid broad labels like "all copilots" if the approval and tool surfaces differ by workflow.
- Separate provider and deployer tasks. Provider design duties and deployer operating duties often land on different teams.
- List every tool with write capability. File writes, API mutations, email send, code execution, and payment actions matter most.
- Mark persistent state. Any memory, RAG, cache, profile store, or learned feedback path changes the risk picture materially.
- Write down the human stop point. If no one can pause, override, or disable the workflow, that is already a likely Article 14 gap.

Minimum evidence to gather before the review

- Risk register or risk log covering the workflow.
- Tool / integration inventory with owners and change history.
- Approval and override records for human oversight.
- Monitoring plan plus incident or escalation process.

Shortcut for prioritization

If the workflow has tool access, persistent memory, or can act on behalf of users, assume multiple ASI risks are in scope until you can disprove them with architecture and evidence.

High-risk use cases under Annex III, or systems materially influencing those use cases, should be reviewed with compliance participation from the start.

2. Mark which ASI risks actually apply

Use Strong only when the control is implemented and you can name the supporting artifact. Use Partial when the control exists but proof or coverage is incomplete. Use Gap when no defensible evidence exists.

STRONG PARTIAL GAP

ASI	RISK	PRIMARY ARTICLES	APPLIES?	CURRENT STATE	EVIDENCE TO VERIFY
ASI01	Agent Goal Hijack Prompt injection, instruction override, adversarial misuse.	Art. 9, Art. 15	Yes / No / Review	Strong / Partial / Gap	Prompt-injection tests, risk register entry, anomaly alerts.
ASI02	Tool Misuse Unsafe use of legitimate tools and chained systems.	Art. 9, Art. 15	Yes / No / Review	Strong / Partial / Gap	Tool allowlists, tool-call logs, destructive-op approvals.
ASI03	Identity and Privilege Abuse Delegated trust and overbroad credentials.	Art. 9, Art. 15	Yes / No / Review	Strong / Partial / Gap	Scoped credentials, rotation logs, named identity owner.
ASI04	Supply Chain Vulnerabilities Compromised MCP, plugins, or update paths.	Art. 17	Yes / No / Review	Strong / Partial / Gap	SBOM, pinned versions, signature verification, change control.
ASI05	Unexpected Code Execution Agent-generated or agent-triggered code runs outside intent.	Art. 15	Yes / No / Review	Strong / Partial / Gap	Sandbox evidence, execution logs, approval gate records.
ASI06	Memory and Context Poisoning Persistent context corrupts future reasoning.	Art. 15, Art. 10	Yes / No / Review	Strong / Partial / Gap	Memory provenance, write audit trail, integrity checks.
ASI07	Insecure Inter-Agent Communication Spoofed or untrusted A2A messages.	Art. 15	Yes / No / Review	Strong / Partial / Gap	Authenticated channels, signed AgentCards, message logs.
ASI08	Cascading Failures Multi-agent or workflow-level blast radius issues.	Art. 9, Art. 15	Yes / No / Review	Strong / Partial / Gap	Circuit-breaker tests, workflow monitoring, incident procedures.
ASI09	Human-Agent Trust Exploitation Automation bias, manipulative output, missing disclosure.	Art. 14	Yes / No / Review	Strong / Partial / Gap	Training, disclosures, FRIA coverage, verification logs.
ASI10	Rogue Agents Drift, compromised autonomy, missing kill switch.	Art. 9, Art. 14, Art. 15	Yes / No / Review	Strong / Partial / Gap	Kill-switch verification, drift reports, incident escalation path.

Scoring rule

If a risk is in scope and you cannot identify a concrete artifact, do not call the control complete. Record it as Partial or Gap.

This prevents the common failure mode where teams confuse stated policy with operating evidence.

Fast escalation triggers

- Any write-capable tool with no approval gate.
- Persistent memory with no provenance or purge policy.
- External MCP or A2A links with no version pinning or identity verification.
- No independent stop path for a live production workflow.

3. Review the seven control areas that usually break first

These are the minimum control domains that tend to span multiple ASI risks and multiple EU AI Act articles at once.

CONTROL AREA	COVERS	MINIMUM EVIDENCE AN AUDITOR SHOULD BE ABLE TO INSPECT	CURRENT STATE	OWNER	DUE
Prompt-injection defense in depth	ASI01, ASI06 Art. 9, 15	Red-team report, instruction hierarchy tests, untrusted-content policy, anomaly alerts.	Strong / Partial / Gap		
Least-privilege tool scoping	ASI02, ASI03 Art. 9, 12, 15	Tool allowlists, full tool-call lineage, destructive-operation approval records.	Strong / Partial / Gap		
Supply-chain and remote-agent trust	ASI04, ASI07 Art. 17, 15	SBOM, pinned versions, signed AgentCards, MCP verification and change logs.	Strong / Partial / Gap		
Sandboxing plus human approval gates	ASI02, ASI05 Art. 14, 15	Sandbox configuration evidence, execution telemetry, approval queue records, stop procedure.	Strong / Partial / Gap		
Memory provenance and feedback-loop controls	ASI06 Art. 10, 15	Memory-write audit trail, provenance metadata, expiration policy, integrity checks.	Strong / Partial / Gap		
Circuit breakers and blast-radius limits	ASI08 Art. 9, 15, 26	Workflow failure tests, correlation traces, escalation and containment procedure.	Strong / Partial / Gap		
Kill switch and drift monitoring	ASI10 Art. 9, 14, 26	Kill-switch verification, drift thresholds, post-market monitoring path, incident reporting route.	Strong / Partial / Gap		

PRIORITY RULES

- Gap + high-impact workflow should move into remediation immediately.
- Partial + no named owner is usually a program management failure, not an engineering one.
- Control exists but evidence is scattered should be treated as an audit-readiness issue.
- Evidence exists but is manual and fragile should be treated as a scaling risk.

WHAT USUALLY CAUSES FALSE CONFIDENCE

- Policy documents without execution-path enforcement.
- Logs without correlation IDs, review decisions, or full tool parameters.
- Approval steps that exist in theory but are bypassed in urgent operations.
- Monitoring dashboards that do not feed incident handling or change control.

4. Turn findings into an owned remediation plan

The meeting is only useful if gaps leave the room with an accountable owner, target date, and evidence expectation.

REMEDIATION BACKLOG

GAP ID	ISSUE	SEVERITY	LINKED ASI / ARTICLE	OWNER	DUE	EVIDENCE NEEDED

30-DAY OPERATING PLAN

- **Week 1**: finalize workflow boundaries, legal role, and named oversight owner.
- **Week 2**: validate control evidence for every in-scope ASI risk and mark all unsupported claims as gaps.
- **Week 3**: collect missing records, tests, and logs or create remediation tickets where proof does not exist.
- **Week 4**: sign off the review, attach supporting artifacts, and schedule the next quarterly re-run.

REVIEW SIGN-OFF

SECURITY OWNER	ENGINEERING OWNER	COMPLIANCE OR LEGAL	BUSINESS / WORKFLOW OWNER
Name / signature / date	Name / signature / date	Name / signature / date	Name / signature / date

Use this workbook as the primary meeting document.
Best for workshops, owner assignment, and operational sign-off.

Use the controls checklist as the control library.
Best for implementation detail after gaps are identified.

Use the CSV as the machine-readable appendix.
Best for sorting, filtering, or importing the crosswalk into internal tooling.