INVICTUS INCIDENT RESPONSE

Frontier Forensics: OpenAI

Part of a series on DFIR in AI environments

A guide to evidence sources, logging, and retention across ChatGPT Workspace, the API Platform, and agentic workloads.

Version 1.0

August 2026

TLP:CLEAR

Executive Summary

OpenAI has two separate evidence environments: ChatGPT Workspace and the API Platform. They have different logs, configuration requirements, and retention. Neither provides a complete record of downstream activity; endpoint, identity, application, and network telemetry remain necessary for a full investigation.

- **ChatGPT Workspace:** Enterprise and Edu organizations have access to 30 days of retained evidence through the Compliance Logs Platform. It covers user, authentication, administrative, and supported product activity. Product specific evidence, such as Codex usage, can provide additional context.

- **API Platform:** API Platform Audit Logs record security and administrative activity involving identities, credentials, permissions, projects, network controls, and organization configuration. However, Audit Logging must be enabled before an incident. API request and response logging, agent traces, and retained application state provide additional workload evidence if and when available.

- **Configuration matters:** Tool access and execution controls, such as MCP, web search, file search, code interpreter, and container network access can help establish what capabilities were available to a user or project at the time of an incident. These controls should be reviewed alongside evidence sources.

- **Evidence availability varies:** Retention is not uniform across OpenAI evidence sources. Some data has defined retention periods, some is retained on a best effort basis, and some may be unavailable depending on configuration, product usage, or data retention settings.

Scope, limitations, and caveats

This guide is intended as a practical starting point for incident responders, not a comprehensive catalogue of every artifact that may exist. OpenAI's products, APIs, controls, and logging capabilities are actively evolving, and the evidence available during an investigation will depend on the services in use, organizational configuration, retention settings, and which logging capabilities were enabled before the incident. Investigators should validate current product behavior and available telemetry at the time of an incident and corroborate findings across multiple sources wherever possible.

A complete investigation will almost always use evidence that OpenAI does not own. For example, identity, endpoint, cloud, application, repositories and MCP logs. Those sources remain essential, but the detailed sections below stay focused on OpenAI native logging and telemetry.

Case Study: In a recent Invictus investigation, endpoint telemetry was central because a personal access token and files the AI could access were present on the affected machine. The OpenAI side of evidence could help reconstruct AI activity, but only host and file telemetry could establish what was present on the endpoint.

Scoping an Investigation

An OpenAI investigation can begin from several different signals. For example, a suspected account compromise, an unexpected administrative change, or suspicious agent activity. The first task is to define what happened, where it happened, and which evidence sources are likely to answer those questions. The table below outlines five common investigative starting points that can help establish an investigation's scope.

Start of an Investigation	ChatGPT Workspace	API Platform
Account compromise Did a threat actor gain access, and what did they do?	**Authentication Logs**: login activity **Conversation Logs:** subsequent user activity	**Audit Logs:** `login.succeeded`, `login.failed`; pivot to subsequent `user.*`, `project.*`, `api_key.*`, and role activity by the same actor
Persistence and Privilege Escalation Did the actor establish durable access or increase privileges?	**Audit Logs:** administrative and access changes	**Audit Logs:** `api_key.created`, `api_key.updated`, `service_account.created`, `service_account.updated`, `invite.*`, `user.*`, `role.assignment.created`, `role.updated`

Start of an Investigation	ChatGPT Workspace	API Platform
Defense Evasion Did the actor weaken logging, identity, or network controls?	**Audit Logs:** security and administrative changes	**Audit Logs**: `scim.disabled`; IP allowlist update or deletion events; external-key/certificate events; organization changes; API Call Logging configuration changes
Data Exposure What data reached OpenAI, and what did the model receive or return?	**Conversation Logs**: prompts and responses; retained files, GPTs, or memories where available	**API Call Logging**: request and response content **Responses, Conversations, Files, Vector Stores**: retained workload state **Audit Logs**: actor or administrative activity
Agent and Codex Abuse What did the agent attempt, who initiated it, and what executed?	**Codex Usage Logs:** coverage, routes, schemas, retention, and permissions **Conversation Logs**: prompts and responses; retained files, GPTs, or memories where available	**Agents SDK Traces**: generations, tool/function calls, handoffs **API Call Logging**: request and response content **Audit Logs**: actor and key context

AI Logging Overview

OpenAI tracks user activity in two different systems, depending on which platform you use:

- **ChatGPT Workspace** splits logs into separate categories (like conversations, logins, audit activity, and Codex usage) and it is delivered through the Compliance Logs Platform.
- **API Platform** groups security activity into a single Audit Log, while handling request and response data through a separate logging setting, and stores other app data in system state.

ChatGPT Workspace Evidence

The workspace side answers what happened in a managed ChatGPT environment. Four sources are important during an investigation:

1. Conversation Logs and Compliance Events focus on user and content activity.
2. Audit Logs focus on administrative changes.
3. Authentication Logs focus on access.
4. Codex Usage Logs focus on coding agent activity.

Evidence source	Retention	Individual / Biz	Enterprise / Edu
Conversation & Compliance Logs	30 days	Not available	Available
ChatGPT Audit Logs	30 days	Not available	Available
ChatGPT Authentication Logs	30 days	Not available	Available
Codex Usage Logs	30 days	Not available	Available

Healthcare and FedRAMP plans do not map cleanly to the standard Enterprise and Edu evidence model. Healthcare documents audit capability, but Compliance Platform entitlement and category coverage should be confirmed in the tenant. As of 4 August 2026, the FedRAMP log endpoint is unavailable; legacy users and conversations remain available, and FedRAMP Codex currently follows an API key path.

API Platform Evidence

Unlike ChatGPT workspace logging, API Platform evidence is not tied to ChatGPT subscription tiers and should be assessed separately. Three sources are important during an investigation.

1. API Platform Audit Logs record security activity across authentication, identities, credentials, privileges, projects, network controls, and administrative changes.
2. API Call Logging can preserve API request and response content when configured.
3. Agents SDK traces provide execution telemetry for supported agent workflows.

Additional evidence may exist outside these logging sources. Stored Responses, Conversations, files, and vector stores can preserve workload content or application state and provide useful investigative pivots, but should not be confused with the platform's primary logging and telemetry sources.

Evidence source	What it gives you	Default On	Retention
API Platform Audit Logs	At least 55 events	No	No fixed TTL
API Call Logging	API request & response content	Yes, Per Call	30 days
Agents SDK Traces	Agent workflows, such as model generations, tool or function calls	Yes	Not publicly specified

Controls that change what logging exists

The following controls are configured per organization and in some cases per project, and can alter logging availability or usefulness:

- **Zero Data Retention.** Excludes customer content from abuse monitoring logs and forces the store parameter to false on the Responses and Chat Completions endpoints. Audit logs are unaffected by ZDR, but agent tracing is unavailable for organizations that use a ZDR policy.
- **Modified Abuse Monitoring.** Excludes customer content from abuse monitoring logs across endpoints while leaving endpoint behaviour otherwise intact. A separate control from ZDR and easily confused with it.
- **API Platform Audit Logs.** These are off by default. Once enabled, it cannot be disabled without contacting OpenAI.
- **API call logging.** A separate organization setting that determines whether the content of supported API calls is retained and visible to organization administrators. These are on by default, but per call rather than for all projects. It can be disabled and/or changed to all projects from the 'data controls' settings.
- **Eyes Off and Safety Retention.** Two further states in which OpenAI may retain content that would otherwise be excluded, with human review either restricted or permitted. Relevant to what OpenAI holds rather than to what the customer can retrieve.
- **Enterprise Key Management.** Application state is encrypted using keys the customer manages in an external key management system, so key availability becomes a precondition for reading stored evidence.
- **Data residency.** Set per project and only at project creation. Determines the region holding customer content at rest, while system data may sit outside it.

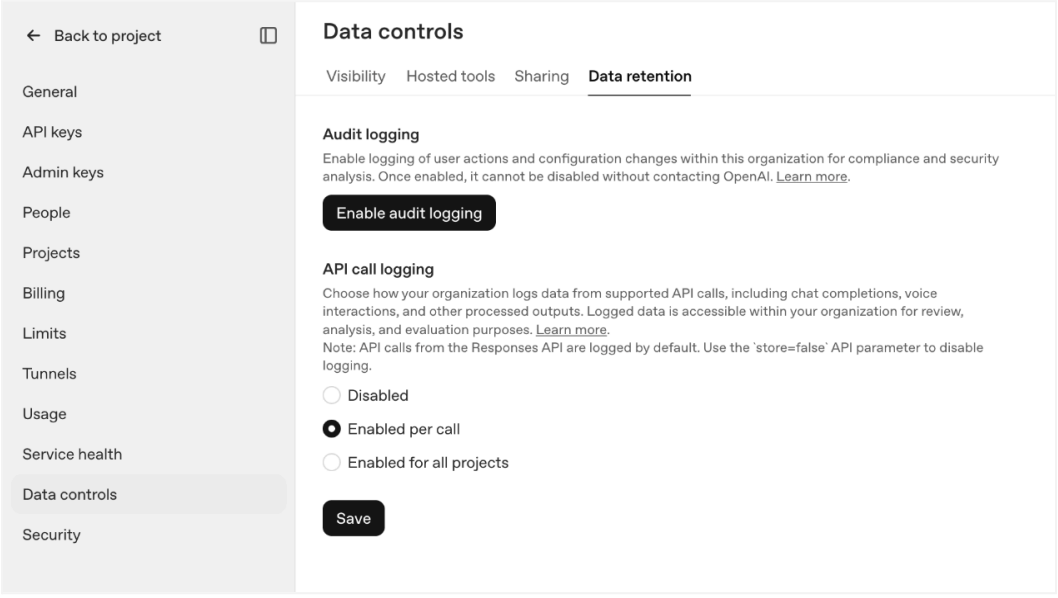

Figure 1 – Data retention settings for the API

Tool Access and Network Controls

OpenAI provides organization level controls that determine which tools and execution capabilities are available to users and projects. These settings are important during an investigation because they help establish what actions were possible in the environment, even when they are not evidence sources.

Administrators can configure access to MCP tools, web search, file search, image generation, and code interpreter. Each capability can generally be disabled, enabled for all projects, or limited to selected projects. Restricting higher risk capabilities to approved projects supports least privilege access.

MCP and web search deserve particular attention because they can introduce external data flows. MCP tools may connect OpenAI workloads to third party services, while web search allows retrieval from the public internet. Investigators should determine whether these capabilities were enabled for the relevant project and, where applicable, identify the external services involved.

File search and code interpreter expand what data a user or application can access and process. File search can retrieve information from configured organizational data sources, while code interpreters can execute code and process uploaded or generated files. Their status can therefore help define the scope of an investigation.

Finally, container network mode controls outbound network access for container based tools. Because outbound connectivity can allow code running in a container to communicate with external systems, it should be treated as a higher risk control. Where enabled, investigators should consider whether external destinations, data transfers, or tool driven network activity are relevant to the incident.

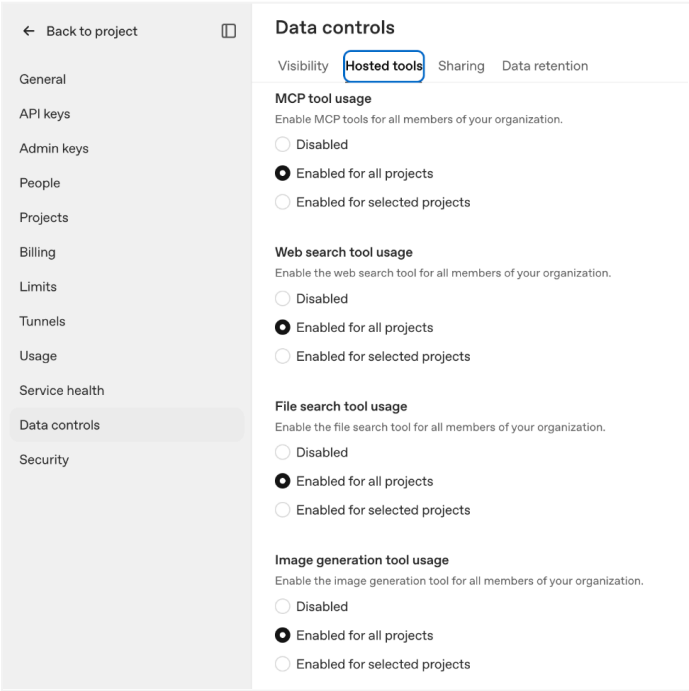

Figure2 – Data controls for hosted tools

OpenAI Logs and Evidence Sources

1. Conversation & Compliance Logs

These are the core ChatGPT workspace activity records delivered through the Compliance Logs Platform for Enterprise and Edu. They capture supported workspace events and can reference conversations, uploaded files, GPTs, memories, users, and other supported objects that can be retrieved through companion compliance interfaces. The delivery platform retains log data for 30 days.

Value	Limitations
This is a top source for reconstructing supported user activity in ChatGPT, tying an account and timestamp to a recorded workspace event, and locating the content or object behind that event.	Coverage can be limited, requiring corroborating evidence from other sources. For example, a download event may show that a file reached an endpoint, but confirming whether it was opened or executed would require endpoint telemetry. Also, object retention is separate from log retention, so an investigator may still see a historical event referencing a file even if that object has been deleted.

2. ChatGPT Audit Logs

ChatGPT Audit Logs are a distinct workspace log category for supported administrative and control changes. They are useful when the incident includes malicious insider activity, permission changes, or attempts to weaken workspace controls. They share the Compliance Logs Platform 30 day retention window.

Value	Limitations
This source builds the ChatGPT workspace timeline, helping to answer administrative changes, which actor or object was associated with it, and when it occurred.	It does not show model requests, generated outputs, or endpoint activity.

3. ChatGPT Authentication Logs

Authentication Logs are a useful starting point for investigating account takeover and suspicious logins. They provide the ChatGPT-side authentication timeline, which can be correlated with the organization's identity and endpoint evidence. The fields available in ChatGPT Authentication Logs differ from those in the API Platform Audit Logs described in Section 5. For example, fields documented for API Platform Audit Logs should not be assumed to be available in ChatGPT Authentication Logs.

Value	Limitations
This source answers whether OpenAI recorded authentication activity and provides the principal and session context needed to scope what followed.	The logs do not provide the full upstream authentication chain. SSO, MFA, and other IdP decisions must be reconstructed from identity provider logs.

4. Codex Usage Logs

Codex usage is available through the Compliance API for supported Codex clients and execution paths. OpenAI's public Codex documentation states that the authenticated Admin API reference is the source of truth for event coverage, routes, schemas, retention, and permissions.

Value	Limitations
This source helps establish which Codex user, session, and activity were recorded, and defines the activity that should then be tested against the developer host, repository, and pipeline evidence.	API authenticated Codex activity should be scoped against the associated API organization and its data controls rather than assuming the ChatGPT workspace is authoritative for that execution path.

5. API Platform Audit Logs

API Platform Audit Logs record security and administrative changes within an OpenAI API organization. The current schema contains at least 55 event types spanning identity and access, service accounts, API keys, projects, roles and permissions, organization configuration, network controls, certificates and external keys, and more.

Value	Limitations
This is a primary OpenAI source for scoping credential and organization control changes in an API compromise. Once enabled, they cannot be turned off without contacting OpenAI, making them resistant to tampering by threat actors.	OpenAI publishes no fixed retention period and retains these logs on a best-effort basis without guaranteeing permanent availability, so durable history depends on the customer exporting them.

Events that matter: API Platform Audit Logs

Investigation	Events to prioritize	What they can establish
Account compromise	login.succeeded, login.failed, logout.succeeded, logout.failed	Authentication timeline plus actor and session context, including source address, geolocation, user agent, and TLS fingerprint
Unauthorized access	invite.sent / accepted / deleted, user.added / updated / deleted	New or modified organization access
Credential persistence	api_key.created / updated / deleted, service_account.created / updated / deleted	Long-lived programmatic access or credential tampering. Check the created service account role
Privilege escalation	role.created / updated / deleted, role.assignment.created / deleted	New or widened roles, the permissions added, and the principals and resources they were bound to
Staging and segregation of activity	project.created, project.updated, project.archived, project.deleted	A new or altered project used to hold keys and activity away from monitored projects, and destructive project changes
Logging tampering	organization.updated where changes_requested includes api_call_logging or api_call_logging_project_ids	Whether API call content logging was disabled, narrowed to per call, or carved down to selected projects, and which projects were excluded
Network control weakening	ip_allowlist.created / updated / deleted, ip_allowlist.config.activated / deactivated	Expansion, removal, or toggling of network access restrictions. Check allowed_ips for overly broad ranges
Identity control weakening	scim.enabled / disabled, group.created / updated / deleted	Provisioning and group-governance changes, including deprovisioning bypass
Key and certificate tampering	certificate.created / updated / deleted, certificates.activated / deactivated, external_key.registered / removed	Certificate lifecycle and enforcement changes, and customer-managed key registration or removal
Destructive and infrastructure activity	resource.deleted, tunnel.created / updated / deleted, checkpoint.permission.created / deleted	Destructive or infrastructure-level administrative activity

6. API Call Logging

API call logging is a separate organization setting from both the Audit Log and abuse monitoring. When enabled it retains the content of supported API calls, meaning prompts, completions, and usage metadata, and the retained data surfaces in the platform logs view for organization administrators.

Value	Limitations
This source answers the content questions for prompt injection, data exposure, and tool abuse attacks.	It is set to per call by default instead of all projects. Unlike audit logging it can be disabled, which makes it an anti-forensics target as well as an evidence source.

7. Agents SDK Traces

The OpenAI Agents SDK are enabled by default and can trace the workflow hierarchy of agents. Use these logs to reconstruct agent behavior, and corroborate actions with the systems the agent interacted with.

Value	Limitations
This source supports reconstructing an agent's workflow, including model generations, tool or function calls, handoffs, and guardrail activity.	A missing trace does not prove the agent did not run. Tracing can be disabled or redirected to another processor, and OpenAI-hosted traces are unavailable under Zero Data Retention.

Other evidence to preserve

OpenAI logs may reference application objects that contain evidence relevant to an investigation. These are not log sources, but they should be preserved when identified because their retention can differ from the telemetry that referenced them.

- **Responses and Conversations** may preserve model output, tool call arguments, or conversation state.
- **Files** may contain the sensitive, malicious, or injected content involved in the incident.
- **Other referenced objects**, such as vector stores or Assistants resources, may provide additional context where the affected application used them.

Retention varies by object and configuration, so preserve relevant application state early rather than assuming the retention period of a related log source applies. Stored objects can provide valuable evidence of content and state, but their existence alone does not establish how they were used. For example, a stored file does not prove that a model accessed it, and a recorded tool call does not by itself prove the downstream action succeeded.

Cheat Sheet

To support investigations, we have also developed a simple guide or a cheat sheet that brings together key evidence sources, retention considerations, and investigation paths across ChatGPT Enterprise, the API Platform, Codex, and agentic workloads.

The guide is intended as a quick operational reference for incident responders and can be downloaded from our GitHub repository: https://github.com/invictus-ir.

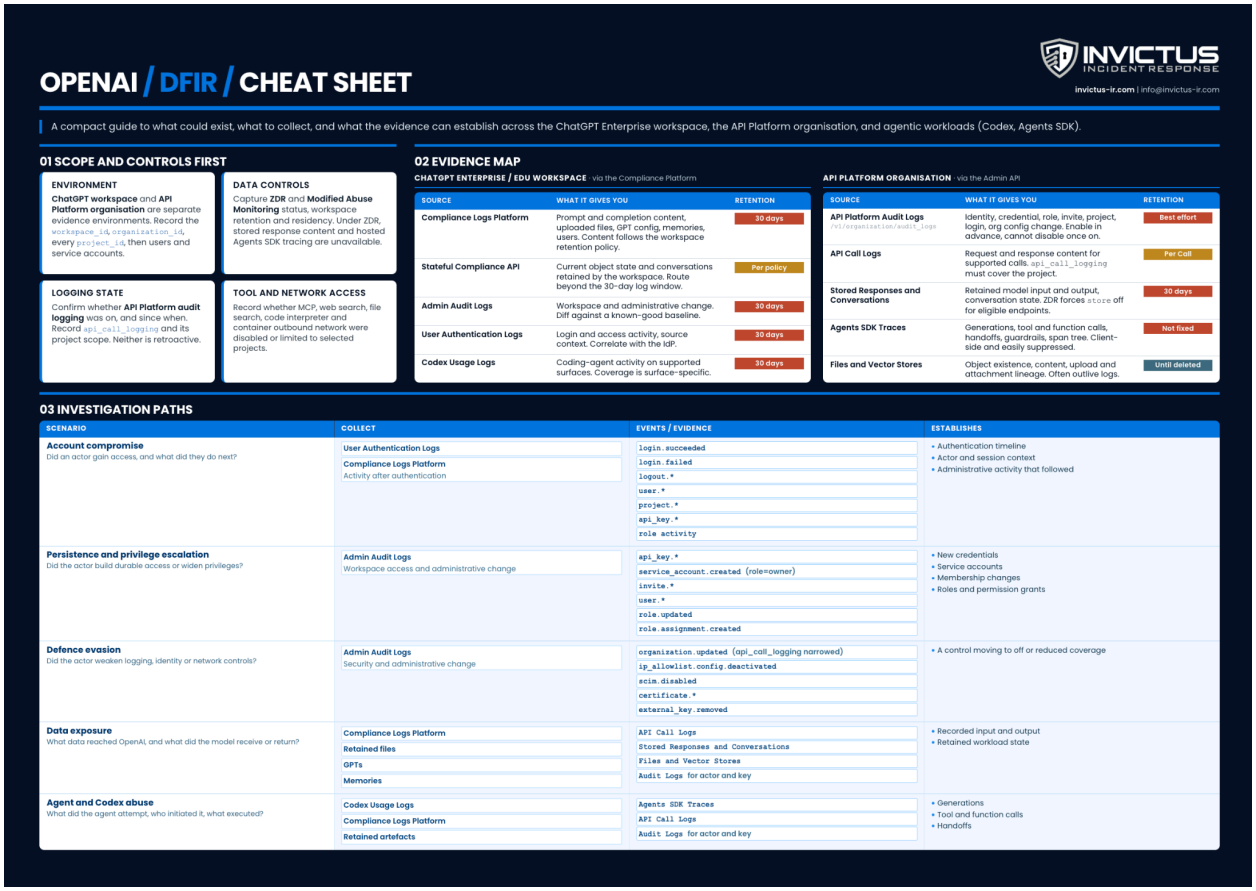

Figure 3 – OpenAI cheat sheet for incident response investigations

References

1. Compliance Platform for Enterprise and Edu Customers: https://help.openai.com/en/articles/9261474-openai-compliance-platform-for-enterprise-and-edu-customers

2. ChatGPT agent: https://help.openai.com/en/articles/11752874-chatgpt-agent

3. GPT Actions introduction: https://developers.openai.com/api/docs/actions/introduction

4. GPTs in ChatGPT: https://help.openai.com/en/articles/8554407-gpts-in-chatgpt

5. Apps in ChatGPT: https://help.openai.com/en/articles/11487775-connectors-in-chatgpt

6. ChatGPT Work Admin FAQ: https://learn.chatgpt.com/docs/enterprise/work-admin-faq

7. ChatGPT Enterprise & Edu release notes: https://help.openai.com/en/articles/10128477-chatgpt-enterprise-edu-release-notes

8. ChatGPT Business release notes: https://help.openai.com/en/articles/11391654-chatgpt-business-release-notes

9. SSO for ChatGPT Business FAQ: https://help.openai.com/en/articles/11489188-sso-for-chatgpt-business-faq

10. SCIM Integration FAQ: https://help.openai.com/en/articles/10011769-scim-integration-faq

11. Global Admin Console: https://help.openai.com/en/articles/12289294-global-admin-console

12. OpenAI Enterprise Key Management: https://help.openai.com/en/articles/20000943-openai-enterprise-key-management-ekm-overview

13. EKM Technical FAQ: https://help.openai.com/en/articles/20000945-ekm-technical-faq

14. Data controls in the OpenAI platform: https://developers.openai.com/api/docs/guides/your-data

15. OpenAI Apps SDK full technical reference: https://developers.openai.com/apps-sdk/llms-full.txt

16. Tracing Logging: https://openai.github.io/openai-agents-python/tracing/

17. AI Tooling Visibility by Monad: https://cdn.prod.website-files.com/64eccd375717312326d818c7/6a693644b48431096abc2b9d_9b243fabb059559a41b70f60a0f77b2e_Monad_%20A%20Security%20Field%20Guide%20to%20AI%20Tooling%20Visibility_eBook.pdf

Disclaimer

This white paper is published by Invictus Incident Response for informational and educational purposes only. It reflects the authors' analysis and opinions as of the date of publication and is subject to change without notice.

The information contained in this document is provided in good faith; however, Invictus Incident Response makes no representations or warranties, express or implied, regarding its accuracy, completeness, or suitability for any particular purpose. Recommendations and examples are intended as general guidance and should be evaluated in the context of each organization's technical environment, legal obligations, regulatory requirements, and risk appetite.

Nothing in this publication constitutes legal, regulatory, financial, or professional advice, nor should it be interpreted as creating a contractual obligation or professional engagement. Organizations should seek advice from appropriately qualified legal, regulatory, and technical professionals before implementing security, forensic, logging, or incident response measures.

To the fullest extent permitted by applicable law, Invictus Incident Response shall not be liable for any direct, indirect, incidental, consequential, or other loss arising from the use of, or reliance upon, this publication.

References to third-party products, platforms, vendors, or services are provided solely for illustrative purposes and do not constitute an endorsement or recommendation unless explicitly stated. Any trademarks mentioned remain the property of their respective owners.

Copyright

About Invictus Incident Response

We founded Invictus Incident Response to meet the surge in cloud cybersecurity incidents and to close the gap in truly effective cloud incident response. As seasoned veterans in cloud security, we've seen firsthand that handling complex incidents takes more than reacting. It demands proactive, expert intervention. Our mission is simple: help you emerge stronger and remain unconquered in the face of cyber adversity. That's the Invictus spirit.

We're an incident response company at heart, and investigating AI platforms is where the discipline is heading. We've handled the full range of incidents across AWS, Azure, and Google Cloud, and because the compute behind AI mostly lives in that same cloud, the hard-won experience we built there is exactly what makes us ready for what comes next. We love the cloud, and we're bringing everything it taught us to the next frontier. We help our clients stay undefeated.

General Enquiries

info@invictus-ir.com

Website

invictus-ir.com

Under Attack? 24×7 Incident Response

Email

cert@invictus-ir.com

North America

+1 470 660 4114

Europe

+31 852 087 843