OpenAI

Workspace agents security overview

Current as of April 29, 2026

Contents

Overview

Workspace agents in ChatGPT Enterprise and Edu help teams turn repeatable work into shared agents that can gather context, follow team processes, take action across approved tools, and keep work moving in ChatGPT or Slack. They can run in the cloud, work on schedules, use connected apps and files, ask for approval when needed, and improve over time as teams use them.

This overview gives security, legal, compliance, IT, and business stakeholders the information they need to evaluate and deploy workspace agents with confidence. It covers the controls admins can use to manage access, app connections, write actions, logging, spend, and rollout, so teams can move faster with AI while staying within their organization's governance requirements.

Further reading

If you're looking for more detail on connected apps, which are a key component of workspace agents, please see our **Apps Security Whitepaper** for a deeper dive on security architecture.

Security

Access, data, systems, and user actions

How is access to data, systems, and user actions protected?

Workspace agents run inside the managed ChatGPT workspace, so access is governed through the same workspace identity, role, app, and connector controls admins already use for ChatGPT Enterprise or Edu. In addition, specifically for workspace agents:

- Admins and owners can use ChatGPT workspace RBAC settings to control who can use agents, build agents, publish agents, publish agents with shared connections, and enable the Slack bot for agents.
- App access and action availability are controlled through workspace app settings and upstream source-system permissions.
- Write-capable actions default to requiring human approval, providing write-action safety control, with flexibility for authors to configure approval behavior.

Workspace agents; Workspace settings

What data can agents access?

An agent can access the data made available through its configured context and tools: instructions, attached files, memory where enabled, connected apps, custom MCPs, Slack channel context for Slack-deployed agents, schedules/triggers, and generated artifacts. The practical data boundary depends on source-system permissions, app scopes, admin-enabled connectors/actions, and authentication mode. Specifically for different agent modes:

- In ChatGPT runs, app access can use the relevant end user connection where supported.
- In Slack and other non-interactive surfaces, the agent generally relies on shared/agent-owned or builder-configured app connections because Slack cannot pause the run to authenticate each invoking user.
- Agent memory is scoped per-user for ChatGPT runs and per-deployed channel for Slack runs; memory is not cross-referenced across ChatGPT and Slack runs.

Workspace agents; Apps in ChatGPT

What high-impact actions are restricted or require approval?

There are two levels of action controls available:

- Admins control which apps and actions are available at the workspace level, including read and write availability where the app control surface supports it. Admins can additionally define parameter constraints that limit the accepted parameters on app action requests across the entire workspace.
- Builders can configure agents to perform write actions only through enabled apps/tools and within the scopes granted to the configured connection. Write-capable actions default to requiring human approval, and builders have the option to configure approval behavior for sensitive workflows. Builders can also define action constraints that limit the accepted parameters on app actions requests for any agent individually.

Workspace agents; Apps in ChatGPT

Compliance

How does this release help support legal, policy, and regulatory obligations?

Workspace agents inherit the broader ChatGPT Enterprise and Edu control plane: workspace identity and role management, SSO/SCIM where configured, app controls, supported data retention and residency commitments, and no training on business data by default.

For Enterprise and Edu customers, the Compliance Platform provides supported logs and metadata for audit, DLP, eDiscovery, or SIEM workflows.

Enterprise privacy; Compliance Platform; Workspace agents

What data is stored, for how long, and how does deletion and retention work?

ChatGPT content and agent-related content follow the same retention, deletion, residency, and no-training commitments applicable to the organization's ChatGPT plan and workspace settings.

Agent-related data can include agent definitions and sanitized snapshots, prompts/instructions, published versions, schedules/triggers, run metadata, agent-authored messages, connector-call metadata, skill usage, memory paths/actions, and generated artifacts/files where applicable.

Enterprise privacy; Compliance Platform

What logs, exports, and documentation are available for audits and review?

The Compliance Platform provides logs and metadata that can connect to eDiscovery, DLP, and SIEM workflows for eligible customers.

- The Logs Platform is designed specifically for compliance needs with: immutable JSONL files, roughly 10-minute windows, p99 under 30 minutes from event time to log inclusion, at-least-once delivery, and event_id-based deduplication.
- Exportable logs include information about agent lifecycle runtime, trigger, connector, skill, and memory events.

Workspace analytics; Compliance Platform; Workspace agents

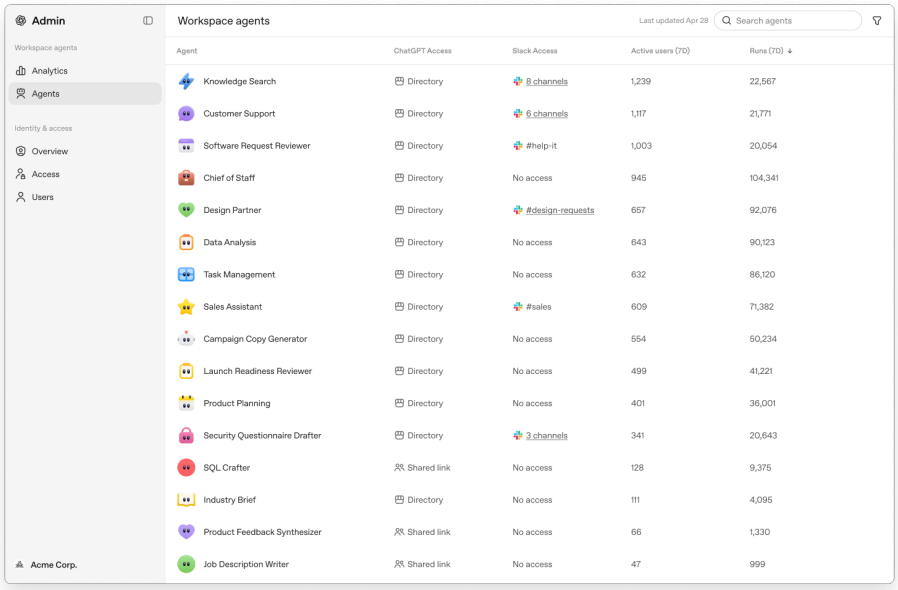
Observability

Admin visibility into usage, actions, failures, and risk

What usage data is available to admins or owners?

Admins and owners have several visibility layers:

- Per-agent level analytics such as run count and unique users for individual agents. This is available for agent builders as well
- Aggregated agent analytics across all agents in the workspace are surfaced in the admin console for admins/owners, including which agents exist and how they are being used. This can be accessed at admin.openai.com.



- The Compliance API additionally exposes the full configuration of every agent, audit logs for every change to every agent, and traces for every run of every agent.

Workspace agents; Global Admin Console

Are prompts, outputs, files, actions, or tool calls logged? Can logs export to monitoring and compliance systems?

Workspace agent compliance logs include agent lifecycle events, run creation/completion/failure, agent-authored messages, connector call requested/completed events, connector OAuth resolution, skill use, trigger create/update/delete, and memory read/write/delete.

Logs can be exported through the Compliance Logs Platform as immutable JSONL files for SIEM, DLP, eDiscovery, data lake, and audit workflows.

Compliance Platform

Can unusual behavior, failures, or usage spikes be detected quickly?

Admins can use the Compliance Logs Platform for recurring ingestion with minutes-level target latency to monitor run failures, connector failures, lifecycle changes, and usage patterns.

Additional observability for agent states and usage patterns is also provided through the admin console for admins/owners and agent analytics for builders.

Compliance Platform

Governance

Rules for who can use, build, publish, and change agents.

How can admins control enablement, permissions, and enforce policies?

Workspace settings and RBAC are the main governance surfaces.

- Admins/owners can control access to agents, agent building, publishing to the workspace directory, publishing agents that use shared connections, and Slack bot usage.
- App and connector availability, including read/write actions where supported, is governed through workspace app controls and source-system permissions.

Workspace agents; Workspace settings, Role-based access controls

Can admins scope access by group, role, workspace, or capability?

Yes. Access can be scoped through workspace roles/RBAC, user or group permissions, app approvals, connector/action availability, sharing/publishing state, Slack channel deployment, and shared connection controls.

Workspace agents; Workspace settings; Apps in ChatGPT

How does approving changes, monitoring use, and managing rollback decisions work?

Builders can preview agents before publishing, keep agents private, share by workspace link, publish where permitted, review version history, duplicate agents, delete agents, and republish earlier versions.

Admins/owners control who can build, publish, run, and use Slack with agents, and can unpublish/delete agents through the Workspace agents API exposed via the Compliance Platform or the admin console at admin.openai.com.

For agents that can write to external systems, admins should define review, approval, rollback, and incident-response expectations before broad rollout.

Workspace agents

Cost

Spend drivers, limits, and monitoring

How does workspace agent usage affect spend?

Workspace agents use credits when they run. Credit use depends on how many people use an agent, how often it runs, how much context or how many large files it processes, and whether it uses tools, apps, custom MCPs, or search. Agents may also use more credits when they run on schedules, work in active Slack channels, or create artifacts like documents, slides, or spreadsheets.

Cumulative credit usage across all workspace agents is visible in ChatGPT Workspace Settings, and per-agent credit usage will be available soon.

Usage limits

What budget, cap, or alert controls are available?

Admins can manage usage and spend through workspace-level spend controls, usage limits, and role-level credit pools or caps. Workspace analytics show aggregate usage patterns, while per-agent usage data is visible in the admin console and in more detail through the Compliance API. Agent-specific budget caps or alerts are not currently exposed as a distinct product surface.

Usage limits; Workspace settings

Change management

Controlled rollout and adoption readiness

How can organizations prepare users and support teams for launch?

Agents are easy to build, with ChatGPT guiding users step by step in conversation. OpenAI also provides starter templates for common use cases, along with learning materials in the OpenAI Academy and Help Center, plus live learning opportunities through events and webinars.

Many teams rely on custom GPTs to power workflows today. OpenAI is exploring options to make it easier to convert existing GPTs into workspace agents and will share more soon.

How can admins help teams build agents smoothly?

Many workspace agent settings use controls that may already be set up in an organization's workspace. Reviewing those settings before launch can help teams build successfully and reduce IT escalations.

Recommended admin steps:

- [ ] **Audit users and groups:** Ensure that users are added to the workspace and in the right groups, if applicable.
- [ ] **Configure custom roles for builders and users:** If using RBAC, map each group to the right ChatGPT custom role, including permissions for using agents, building agents, publishing agents, and publishing agents with shared or agent-owned connections.
- [ ] **Enable approved apps:** Turn on the apps builders and users need for their key workflows, such as Slack, Google Drive, SharePoint, Gmail, Calendar, GitHub, Jira, or Confluence. Confirm which app actions are read-only, which can write to source systems, and whether new actions should be auto-enabled or require admin review.
- [ ] **Make required skills available:** Enable skill permissions for the relevant builder and user groups so agents can use approved repeatable workflows.
- [ ] **Review MCP availability:** If builders need custom MCPs, enable Developer Mode or approved MCP connectors only for the right builder groups.
- [ ] **Confirm Slack setup if needed:** If agents will be deployed to Slack, approve the ChatGPT Agents app and define who can deploy agents in Slack.

Workspace agents, OpenAI Academy

Conclusion

With the right workspace controls in place, organizations can help teams build useful agents while keeping rollout, permissions, app access, and monitoring manageable.

To learn more about deploying workspace agents, contact the OpenAI sales team.