INTELLIGENCE BRIEFING
Security Command Center
TLP:CLEAR
2026-06-01 18:43 UTC

Shadow AI Governance Gap: 70% of Enterprise AI Operates Outside Security Controls as Attack Surface Expands

GOVERNANCE | HIGH | CVSS 7.5

SCC Item ID	SCC-GOV-2026-0043
Type	Governance
Severity	HIGH
CVSS Base Score	7.5
Affected Products	Enterprise SaaS platforms (generic), LLM/GenAI tools (generic), agentic AI workflows (generic); CrowdStrike Falcon AIDR and Falcon Platform referenced as detection/response tooling
Discovery Source	Rss:T1 Threatintel

Executive Summary

Enterprises are deploying AI tools, APIs, and autonomous workflows at a pace that security teams cannot track, Lenovo research indicates 70% of enterprise AI operates outside governance controls, and a Salesforce survey found 55% of employees use unapproved AI tools. These unmanaged deployments create non-human identities with excessive permissions, expose sensitive data to unvetted third-party APIs, and introduce prompt injection vectors that bypass traditional DLP and CASB controls entirely. The business risk is quantifiable: data loss, regulatory violation, and supply-chain compromise through channels security programs were not designed to monitor.

Technical Analysis

This is a structural governance and architecture risk, not a discrete vulnerability. No CVE is assigned. The failure modes map to CWE-284 (Improper Access Control), CWE-285 (Improper Authorization), CWE-276 (Incorrect Default Permissions), CWE-359 (Exposure of Private Personal Information), and CWE-532 (Insertion of Sensitive Information into Log Files). The primary attack surface components are: (1) unmanaged non-human identities (NHIs), OAuth tokens, API keys, and service accounts created by unsanctioned AI integrations with excessive permission scopes; (2) prompt injection in LLM-integrated pipelines (T1059) enabling instruction manipulation or data exfiltration (T1048, T1567); (3) sensitive data routed to third-party AI APIs without data classification or contractual controls (T1530, T1552, T1552.001); (4) agentic workflows with autonomous execution capability that do not trigger DLP or CASB signatures (T1199, T1078, T1078.004); and (5) exposed AI

API endpoints accessible without authentication controls (T1190). Traditional perimeter and endpoint controls were not designed to cover these vectors. CrowdStrike Falcon AIDR and Falcon Platform are referenced in source material as detection and response tooling for AI-specific threat activity, though no vendor patch or remediation advisory exists, this risk requires governance intervention, not a software update.

Action Checklist

1. Step 1: Containment, Enumerate all AI-related SaaS applications, API integrations, and browser extensions in use across the organization using CASB telemetry and DNS/proxy logs; block unapproved AI tool domains at the proxy or firewall until reviewed. Apply NIST AC-20 (Use of External Systems) to require documented approval before any external AI system connects to enterprise data or identity infrastructure.
2. Step 2: Detection, Query identity provider logs (Entra ID, Okta, etc.) for OAuth grants and service account creation events linked to AI vendor domains. Search endpoint and proxy logs for traffic to known LLM API endpoints (api.openai.com, api.anthropic.com, generativelanguage.googleapis.com, and equivalents) from unapproved sources. Review CASB shadow IT reports for AI tool categorization. Implement CIS 8.2 (Collect Audit Logs) coverage across SaaS connectors. Flag NHIs with scopes exceeding least-privilege baselines per NIST AC-6 (Least Privilege) and AC-3 (Access Enforcement).
3. Step 3: Eradication, Revoke OAuth grants and API keys associated with unapproved AI tools identified in Step 2 using credential management practices and account permission review procedures. Remove or scope-down NHI permissions to minimum required access. Enforce NIST AC-5 (Separation of Duties) for agentic workflow service accounts. Document exceptions through a formal AI tool intake process per NIST AC-1 (Policy and Procedures) and CIS 6.1 (Establish an Access Granting Process).
4. Step 4: Recovery, Validate that approved AI integrations operate under documented data handling agreements and that NHIs have been inventoried per CIS 1.1 (Establish and Maintain Detailed Enterprise Asset Inventory) and CIS 5.1 (Establish an Inventory of Accounts). Confirm DLP rules cover API egress paths and that CASB policies flag new AI tool adoption in real time. Re-audit OAuth grant inventory monthly per CIS 6.2 (Establish an Access Revoking Process).
5. Step 5: Post-Incident, Establish a formal AI governance policy addressing uncontrolled AI: required security review before AI tool adoption, data classification requirements for AI-connected systems, NHI lifecycle management, and prompt injection testing requirements for LLM-integrated pipelines. Map controls to NIST AC-20 (External Systems), AC-1 (Policy), AC-6 (Least Privilege), AU-2 (Event Logging), AU-6 (Audit Record Review), and SI-4 (System Monitoring). The core control gap this risk exposes is the absence of a defined intake and authorization process for AI tools, treat AI adoption like third-party software procurement, not self-service SaaS.

IR / Forensic Enrichment

Triage Priority	URGENT

Escalation Criteria	Escalate immediately to legal, privacy counsel, and executive leadership if proxy or CASB logs confirm that data classified as PII, PHI, PCI-DSS cardholder data, or regulated IP was transmitted to unapproved LLM API endpoints, as this may trigger breach notification obligations under GDPR Article 33, HIPAA §164.400, or state privacy statutes, or if any NHI is found with tenant-wide administrative scopes (e.g., Global Administrator, Organization.ReadWrite.All) granted to an unapproved AI vendor application.
Recovery Notes	Post-containment, monitor OAuth grant creation events and proxy logs to LLM API endpoints daily for a minimum of 60 days to detect re-emergence of shadow AI tool adoption, as employees frequently re-authorize tools after revocation if no approved alternative is provided. Verify that any agentic workflow pipelines (LangChain agents, AutoGPT-style orchestrators, Zapier/Make automations connecting to AI APIs) have been re-inventoried and operate only under service accounts with documented, scoped permissions — these are the highest-risk NHIs because they can autonomously exfiltrate data through chained API calls without human review. Confirm that CASB or proxy policies are configured to alert in real time on first-seen access to net-new LLM API hostnames, not just the known domains enumerated in Step 1, as the AI vendor landscape changes rapidly and new endpoints emerge frequently.
Forensic Artifacts	Entra ID / Okta OAuth 2.0 consent grant audit logs — contains application display name, client ID, granted permission scopes, consenting user UPN, and consent timestamp; this is the primary artifact establishing which unapproved AI tools received delegated or application-level access to enterprise data and for how long | Proxy or Zscaler HTTPS inspection logs for POST requests to LLM API endpoints — User-Agent strings distinguish human browser-based tool use from programmatic NHI/agentic API calls; request body size indicates volume of data sent as prompt context to external AI models | DNS resolver query logs (Pi-hole, Windows DNS debug logging, or resolver export) filtered for LLM vendor hostnames — provides the most complete enumeration of AI tool usage across the environment including tools that operate outside the browser and may bypass CASB coverage | Git repository scan results from truffleHog or git-secrets against internal code repositories — surfaces hardcoded OpenAI, Anthropic, or other LLM API keys committed by developers integrating shadow AI into internal tooling, which represent persistent credential exposure surviving OAuth revocation | Browser extension manifests and permissions from managed endpoints — Chrome extension JSON manifests stored at %LOCALAPPDATA%\Google\Chrome\User Data\Default\Extensions\[extension_id]\manifest.json; AI-integrated extensions (e.g., Compose AI, Grammarly with GPT backend, browser-based Copilot extensions) with 'storage', 'tabs', or '' permissions represent unmonitored data egress channels that bypass proxy inspection when operating on HTTPS pages

Per-Action IR Details

Step 1: Containment — Enumerate all AI-related SaaS applications, API integrations, and browser extensions in use across the organization using CASB telemetry and DNS/proxy logs; block unapproved AI tool domains at the proxy or firewall until reviewed. Apply NIST AC-20 (Use of External Systems) to require documented approval before any external AI system connects to enterprise data or identity infrastructure.

NIST Phase: Containment

Reference: NIST 800-61r3 §3.3 — Containment Strategy

Controls: NIST AC-20 (Use of External Systems), NIST AC-4 (Information Flow Enforcement), CIS 4.4 (Implement and Manage a Firewall on Servers), CIS 4.5 (Implement and Manage a Firewall on End-User Devices), CIS 2.3 (Address Unauthorized Software)

Compensating: Export Squid or pfSense proxy logs and filter for DNS queries or CONNECT requests to known LLM API hostnames (api.openai.com, api.anthropic.com, generativelanguage.googleapis.com, api.cohere.ai, api.mistral.ai).

On Windows endpoints, run: Get-DnsClientCache | Where-Object {$_.Entry -match 'openai|anthropic|gemini|cohere|mistral'} to surface AI traffic without a CASB. For browser extensions, query Chrome's local extension manifest path (%LOCALAPPDATA%\Google\Chrome\User Data\Default\Extensions) or use osquery: SELECT name, identifier, version FROM chrome_extensions WHERE permissions LIKE '%api%'; to enumerate extensions with broad API access.

Evidence: Before blocking, capture a full DNS query log export from your recursive resolver or proxy covering the prior 90 days — AI tool discovery depends on hostname patterns since SaaS apps rarely use static IPs. Preserve CASB shadow IT category reports as a dated snapshot showing the pre-containment AI tool footprint. Export firewall flow logs showing destination ASNs associated with OpenAI (AS54113), Anthropic, and Google AI (AS15169) to establish a baseline of data volume transmitted to these endpoints before any intervention.

Step 2: Detection — Query identity provider logs (Entra ID, Okta, etc.) for OAuth grants and service account creation events linked to AI vendor domains. Search endpoint and proxy logs for traffic to known LLM API endpoints (api.openai.com, api.anthropic.com, generativelanguage.googleapis.com, and equivalents) from unapproved sources. Review CASB shadow IT reports for AI tool categorization. Implement CIS 8.2 (Collect Audit Logs) coverage across SaaS connectors. Flag NHIs with scopes exceeding least-privilege baselines per NIST AC-6 (Least Privilege) and AC-3 (Access Enforcement).

NIST Phase: Detection Analysis

Reference: NIST 800-61r3 §3.2 — Detection and Analysis

Controls: NIST AC-6 (Least Privilege), NIST AC-3 (Access Enforcement), NIST AU-2 (Event Logging), NIST AU-6 (Audit Record Review, Analysis, and Reporting), CIS 8.2 (Collect Audit Logs), CIS 5.1 (Establish and Maintain an Inventory of Accounts)

Compensating: For Entra ID without a SIEM: use the Microsoft Graph API or Azure AD PowerShell to enumerate OAuth grants: Get-MgOAuth2PermissionGrant -All | Where-Object {$_.Scope -match 'Mail|Files|User.ReadWrite'} | Export-Csv oauth_grants.csv. For Okta: pull the System Log via API filtering on event_type eq 'app.oauth2.token.grant' and correlate client_id values against a manually maintained list of approved AI vendor application IDs. For NHI detection, run: Get-MgServicePrincipal -All | Select DisplayName, AppId, CreatedDateTime | Where-Object {$_.DisplayName -match 'ai|gpt|llm|bot|agent'} to surface non-human identities created by AI tool OAuth flows. Use the free Sigma rule 'oauth_application_consent_phishing' as a detection baseline and adapt it for AI vendor app IDs.

Evidence: Preserve Entra ID or Okta audit logs showing OAuth consent grant events — specifically the application display name, granted scopes (e.g., Files.ReadWrite.All, Mail.Read), the consenting user's UPN, and the timestamp. Capture service principal creation logs that show NHIs instantiated outside your standard provisioning workflow. Export proxy or Zscaler access logs showing the User-Agent strings from API calls to LLM endpoints — programmatic API calls from scripts or agentic workflows will show non-browser User-Agent values (e.g., 'python-httpx/0.24.0', 'langchain/0.1.x') which distinguish automated NHI traffic from human browser-based tool use.

Step 3: Eradication — Revoke OAuth grants and API keys associated with unapproved AI tools identified in Step 2 using D3-CRO (Credential Rotation) and D3-UAP (User Account Permissions) countermeasures. Remove or scope-down NHI permissions to minimum required access. Enforce NIST AC-5 (Separation of Duties) for agentic workflow service accounts. Document exceptions through a formal AI tool intake process per NIST AC-1 (Policy and Procedures) and CIS 6.1 (Establish an Access Granting Process).

NIST Phase: Eradication

Reference: NIST 800-61r3 §3.4 — Eradication

Controls: NIST AC-5 (Separation of Duties), NIST AC-1 (Policy and Procedures), NIST AC-2 (Account Management), NIST AC-6 (Least Privilege), CIS 6.1 (Establish an Access Granting Process), CIS 6.2 (Establish an Access Revoking Process), CIS 5.3 (Disable Dormant Accounts)

Compensating: Revoke OAuth grants in bulk via Microsoft Graph: Get-MgOAuth2PermissionGrant -All | Where-Object {$_.ClientId -in $unapprovedAppIds} | ForEach-Object { Remove-MgOAuth2PermissionGrant -OAuth2PermissionGrantId $_.Id }. For Okta, use the Okta API: DELETE /api/v1/apps/{appId}/grants/{grantId} for each unapproved AI application. For externally held API keys (OpenAI, Anthropic platform keys), instruct users to rotate or delete keys directly in vendor dashboards and confirm via a re-scan of code repositories using git-secrets or truffleHog

(free) to detect any hardcoded API keys committed to internal repos: trufflehog git file://./repo --only-verified. For agentic service accounts, apply the minimum scope by removing wildcard permissions and replacing with resource-specific delegated permissions only.

Evidence: Before revoking, export the full OAuth grant manifest for each unapproved NHI — record the application ID, granted scopes, resource owner, and consent timestamp as a forensic record of the exposure window. Capture any pipeline configuration files (GitHub Actions .yml, Azure DevOps pipeline definitions, Airflow DAGs) that reference AI API endpoints or store credentials as environment variables, as these represent persistence mechanisms for shadow AI integrations that will survive OAuth revocation unless the pipeline itself is modified.

Step 4: Recovery — Validate that approved AI integrations operate under documented data handling agreements and that NHIs have been inventoried per CIS 1.1 (Establish and Maintain Detailed Enterprise Asset Inventory) and CIS 5.1 (Establish and Maintain an Inventory of Accounts). Confirm DLP rules cover API egress paths and that CASB policies flag new AI tool adoption in real time. Re-audit OAuth grant inventory monthly per CIS 6.2 (Establish an Access Revoking Process).

NIST Phase: Recovery

Reference: NIST 800-61r3 §3.5 — Recovery

Controls: NIST AC-20 (Use of External Systems), NIST AU-11 (Audit Record Retention), CIS 1.1 (Establish and Maintain Detailed Enterprise Asset Inventory), CIS 5.1 (Establish and Maintain an Inventory of Accounts), CIS 6.2 (Establish an Access Revoking Process), CIS 3.2 (Establish and Maintain a Data Inventory)

Compensating: Maintain the NHI inventory as a version-controlled CSV or YAML file in a private Git repository, with fields for: service principal name, application ID, granted scopes, owning team, data classification of resources accessed, and review date. Schedule a monthly PowerShell or Graph API job to diff the current OAuth grant list against the approved inventory and alert on new entries: Compare-Object (Import-Csv approved_nhis.csv | Select AppId) (Get-MgOAuth2PermissionGrant -All | Select ClientId). For DLP coverage of API egress, configure HTTPS inspection on your proxy for LLM API endpoints and create a Sigma-compatible log alert that fires when POST request body size to api.openai.com or equivalents exceeds a defined threshold (e.g., 50KB), which may indicate bulk data exfiltration via prompt context.

Evidence: Validate recovery completeness by re-running the same OAuth grant query from Step 2 and confirming the delta shows only approved application IDs remain. Retain a post-remediation snapshot of proxy logs showing traffic to LLM API endpoints is reduced to only approved sources — this establishes a clean baseline for future anomaly detection. Confirm that CASB or proxy logs now show HTTP 403 or DNS NXDOMAIN responses for blocked AI tool domains that were enumerated in Step 1, proving the firewall/proxy policy is enforced and not bypassed.

Step 5: Post-Incident — Establish a formal AI governance policy addressing Shadow AI: required security review before AI tool adoption, data classification requirements for AI-connected systems, NHI lifecycle management, and prompt injection testing requirements for LLM-integrated pipelines. Map controls to NIST AC-6, AC-20, AU-2 (Event Logging), AU-6 (Audit Record Review), and SI-4 (System Monitoring). The core control gap this risk exposes is the absence of a defined intake and authorization process for AI tools — treat AI adoption like third-party software procurement, not self-service SaaS.

NIST Phase: Post Incident

Reference: NIST 800-61r3 §4 — Post-Incident Activity

Controls: NIST AC-6 (Least Privilege), NIST AC-20 (Use of External Systems), NIST AU-2 (Event Logging), NIST AU-6 (Audit Record Review, Analysis, and Reporting), NIST AC-1 (Policy and Procedures), CIS 7.1 (Establish and Maintain a Vulnerability Management Process), CIS 7.2 (Establish and Maintain a Remediation Process), CIS 4.6 (Securely Manage Enterprise Assets and Software)

Compensating: For prompt injection testing of internal LLM-integrated pipelines without commercial tooling, use the open-source Garak framework (pip install garak) to run adversarial probe suites against any internally deployed or API-connected LLM endpoint — document test results as part of the AI tool intake security review artifact. For ongoing NHI lifecycle monitoring without a PAM solution, implement a scheduled osquery query against workstations to detect new credential files in user home directories: SELECT * FROM file WHERE path LIKE '/home/%/.config/openai%' OR path LIKE 'C:\Users\%\AppData\Roaming\openai%'; and alert on new findings. Develop an AI Tool Intake Checklist as

a Markdown template in your internal wiki covering: vendor data processing agreement status, data classification of inputs, NHI scope justification, egress monitoring confirmation, and annual review date.

Evidence: As the primary lessons-learned artifact, preserve the complete timeline of OAuth grant creation dates for all unapproved NHIs discovered — this establishes the governance gap duration and informs the scope of any data handling agreement retrospective. Retain proxy log evidence of which internal user accounts or systems generated traffic to unapproved LLM APIs, categorized by data sensitivity of the source system, as this is the core input for any breach notification assessment if sensitive data classes were transmitted to unvetted third-party AI APIs.

Detection Guidance

Detection requires visibility across four planes that traditional tooling often misses for AI workloads. (1) Identity plane, query IdP logs for OAuth application grants created in the past 90 days; filter for scopes including mail.read, files.readwrite, or data access permissions granted to AI vendor application IDs. Flag any NHI (service account, API key, bot identity) created outside a change management ticket. Relevant NIST control: AU-2, AU-6, AC-3. (2) Network/proxy plane, extract DNS and proxy logs for hostnames matching known LLM API providers; correlate against approved application inventory. Large or frequent POST requests to /v1/chat/completions, /v1/messages, or equivalent LLM inference endpoints from endpoints not running approved AI clients require investigation. MITRE T1048, T1567. (3) Data egress plane, CASB and DLP rules should flag uploads of structured data (CSV, JSON, database exports) to AI API endpoints. CWE-359 and CWE-532 failure modes often appear as log files or debug outputs sent to external APIs containing PII or credentials. MITRE T1530, T1552.001. (4) Agentic workflow plane, monitor for automated sequences: a service account authenticating, querying internal APIs, then making outbound calls in rapid succession without human-initiated sessions. This pattern maps to T1078.004 (Cloud Accounts) and T1199 (Trusted Relationship). CIS 8.2 coverage must include cloud audit logs, not just on-premises sources. Cloud account monitoring and configuration review apply to endpoint-side detection of AI tool installation and configuration artifacts.

Framework Mappings

MITRE-ATTACK

- **T1190** — Exploit Public-Facing Application
- **T1059** — Command and Scripting Interpreter
- **T1552** — Unsecured Credentials
- **T1567** — Exfiltration Over Web Service
- **T1199** — Trusted Relationship
- **T1078** — Valid Accounts
- **T1552.001** — Credentials In Files
- **T1048** — Exfiltration Over Alternative Protocol
- **T1078.004** — Cloud Accounts
- **T1530** — Data from Cloud Storage

NIST-800-53R5

- **CA-8** — Penetration Testing
- **RA-5** — Vulnerability Monitoring and Scanning

- **SC-7** — Boundary Protection
- **SI-2** — Flaw Remediation
- **SI-7** — Software, Firmware, and Information Integrity
- **CM-7** — Least Functionality
- **SI-3** — Malicious Code Protection
- **SI-4** — System Monitoring
- **AC-2** — Account Management
- **AC-6** — Least Privilege
- **IA-2** — Identification and Authentication (Organizational Users)
- **IA-5** — Authenticator Management
- **AC-3** — Access Enforcement
- **SR-2** — Supply Chain Risk Management Plan

OWASP-TOP10-2021

- **A01:2021** — Broken Access Control

CIS-V8

- **6.1** — Establish an Access Granting Process
- **6.2** — Establish an Access Revoking Process
- **15.1** — Establish and Maintain an Inventory of Service Providers

SOC2-TSC

- **CC6.1** — The entity implements logical access security software, infrastructure, and architectures over protected information assets
- **CC9.2** — Manages risks associated with vendors and business partners

HIPAA-SECURITY

- **164.312(a)(1)** — Access Control

ISO-27001-2022

- **A.8.8** — Management of technical vulnerabilities
- **A.5.21** — Managing information security in the ICT supply chain

NIST-CSF-2

- **GV.SC-01** — Cybersecurity supply chain risk management program

MITRE ATT&CK Mapping

Technique ID	Technique Name	Tactic
T1190	Exploit Public-Facing Application	Initial-Access
T1059	Command and Scripting Interpreter	Execution

Technique ID	Technique Name	Tactic
T1552	Unsecured Credentials	Credential-Access
T1567	Exfiltration Over Web Service	Exfiltration
T1199	Trusted Relationship	Initial-Access
T1078	Valid Accounts	Defense-Evasion
T1552.001	Credentials In Files	Credential-Access
T1048	Exfiltration Over Alternative Protocol	Exfiltration
T1078.004	Cloud Accounts	Defense-Evasion
T1530	Data from Cloud Storage	Collection

Sources

Source	URL	Tier
Blog	https://www.crowdstrike.com/en-us/blog/shadow-ai-hidden-risk-expand...	T3
	http://news.lenovo.com/pressroom/press-releases/70-enterprise-ai-un...	T3
	https://www.varindia.com/news/mythos-dpdpa-are-enterprises-ready-fo...	T3
	https://thehackernews.com/2026/04/the-hidden-security-risks-of-shad...	T3
CrowdStrike Falcon AIDR: AI Detection & Response	https://www.crowdstrike.com/en-us/platform/falcon-aidr-ai-detection...	T3

Generated 2026-06-01 18:43 UTC by TJS Security Command Center