A REVIEW OF PUBLIC DOCUMENTATION FOR 17 ENTERPRISE AGENT PLATFORMS

# Agent Audit Readiness 2027

What enterprise AI agent platforms document about run records, approvals, limits, audit logs and data lineage, and what buyers should ask.

01 Run records

02 Data lineage

03 Approvals and limits

04 Audit logs and retention

05 Buyer questions

Gautam Mehra · Aman Khanna

Whiteboxx AI

SUMMARY

# Runs are documented. Data versions are not.

We reviewed the public documentation of 17 enterprise AI agent platforms against nine audit criteria on 20 November 2026. Most document what an agent did. None documents which version of the source data it used.

01 **13 of 17 document per-run traces.**

Steps, tool calls and model calls are recorded for each run. The other platforms document narrower traces.

02 **0 of 17 document data version per run.**

No platform documents recording which version or retrieval time of source data a run used. 6 document something about the data, such as retrieval timestamps.

03 **11 of 17 document human approval before actions.**

Every platform documents approvals in some form; six only partly, for example inside workflows only.

04 **6 of 17 document an enforced budget per agent or run.**

Most others document limits only at workspace or account level, or spend alerts without enforcement.

05 **6 of 17 document all three record-keeping basics.**

Run traces, audit log export and configurable retention together.

0 / 17
platforms documenting the data version used in each run

13 / 17
platforms documenting per-run traces

6 / 17
platforms documenting traces, audit export and retention

108
official documentation pages cited as evidence

WHY IT MATTERS IN 2027

# The accountability gap.

When an AI agent acts in business systems, someone will eventually ask what it did, with which data and who approved it. The answer depends on records the platform keeps.

## 49%

of 204 security leaders said they could fully reconstruct every action an agent had taken over the previous 30 days.

SECURITYHQ AND DATADOG, 2026

## 45%

had had an AI-related incident where it was unclear who was accountable.

SECURITYHQ AND DATADOG, 2026

## ~30%

of executives in Europe and the Middle East expect to need real-time monitoring and audit trails of agent decisions by 2027.

IBM, OCTOBER 2026

## 2 Dec 2027

EU AI Act high-risk rules apply to Annex III uses such as hiring and credit. High-risk systems must keep logs (Article 12).

EU AI ACT, AS AMENDED IN 2026

**Why data lineage matters most**

A trace shows that an agent read a document or queried a table. Without the version or retrieval time of that data, a reviewer cannot check the result against the data as it was, and cannot tell which past results to recheck when a source is corrected.

METHOD

# What we checked and how.

For each platform we searched the vendor's official public documentation for nine capabilities. We did not test the products.

| CRITERION | DEFINITION |
|---|---|
| **Run/trace logs** | Per-run record of steps, tool calls and model calls available to customers. |
| **Source attribution** | Outputs cite the data sources or documents used. |
| **Data lineage/versioning** | Records which version or retrieval time of source data was used in a run (agent/configuration versioning alone does not qualify). |
| **Human approval controls** | Configurable human approval before actions / human-in-the-loop. |
| **Permission scoping** | Per-agent limits on tools, data and actions (least privilege). |
| **Budget/usage limits** | Enforced budget or usage limits per agent or per run (spend, credits, tokens, iterations); org-wide caps only count as partial. |
| **Audit log export/API** | Admin audit logs exportable to a SIEM or available via API. |
| **Retention controls** | Configurable retention of logs and transcripts. |
| **Evaluation/testing tools** | Built-in evaluation tooling or test sets. |

## Scoring

**Documented:** clear official documentation. **Partial:** limited, preview or beta, or covering only part of the criterion. **Not found:** no official public documentation found. Where the choice was unclear, we recorded the lower status.

## Sources

Only first-party sources: vendor documentation, admin guides, API references, release notes and trust pages. 108 documentation pages are cited. Every link is listed in the scorecard data file.

THE SCORECARD

# Nine criteria, 17 platforms.

Public documentation checked on 20 November 2026. This is not a ranking: platforms are listed alphabetically.

| PLATFORM | Run traces | Source citations | Data version per run | Human approval | Per-agent permissions | Per-agent budgets | Audit log export | Retention settings | Evaluation tools |
|---|---|---|---|---|---|---|---|---|---|
| Amazon Bedrock AgentCore | ● | ● | ○ | ● | ● | ● | ◐ | ◐ | ● |
| Anthropic Claude Managed Agents (beta) | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ | ◐ |
| Glean Agents | ● | ● | ○ | ● | ◐ | ● | ● | ● | ● |
| Google Gemini Enterprise Agent Platform | ● | ● | ◐ | ● | ● | ◐ | ● | ● | ● |
| IBM watsonx Orchestrate | ● | ● | ○ | ◐ | ● | ◐ | ● | ● | ● |
| Microsoft Copilot Studio | ● | ● | ○ | ● | ● | ● | ● | ● | ● |
| OpenAI ChatGPT Workspace Agents | ● | ◐ | ○ | ● | ● | ◐ | ● | ◐ | ○ |
| Oracle AI Agent Studio | ● | ◐ | ○ | ● | ● | ● | ○ | ○ | ● |
| Palantir AIP | ● | ● | ◐ | ● | ● | ◐ | ● | ● | ● |
| Salesforce Agentforce | ● | ● | ○ | ● | ● | ◐ | ◐ | ◐ | ◐ |
| SAP Joule Studio | ● | ◐ | ○ | ● | ◐ | ● | ◐ | ◐ | ● |
| ServiceNow AI Agents | ◐ | ○ | ○ | ◐ | ● | ○ | ○ | ○ | ● |
| Sierra Agent OS | ● | ○ | ◐ | ◐ | ○ | ○ | ◐ | ○ | ● |
| UiPath Agentic Automation | ● | ● | ◐ | ● | ● | ● | ● | ● | ● |
| Workday agents | ◐ | ◐ | ○ | ◐ | ● | ○ | ◐ | ◐ | ○ |
| WRITER Agent | ◐ | ● | ○ | ● | ◐ | ○ | ● | ● | ◐ |
| Zendesk AI agents | ● | ● | ◐ | ◐ | ◐ | ◐ | ◐ | ● | ◐ |

● Documented ◐ Partial: limited, preview or beta, or covers part of the criterion

○ Not found in public documentation. Not a finding that the capability is absent.

RESULTS BY CRITERION

# Strong on traces, weak on data.

Run traces, approvals, permissions and evaluation tools are widely documented. Data versioning, enforced budgets and retention settings are not.

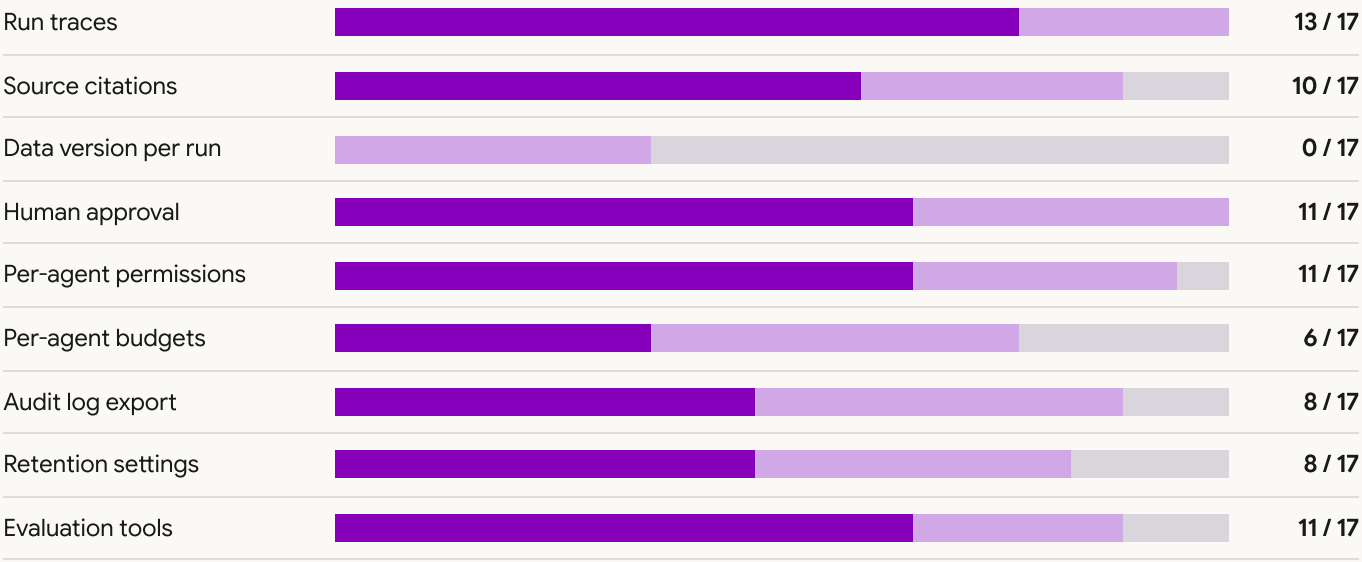



Figure: platforms documenting each criterion in full, out of 17.

| CRITERION | DOCUMENTED | PARTIAL | NOT FOUND |
|---|---|---|---|
| **Run/trace logs** | 13 | 4 | 0 |
| **Source attribution** | 10 | 5 | 2 |
| **Data lineage/versioning** | 0 | 6 | 11 |
| **Human approval controls** | 11 | 6 | 0 |
| **Permission scoping** | 11 | 5 | 1 |
| **Budget/usage limits** | 6 | 7 | 4 |
| **Audit log export/API** | 8 | 7 | 2 |
| **Retention controls** | 8 | 6 | 3 |
| **Evaluation/testing tools** | 11 | 4 | 2 |

FINDINGS 1 TO 5

# What the documentation shows.

1. **Per-run tracing is the most consistently documented criterion.**

   13 of 17 platforms document per-run traces covering steps, tool calls and model calls. The other 4 document narrower traces, for example model calls only. No platform scored “not found”.

2. **No platform documents data lineage at the level of source-data version or retrieval time per run.**

   6 scored partial because their documentation records something about the data, such as retrieval timestamps. Of the 11 scored “not found”, most document agent or configuration versioning, which shows which agent ran but not which data it read.

3. **Human approval controls are documented on every platform, in varied forms.**

   11 of 17 document configurable approval before actions. Partial scores reflect approvals only inside workflows, approvals on configuration changes rather than actions, or early-access programmes.

4. **Per-agent permission scoping is widely documented.**

   11 of 17 document per-agent limits on tools, data or actions. Partial scores mostly reflect beta features or limits set per team or connector rather than per agent.

5. **Enforced budgets per agent or per run are documented by about a third.**

   6 of 17 document an enforced limit per agent or run. 7 more document limits only at workspace, project or account level, fixed limits, or limits in preview.

FINDINGS 6 TO 10

# What the documentation shows.

6. **Full audit log export is documented by fewer than half.**

   8 of 17 document admin audit logs exportable by API or to security tools. Partial scores include exports that cover only some features, are time-limited or are in beta.

7. **Configurable retention is less common than log availability.**

   8 of 17 document configurable retention for agent logs or transcripts. Some partial scores reflect fixed retention windows that cannot be changed.

8. **Built-in evaluation tools are documented by most platforms.**

   11 of 17 document built-in evaluation features or test sets.

9. **Few platforms document the full set of record-keeping basics.**

   6 of 17 document run traces, audit log export and configurable retention together, the combination most relevant to producing records for an after-the-fact review.

10. **Preview and beta labels are common.**

    At least 7 of 17 platforms label one or more of the features assessed here as preview, beta or early access, and one assessed product is in beta as a whole. Several products were renamed during 2026. Scores are likely to change.

FOR BUYERS

# What to ask your vendor.

Use the gaps in the documentation as questions. A vendor may support more than it documents, so ask for a demonstration and keep the answers.

| GAP | ASK THE VENDOR | EVIDENCE TO REQUEST |
|---|---|---|
| **Data version per run** | Can a run record show which version or retrieval time of each source was used? | A run record with source versions or dates |
| **Per-agent budgets** | Can we set a hard limit per agent or per run that stops it? | The setting, and a run that hit the limit |
| **Audit log coverage** | Do your audit logs include agent actions, and can we export them to our security tools? | A sample export containing agent events |
| **Retention** | Can we set how long traces and transcripts are kept? | The retention setting and its limits |
| **Preview features** | Which of the controls we rely on are in preview or beta? | The product's current release status |
| **Approvals** | Can a person approve an individual action, not only a whole workflow? | An approval on a single tool call |

Our 32-question vendor due-diligence questionnaire covers these points and more, with a scoring sheet: whiteboxx.ai/resources/ai-agent-vendor-questionnaire

LIMITATIONS AND RIGHT OF REPLY

# How to read this report.

- Documentation is not capability. A product may do more, or less, than its public documentation describes. We did not test any product.
- “Not found” can mean the capability is absent, undocumented, documented only behind a customer login, described in terms we did not search for, or on pages we could not retrieve.
- Some documentation sites block automated retrieval or need JavaScript. For those, scores rest on official page text shown in search results from the vendor’s own domain, or on official PDFs, and were scored conservatively.
- Sierra has no public documentation site; its scores rest on its own product pages and blog posts.
- Several vendors offer more than one agent product. Scores apply to the product named in the scorecard, not to the vendor’s whole portfolio.
- Products change fast. Results reflect 20 November 2026 only.
- A three-level scale cannot capture depth, defaults, licensing tiers or ease of use.

### Vendor right of reply

Vendors can send corrections with links to official documentation to info@whiteboxx.ai. We will review them, update the scorecard and note each change and its date in the next edition.

Disclosure: Whiteboxx sells an enterprise AI platform, Evidence OS. We did not score our own product.

ABOUT WHITEBOXX

# Evidence OS, extended for your workflows.

Whiteboxx builds Evidence OS, an enterprise AI platform, and extends it for each customer with forward deployed engineers. Each run records its inputs, the sources used with the references needed to find the same data again, the steps, the calculations and the results.

START WITH THE ARCHITECTURE SPRINT

## 30 days · USD 15,000, fixed

We review your AI tools and pilots, choose one workflow with you and deliver a build plan with costs. You can use the plan with us, with your own team or with another partner.

Book a scope call: **whiteboxx.ai/contact** · info@whiteboxx.ai

# Sources and data

Every score in this report links to the official documentation page it relies on. The full scorecard, with 108 documentation links and a one-sentence note for each score, is published as a data file.

1. Scorecard data (JSON), with every documentation link and note
   https://whiteboxx.ai/assets/downloads/whiteboxx-agent-audit-readiness-2027-scorecard.json
2. SecurityHQ, “The accountability gap”, 2026
   https://www.securityhq.com/blog/research-the-accountability-gap/
3. IBM, study of executives in Europe and the Middle East, 14 October 2026
   https://newsroom.ibm.com/2026-10-14-ibm-study-eight-in-ten-emea-enterprises-escalate-ai-incidents-to-board-or-c-suite-as-ai-agents-test-governance
4. EU AI Act, Article 12: record-keeping
   https://artificialintelligenceact.eu/article/12/
5. Loyens & Loeff, EU AI Act amendments under the Digital Omnibus, 2026
   https://www.loyensloeff.com/insights/news--events/news/eu-ai-act-key-amendments-adopted-under-digital-omnibus-package/

### Platforms reviewed

Amazon Bedrock AgentCore · Anthropic Claude Managed Agents (beta) · Glean Agents · Google Gemini Enterprise Agent Platform · IBM watsonx Orchestrate · Microsoft Copilot Studio · OpenAI ChatGPT Workspace Agents · Oracle AI Agent Studio · Palantir AIP · Salesforce Agentforce · SAP Joule Studio · ServiceNow AI Agents · Sierra Agent OS · UiPath Agentic Automation · Workday agents · WRITER Agent · Zendesk AI agents

This report is general information, not legal advice or a product recommendation. Edition 1 · November 2026, published 2026-11-20. Authors: Gautam Mehra and Aman Khanna, Whiteboxx AI. Product names are trademarks of their owners.

You may share this report in full, unchanged, with attribution to Whiteboxx AI. The latest edition is at whiteboxx.ai/resources/agent-audit-readiness-2027.