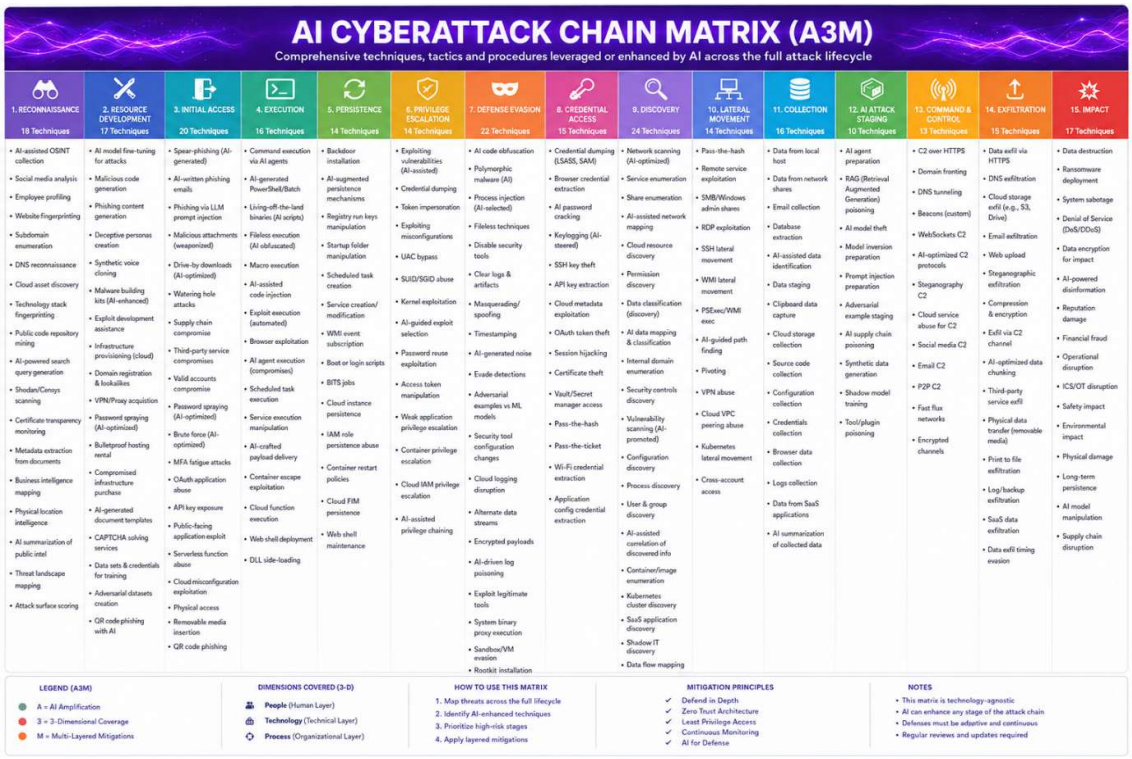
A3M - Agent AI Attack Matrix

Introduction: Why an Agentic AI Attack Matrix Is Needed

Agentic AI systems are no longer “just models.”
They are operational actors: they interpret instructions, browse, retrieve enterprise knowledge (RAG), call tools, trigger automations, and execute actions using real identities and permissions. This shift fundamentally changes cyber risk. In many cases, the attacker does not need to compromise an endpoint or move laterally through a network.
Instead, they can **compromise the decision-and-action layer**: the agent’s planning, tool use, and access context.



Traditional frameworks such as MITRE ATT&CK and the broader ecosystem of threat catalogs remain extremely valuable for describing classic attacker behavior.

But agentic AI introduces new paradigms where common security questions change from *“Was something exploited?”* to *“Was the agent manipulated into doing the wrong thing with legitimate access?”* This is a different problem space with different primitives:

- **Instruction manipulation** (direct and indirect prompt injection via email, docs, web, tool outputs)
- **Tool misuse** (coercing function calls, parameter injection, browser automation abuse, workflow chaining)
- **Access abuse** (OAuth grants, tokens, delegated permissions, service principals/non-human identities)

- **Persistence in memory and automations** (poisoned long-term memory, poisoned RAG corpus, webhooks, mail rules)
- **Human–agent trust failures** (approval exploitation, social engineering amplified by AI-generated content)

What this white paper contributes

This whitepaper introduces the **A3M (Agentic AI Attack Matrix)** to close that gap. A3M is designed to be:

- **Agent-native**: it treats the agent's reasoning, memory, and context handling as first-class attack surfaces.
- **Action-native**: it models how attackers exploit tool calls, browser actions, automations, and workflow chaining.
- **Access-native**: it captures identity, delegated authorization, tokens/scopes, and non-human identity abuse as primary routes to impact.

Most importantly, A3M is built to answer the question security leaders actually need to answer:
"Is our agent stack covered?"
Not just by policies on paper, but by enforceable controls, telemetry, and response actions for the new agentic threat landscape.

Why "a matrix" format?

A matrix is not just a list: it is an operational instrument. It allows teams to:

- map threats to the full lifecycle (from reconnaissance to impact),
- identify **uncovered technique families** unique to agents,
- assign ownership across security, IAM, platform engineering, and governance,
- measure prevention/detection/response maturity over time.

Resume

In short, A3M exists because agentic AI is creating a new security domain where the absence of mappings in traditional frameworks is itself a measurable signal: the controls, detections, and mitigations must evolve. This white paper provides the structure to do that systematically.

A3M stands for Agentic AI Attack Matrix. The name is deliberate: it highlights that the matrix is purpose-built for agentic AI systems: AI that can plan, use tools, and act on behalf of users, rather than traditional endpoints or networks.

The following tables provide the complete list of newly defined techniques. Where the *Remarks* column does not include a reference, no direct mapping to **OWASP Agentic**, **MITRE ATLAS**, or **MITRE ATT&CK** has been identified.

Reconnaissance

Reference	Topic / Agentic Technique	Remark
AAT-1001	Agent surface mapping (chat UIs, APIs, copilots, automations)	Create triggers/callbacks for persistence, C2, or exfiltration.
AAT-1002	Capability probing (tools, plugins, connectors, models)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse | MITRE ATLAS: AML.T0053 LLM Plugin Compromise
AAT-1003	RAG target mapping (what sources are indexed; freshness; scope)	Enumerate assets, permissions, configurations, and data sources. OWASP Agentic: ASI05 Unexpected Code Execution
AAT-1004	Identity perimeter mapping (IdP, tenants, guests, service principals)	Enumerate assets, permissions, configurations, and data sources. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1005	Approval workflow recon (who approves, thresholds, time windows)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1006	Telemetry probing (what gets logged; detection/alert behavior)	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability
AAT-1007	Policy/guardrail boundary probing (refusals, tool gating)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool
AAT-1008	Model fingerprinting (provider, safety mode, tool schemas)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1009	Search open sources for agent configs (docs, repos, tickets)	Technique used to advance objectives within this phase. OWASP Agentic: ASI05 Unexpected Code Execution
AAT-1010	Query public AI services to refine attacks (prompt/voice/deepfake)	Impersonate humans to gain trust or pass approvals. OWASP Agentic: ASI09 Human–Agent Trust Exploitation

Resource Development

Reference	Topic / Agentic Technique	Remark
AAT-2001	Generate social-engineering content (phish, pretexts, scripts)	Technique used to advance objectives within this phase. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-2002	Craft indirect prompt injection payloads (docs, email, web pages)	Inject instructions to override goals or trigger unsafe actions. OWASP Agentic: ASI01 Agent Goal Hijack | MITRE ATLAS: AML.T0052 Phishing (Indirect Prompt Injection); AML.T0051 Prompt Injection (Direct Prompt Injection)
AAT-2003	Acquire infrastructure (lookalike domains, SaaS tenants, webhook endpoints)	Create triggers/callbacks for persistence, C2, or exfiltration. OWASP Agentic: ASI02 Tool Misuse
AAT-2004	Stage malicious OAuth app (branding, scopes, redirect URIs)	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1671 Cloud Application Integration
AAT-2005	Prepare poisoned RAG content (wikis, KBs, tickets, PDFs)	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-2006	Create stealth payloads (unicode/homoglyphs, hidden text, stego)	Blend in or hide intent to evade correlation and monitoring. OWASP Agentic: (no direct mapping assigned)
AAT-2007	Develop tool abuse kit (schema confusion, param injection patterns)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-2008	Build rogue agent persona (docs, onboarding, legitimacy signals)	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-2009	Acquire compromised accounts/sessions (brokered access)	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-2010	Supply-chain positioning (publish poisoned tools/agents/packages)	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI02 Tool Misuse :
AAT-2011	Deepfake assets for approvals (voice/video, chat impersonation)	Impersonate humans to gain trust or pass approvals. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-2012	Adversarial content crafting to trigger unsafe tool routes	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse

Initial Access

Reference	Topic / Agentic Technique	Remark
AAT-1101	Valid account takeover for agent driving	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1102	AiTM / reverse-proxy phishing to capture sessions	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1557 Adversary-in-the-Middle
AAT-1103	OAuth consent phishing / malicious app authorization	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1671 Cloud Application Integration
AAT-1104	Supply-chain compromise of plugin/connector/agent registry	Technique used to advance objectives within this phase. OWASP Agentic: ASI04 Supply Chain & Dependency Risk | MITRE ATLAS: AML.T0053 LLM Plugin Compromise
AAT-1105	Indirect prompt injection via ingested email/document/webpage	Inject instructions to override goals or trigger unsafe actions. OWASP Agentic: ASI01 Agent Goal Hijack | MITRE ATLAS: AML.T0052 Phishing (Indirect Prompt Injection); AML.T0051 Prompt Injection (Direct Prompt Injection)
AAT-1106	Rogue agent invitation into shared workspace/channel	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1107	Shared link delivery with embedded instructions (Drive/SharePoint/Git)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1108	Trusted relationship abuse (B2B guests, inbound federation foothold)	Technique used to advance objectives within this phase. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-1109	Support/ticket channel compromise (instructions as data)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1110	Drive-by content injection (compromised pages the agent retrieves)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1111	Device code phishing (token issued via user out-of-band approval)	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1112	Malicious browser extension / workspace add-on to influence agent flows	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)

AI Model Access

Reference	Topic / Agentic Technique	Remark
AAT-1201	AI inference API access using stolen or mis-scoped credentials	Increase permissions/roles/scopes to expand control. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1202	Full model access via misconfigured hosting / exposed endpoints	Technique used to advance objectives within this phase.
AAT-1203	AI-enabled product/service access (attacker uses built-in agent features)	Technique used to advance objectives within this phase.
AAT-1204	Model endpoint enumeration (routes, versions, safety modes)	Enumerate assets, permissions, configurations, and data sources.
AAT-1205	Prompt crafting to elicit restricted behaviors (jailbreak attempts)	Technique used to advance objectives within this phase. OWASP Agentic: ASI01 Agent Goal Hijack
AAT-1206	Exploit model routing (force higher-capability model or toolset)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse

E

Execution

Reference	Topic / Agentic Technique	Remark
AAT-1301	Direct prompt injection (override intended task/goal)	Inject instructions to override goals or trigger unsafe actions. OWASP Agentic: ASI01 Agent Goal Hijack
AAT-1302	Indirect/tool-output prompt injection (data treated as instructions)	Inject instructions to override goals or trigger unsafe actions. OWASP Agentic: ASI01 Agent Goal Hijack | MITRE ATLAS: AML.T0052 Phishing (Indirect Prompt Injection); AML.T0051 Prompt Injection (Direct Prompt Injection)
AAT-1303	Tool parameter injection (attacker-controlled tool arguments)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1304	Tool selection steering (force powerful tools)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1305	Function-calling / schema confusion (coerce wrong function)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1306	Browser automation abuse (agent navigates to attacker flow)	Create triggers/callbacks for persistence, C2, or exfiltration. OWASP Agentic: ASI02 Tool Misuse
AAT-1307	Unexpected code execution via toolchain (RCE through tools)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1308	Workflow/CI execution triggered by agent actions (pipelines, serverless)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1309	UI redressing against approvals (clickjacking, consent misdirection)	Technique used to advance objectives within this phase. OWASP Agentic: ASI09 Human–Agent Trust Exploitation | MITRE ATT&CK: T1671 Cloud Application Integration
AAT-1310	Command generation for operators (agent produces harmful commands)	Technique used to advance objectives within this phase.
AAT-1311	Cascading loop induction (retry/fan-out amplifies actions)	Technique used to advance objectives within this phase. OWASP Agentic: ASI08 Cascading Failures / Agentic DoS
AAT-1312	Cross-context instruction smuggling (embed directives in structured data)	Technique used to advance objectives within this phase.)

Persistence

Reference	Topic / Agentic Technique	Remark
AAT-1501	Durable OAuth grants / refresh tokens retained	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1671 Cloud Application Integration
AAT-1502	API keys / service principal credentials planted	Technique used to advance objectives within this phase. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1503	Malicious mail rules / forwarding / inbox persistence	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1504	Webhooks / automation triggers created for re-entry	Create triggers/callbacks for persistence, C2, or exfiltration. OWASP Agentic: ASI02 Tool Misuse
AAT-1505	Agent configuration tampering (default tools, policies, allowlists)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI02 Tool Misuse
AAT-1506	Prompt template backdoors (shared system prompts/packs)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1507	Memory poisoning (long-term instruction persistence)	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-1508	Vector DB / RAG poisoning (persist in knowledge base)	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-1509	Rogue agents / scheduled autonomous runs (shadow workflows)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1510	Trusted integration persistence (evil-twin integration remains installed)	Technique used to advance objectives within this phase. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-1511	Session persistence via stolen cookies/tokens	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1512	Service desk persistence (rules/macros keep reinfecting content)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)

Privilege Escalation

Reference	Topic / Agentic Technique	Remark
AAT-1401	Role escalation via cloud directory (add roles/groups)	Increase permissions/roles/scopes to expand control. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1402	Scope escalation (expand OAuth scopes / app permissions)	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1403	Service principal abuse (app roles, delegated permissions)	Increase permissions/roles/scopes to expand control. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1404	Delegation/impersonation abuse (act-as / on-behalf-of)	Impersonate humans to gain trust or pass approvals. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-1405	Conditional access/tenant policy modification	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1406	Cross-tenant / federation privilege pivot	Increase permissions/roles/scopes to expand control. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1407	Approval bypass (social engineering of human-in-the-loop)	Technique used to advance objectives within this phase. OWASP Agentic: ASI08 Cascading Failures / Agentic DoS
AAT-1408	Token exchange abuse (convert low privilege to higher privilege token)	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1409	Abuse break-glass / emergency access accounts	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-1410	Privilege escalation through tool misconfiguration (admin tools exposed)	Increase permissions/roles/scopes to expand control. OWASP Agentic: ASI02 Tool Misuse

Stealth

Reference	Topic / Agentic Technique	Remark
AAT-1601	Low-and-slow multi-turn manipulation (avoid spikes)	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1602	Living-off-the-land SaaS actions (blend with normal usage)	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1603	Masquerading as legitimate workflow (benign naming/labels)	Blend in or hide intent to evade correlation and monitoring. OWASP Agentic: ASI02 Tool Misuse | MITRE ATT&CK: T1036 Masquerading
AAT-1604	Prompt/command obfuscation (unicode, encoding, homoglyphs)	Blend in or hide intent to evade correlation and monitoring. OWASP Agentic: ASI10 Rogue Agents / Observability Failures | MITRE ATT&CK: T1027 Obfuscated Files or Information
AAT-1605	Hidden directives in documents (white text, comments, stego)	Blend in or hide intent to evade correlation and monitoring. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1606	Delay execution (time-shift tool calls)	Blend in or hide intent to evade correlation and monitoring. OWASP Agentic: ASI02 Tool Misuse
AAT-1607	Split actions across identities/tools to reduce correlation	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1608	Data minimization exfil (small periodic exports mimicking routine)	Move sensitive data out via allowed channels or covert paths. OWASP Agentic: ASI10 Rogue Agents / Observability Failures)
AAT-1609	Use legitimate cloud services as staging/C2 (avoid suspicious domains)	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1610	Manipulate chat history to conceal prior malicious prompts	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures

Defense Impairment

Reference	Topic / Agentic Technique	Remark
AAT-1701	Disable/modify cloud audit logs / logging sinks	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI10 Rogue Agents / Observability Failures | MITRE ATT&CK: T1562 Impair Defenses
AAT-1702	Disable/modify monitoring integrations (SIEM/SOAR apps)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI10 Rogue Agents / Observability Failures |MITRE ATT&CK: T1562 Impair Defenses
AAT-1703	Alert suppression (routing rules, muting, notification tamper)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI10 Rogue Agents / Observability Failures | MITRE ATT&CK: (no direct equivalent identified)
AAT-1704	Modify DLP / egress controls to permit exfil	Move sensitive data out via allowed channels or covert paths. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1705	Weaken authentication controls (MFA settings, legacy auth enablement)	Technique used to advance objectives within this phase. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1706	Tamper retention settings (shorten log retention)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1707	Delete/alter agent conversation traces	Damage integrity/availability to disrupt operations or extort. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-1708	Disable security tooling in connected platforms (SaaS security/EDR)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI02 Tool Misuse
AAT-1709	Subvert trust controls (allowlist attacker app/endpoint)	Technique used to advance objectives within this phase. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-1710	Poison detection signals (chaff to overwhelm/raise noise floor)	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI06 Memory & Context Poisoning | MITRE ATLAS: AML.T0046 Spamming ML System with Chaff Data

Credential Access

Reference	Topic / Agentic Technique	Remark
AAT-1801	Steal web session cookie (browser/session theft)	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1539 Steal Web Session Cookie
AAT-1802	Replay cookies (alternate authentication material)	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1550.004 Use Alternate Authentication Material: Web Session Cookie
AAT-1803	Steal application access tokens (OAuth/API tokens)	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1528 Steal Application Access Token; T1671 Cloud Application Integration
AAT-1804	Credential stuffing against SaaS accounts that drive agents	Technique used to advance objectives within this phase. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: (no direct equivalent identified)
AAT-1805	Password scraping/harvesting from corp docs/repos	Technique used to advance objectives within this phase
AAT-1806	Prompt-based secret extraction (coax keys from context)	Technique used to advance objectives within this phase.
AAT-1807	AiTM interception (credentials + session capture)	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATT&CK: T1557 Adversary-in-the-Middle
AAT-1808	Device code phishing (token minted via user approval)	Abuse delegated authorization to gain durable access. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-1809	API secret theft from CI/CD or configuration repositories	Cause direct monetary loss via agent-enabled actions.
AAT-1810	Credential harvest from agent config (stored tool creds)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-1811	RAG credential harvesting (secrets exposed via retrieval)	Technique used to advance objectives within this phase. OWASP Agentic: ASI03 Identity & Privilege Abuse

Discovery

Reference	Topic / Agentic Technique	Remark
AAT-1901	Cloud service discovery (apps, dashboards, tenants)	Enumerate assets, permissions, configurations, and data sources.
AAT-1902	Account & group discovery (roles, owners, approvers)	Enumerate assets, permissions, configurations, and data sources. OWASP Agentic: ASI03 Identity & Privilege Abuse | MITRE ATLAS: (no direct mapping assigned) | MITRE ATT&CK: (no direct equivalent identified)
AAT-1903	Cloud storage object discovery (drives/buckets/sites)	Enumerate assets, permissions, configurations, and data sources. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-1904	Information repository discovery (wiki/tickets/CRM/HR)	Enumerate assets, permissions, configurations, and data sources.
AAT-1905	Permission group discovery (who can access what)	Enumerate assets, permissions, configurations, and data sources.
AAT-1906	Integration discovery (installed apps, webhooks, automations)	Create triggers/callbacks for persistence, C2, or exfiltration. OWASP Agentic: ASI02 Tool Misuse
AAT-1907	Secret location discovery (vaults, env vars, pipelines)	Enumerate assets, permissions, configurations, and data sources.
AAT-1908	Policy discovery (CA policies, sharing policies, DLP policies)	Enumerate assets, permissions, configurations, and data sources.
AAT-1909	Agent configuration discovery (tools enabled, model routing)	Enumerate assets, permissions, configurations, and data sources. OWASP Agentic: ASI02 Tool Misuse
AAT-1910	Model artifact discovery (prompts, evals, embeddings, configs)	Enumerate assets, permissions, configurations, and data sources. OWASP Agentic: ASI06 Memory & Context Poisoning

Lateral Movement

Reference	Topic / Agentic Technique	Remark
AAT-2101	Taint shared content to propagate indirect prompt injection	Inject instructions to override goals or trigger unsafe actions. OWASP Agentic: ASI01 Agent Goal Hijack | MITRE ATLAS: AML.T0052 Phishing (Indirect Prompt Injection); AML.T0051 Prompt Injection (Direct Prompt Injection)
AAT-2102	Share poisoned templates/prompts across teams/workspaces	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-2103	Invite attacker/rogue agent into shared channels/workspaces	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-2104	Pivot via valid accounts across SaaS (SSO sprawl)	Technique used to advance objectives within this phase.
AAT-2105	Remote session hijacking across cloud consoles	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-2106	Tool-to-tool pivot chains (ticket → repo → CI → cloud)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-2107	Inter-agent message forgery/injection (multi-agent systems)	Technique used to advance objectives within this phase. OWASP Agentic: ASI07 Insecure Inter-Agent Communication
AAT-2108	Abuse trusted relationship / B2B guest to move tenants	Technique used to advance objectives within this phase. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-2109	Lateral tool transfer (agent uploads payloads into other systems)	Move sensitive data out via allowed channels or covert paths. OWASP Agentic: ASI02 Tool Misuse

Collection

Reference	Topic / Agentic Technique	Remark
AAT-2201	Email collection (mailboxes, threads, attachments)	Technique used to advance objectives within this phase.
AAT-2202	Data from cloud storage (Drive/SharePoint/S3/etc.)	Technique used to advance objectives within this phase. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-2203	Data from information repositories (wiki/tickets/CRM/HR)	Technique used to advance objectives within this phase.
AAT-2204	Data from configuration repositories (Git/IaC/CI variables)	Technique used to advance objectives within this phase.
AAT-2205	Conversation history collection (agent transcripts)	Technique used to advance objectives within this phase.
AAT-2206	Memory dump/harvest (long-term memory, embeddings)	Technique used to advance objectives within this phase. OWASP Agentic: ASI06 Memory & Context Poisoning
AAT-2207	Browser/session state capture (screenshots, exports where supported)	Steal or replay sessions to operate as a legitimate user. OWASP Agentic: ASI03 Identity & Privilege Abuse
AAT-2208	Automated bulk collection via tool loops	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-2209	Artifact collection (models, prompts, datasets where accessible)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)

AI Attack Staging

Reference	Topic / Agentic Technique	Remark
AAT-3001	Craft adversarial data/instructions to verify exploitability	Technique used to advance objectives within this phase.
AAT-3002	Create proxy AI model / shadow agent for testing & evasion	Technique used to advance objectives within this phase. OWASP Agentic: ASI10 Rogue Agents / Observability Failures
AAT-3003	Generate deepfakes (voice/video) for identity and approvals	Impersonate humans to gain trust or pass approvals. OWASP Agentic: ASI09 Human–Agent Trust Exploitation
AAT-3004	Generate malicious commands/scripts for operator execution	Technique used to advance objectives within this phase.
AAT-3005	Verify attack path (dry-run tool calls, safe probes)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse | MITRE ATLAS: AML.T0042 Verify Attack
AAT-3006	Chaff generation (spam data/prompts to mask signals)	Technique used to advance objectives within this phase. OWASP Agentic: ASI08 Cascading Failures / Agentic DoS | MITRE ATLAS: AML.T0046 Spamming ML System with Chaff Data
AAT-3007	Retrieval content crafting (false RAG entries, deceptive citations)	Technique used to advance objectives within this phase. OWASP Agentic: ASI06 Memory & Context Poisoning

Command and Control

Reference	Topic / Agentic Technique	Remark
AAT-2301	Webhooks as C2 (callbacks to attacker endpoints)	Create triggers/callbacks for persistence, C2, or exfiltration. OWASP Agentic: ASI02 Tool Misuse
AAT-2302	Chat/IM as C2 (bot channels)	Technique used to advance objectives within this phase. OWASP Agentic: (no direct mapping assigned)
AAT-2303	Email as C2 (automated command mailbox)	Technique used to advance objectives within this phase.
AAT-2304	Web service/proxy relays using legitimate SaaS	Technique used to advance objectives within this phase.
AAT-2305	Dynamic resolution (rotate endpoints/domains)	Technique used to advance objectives within this phase
AAT-2306	Traffic signaling (low-volume beacons via benign actions)	Technique used to advance objectives within this phase.
AAT-2307	Multi-stage channels (stage in one tool, execute in another)	Technique used to advance objectives within this phase. OWASP Agentic: ASI02 Tool Misuse
AAT-2308	Agent-to-agent C2 (commands relayed through compromised agents)	Technique used to advance objectives within this phase. OWASP Agentic: ASI07 Insecure Inter-Agent Communication

Exfiltration

Reference	Topic / Agentic Technique	Remark
AAT-2401	Transfer data to attacker cloud account (share/sync/export)	Move sensitive data out via allowed channels or covert paths.
AAT-2402	Automated exfiltration via integrations/webhooks	Create triggers/callbacks for persistence, C2, or exfiltration. OWASP Agentic: ASI02 Tool Misuse
AAT-2403	Exfiltration over web service (POST to attacker)	Move sensitive data out via allowed channels or covert paths.
AAT-2404	Exfiltration via email (external recipients/forwarding)	Move sensitive data out via allowed channels or covert paths.
AAT-2405	Exfiltration via tickets/CRM updates to external parties	Move sensitive data out via allowed channels or covert paths.
AAT-2406	Covert exfil in code artifacts (PRs/commits/releases)	Move sensitive data out via allowed channels or covert paths.
AAT-2407	Scheduled transfer (drip exfil)	Move sensitive data out via allowed channels or covert paths.
AAT-2408	Exfiltration over C2 channel	Move sensitive data out via allowed channels or covert paths.
AAT-2409	LLM data leakage (sensitive responses rendered to attacker)	Move sensitive data out via allowed channels or covert paths | MITRE ATLAS: AML.T0057 LLM Data Leakage
AAT-2410	Extract system prompt / hidden policies and exfiltrate	Move sensitive data out via allowed channels or covert paths. MITRE ATLAS: AML.T0056 LLM Meta Prompt Extraction

Impact

Reference	Topic / Agentic Technique	Remark
AAT-2501	Financial theft / fraudulent transactions via agent tools	Cause direct monetary loss via agent-enabled actions. OWASP Agentic: ASI02 Tool Misuse
AAT-2502	Data manipulation (silent edits to docs/config/code)	Technique used to advance objectives within this phase.
AAT-2503	Data destruction (delete repos/files, wipe)	Damage integrity/availability to disrupt operations or extort.)
AAT-2504	Data encrypted for impact (trigger ransomware-like outcomes)	Damage integrity/availability to disrupt operations or extort.
AAT-2505	Service stop / operational disruption (disable pipelines/workflows)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI02 Tool Misuse
AAT-2506	Denial-of-wallet (cost exhaustion / quota depletion)	Technique used to advance objectives within this phase. OWASP Agentic: ASI08 Cascading Failures / Agentic DoS | MITRE ATLAS: AML.T0034 Cost Harvesting | MITRE ATT&CK: T1496 Resource Hijacking
AAT-2507	Erode integrity (poison KBs/models; degrade decision quality)	Poison data/context to influence agent behavior or retrieval. OWASP Agentic: ASI06 Memory & Context Poisoning)
AAT-2508	External harms (misinformation, unsafe actions, reputational damage)	Technique used to advance objectives within this phase.
AAT-2509	Inhibit recovery (tamper backups/restore paths)	Weaken defenses or visibility to avoid detection and response. OWASP Agentic: ASI10 Rogue Agents / Observability Failures