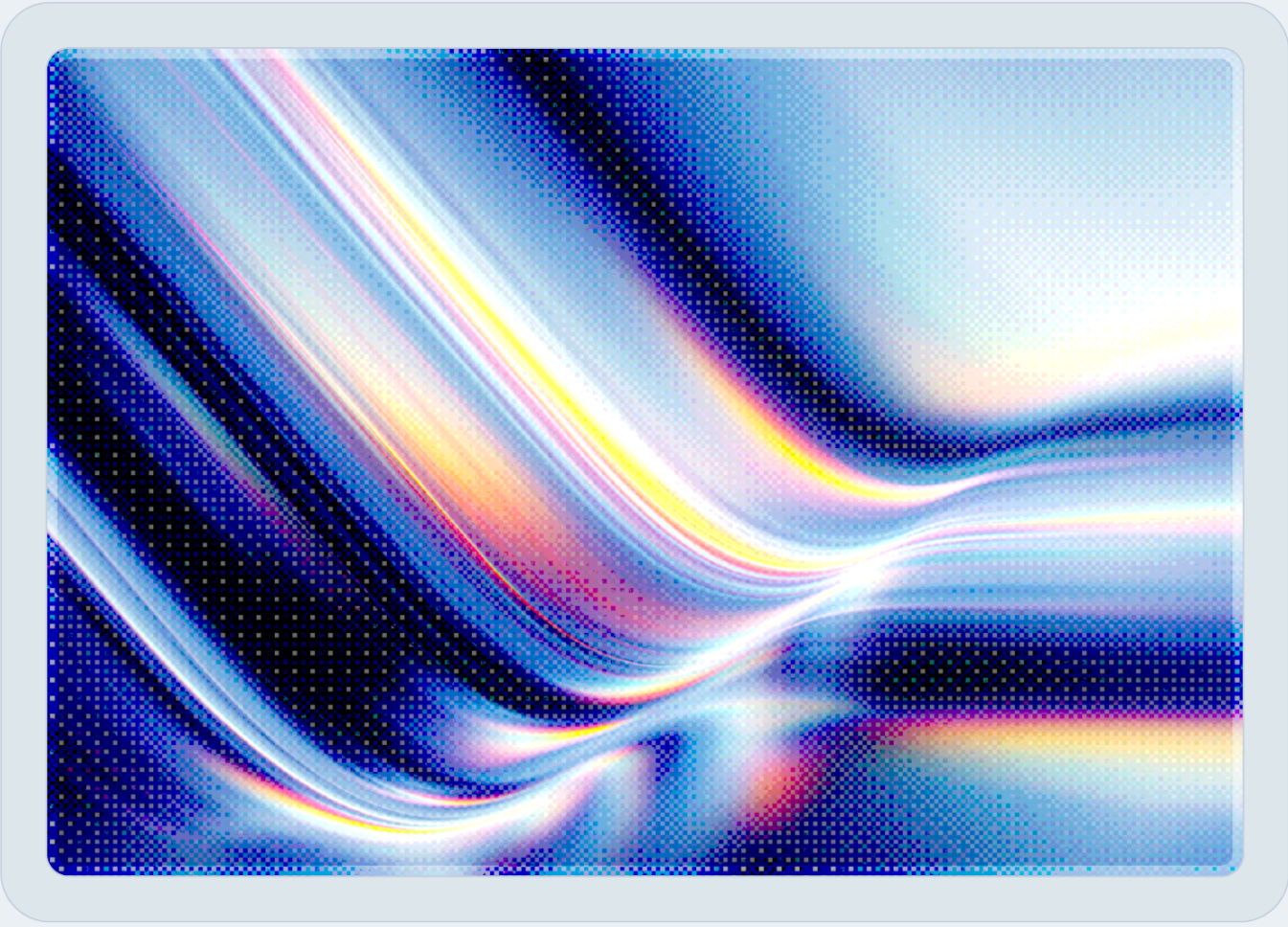
CSAI Foundation | Cloud Security Alliance

OAuth Ghost Tokens: Enterprise AI Integration Supply Chain Risk

Structural Lessons from the Context.ai–Vercel Double Supply Chain Attack

2026-05-05

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

- The April 2026 Context.ai–Vercel incident demonstrates a "double supply chain" attack pattern: compromise a small AI SaaS vendor, extract the OAuth tokens it holds on behalf of enterprise customers, and use those tokens to walk directly into downstream organizations.
- OAuth ghost tokens — dormant authorizations granted to third-party AI tools by employees acting outside IT oversight — persist indefinitely after the original authorization intent has passed, creating an invisible, durable attack surface.
- Research indicates enterprises average 17 unique AI application integrations per organization in Microsoft 365 and Google Workspace, the majority of which have not been reviewed through formal security or procurement channels [1].
- Non-human identities now outnumber human identities by as many as 144 to 1 in some enterprise environments, yet identity governance programs remain oriented around human login-and-logout patterns [2].
- Reversing this exposure requires treating OAuth token lifecycle management as a first-class identity discipline, not an afterthought to endpoint or network security.

Background

On April 19, 2026, Vercel, a widely deployed cloud platform-as-a-service used by development teams globally, disclosed that an unauthorized actor had accessed internal systems and a limited set of non-sensitive customer environment variables [3]. The investigation traced the intrusion not to a direct compromise of Vercel's own infrastructure but to a third-party AI productivity tool called Context.ai, whose OAuth tokens had been compromised months earlier.

The initial breach of Context.ai began in approximately February 2026, when a company employee was infected with Lumma Stealer malware — reportedly introduced through a download of Roblox game exploit scripts on a personal or lightly managed device [4]. The infostealer harvested a broad set of credentials from that device, including Google Workspace credentials and access keys for Context.ai's cloud infrastructure, including its Supabase database and AWS environment. By March 2026, the

attacker had gained unauthorized access to Context.ai's AWS environment and exfiltrated OAuth tokens stored there — tokens issued by enterprise users who had previously authorized Context.ai's AI Office Suite, a consumer-oriented productivity product the company had launched in June 2025 [4].

The pivot to Vercel followed from a straightforward fact: at least one Vercel employee had signed up for the Context.ai AI Office Suite using their enterprise Google Workspace credentials and had granted the application broad OAuth permissions, including access to Google Drive and email [5]. That OAuth token, once exfiltrated from Context.ai's storage, gave the attacker the ability to impersonate the Vercel employee within Google Workspace. From that foothold, the attacker navigated to Vercel's internal dashboards and systems, ultimately accessing non-sensitive plaintext environment variables for a limited subset of customer projects [3]. Sensitive encrypted variables were not accessed [3]. A threat actor claiming to be ShinyHunters later advertised the stolen database access credentials on BreachForums for $2 million, though security researchers assessed the claim of group identity as likely fabricated to inflate notoriety [4].

Context.ai had already deprecated the AI Office Suite product by the time the breach occurred, and Vercel itself was not a registered customer of the product [1]. The OAuth authorization had been established informally, by one employee, through self-service signup — a pattern that characterizes the broader shadow AI adoption dynamic reshaping enterprise attack surfaces in 2026.

Security Analysis

The Ghost Token Problem: Authorization That Outlives Intent

At the center of this incident is what has been termed "authorization drift" — the condition in which machine credentials and delegated access grants outlive the operational purpose for which they were issued [2]. When an employee grants a third-party AI application access to their corporate Google Workspace or Microsoft 365 account, that OAuth authorization does not expire when the employee stops using the tool, when the vendor deprecates the product, or even when the employee leaves the organization. OWASP's Non-Human Identities research, synthesized by CSA [2], indicates that credentials remain active an average of 47 days after they are no longer operationally necessary, and in practice, OAuth grants to AI tools often persist for months or years with no active monitoring or revocation.

These dormant, forgotten authorizations are what this report terms OAuth ghost tokens: active access grants that are invisible to security teams, unmonitored by identity governance programs, and indistinguishable from legitimate use when an attacker deploys them. Ghost tokens do not trigger multi-

factor authentication challenges — they represent already-approved delegated trust. They do not appear in traditional endpoint detection telemetry, because no credentials are being entered or phished. They simply exist, quietly, in the OAuth authorization tables of enterprise identity providers, waiting.

The structural danger is that ghost tokens inherit the permission scope of the authorizing user at the time of authorization. In the Vercel incident, the token issued to Context.ai reflected the full scope of the Vercel employee's account access [5], which apparently included access to internal dashboards, API keys, NPM package management tokens, and GitHub credentials. An OAuth grant that appeared to be for an AI productivity tool thus became a master key to significant portions of Vercel's development infrastructure. This scope inheritance is not a misconfiguration in the traditional sense — it is the predictable consequence of OAuth's delegation model applied to users with broad, legitimate access to enterprise systems.

Double Supply Chain: Compromising Vendors to Reach Enterprises

The Context.ai–Vercel incident belongs to an emerging class of attacks that exploit the trust architecture of the modern SaaS ecosystem rather than attacking target organizations directly. In this pattern, the attacker targets a smaller AI or SaaS vendor — one whose security posture may be commensurate with its headcount and security budget rather than the mature controls of a large enterprise — that holds OAuth tokens issued by enterprise users of its products. Once that vendor is compromised, those tokens provide authenticated access to hundreds or thousands of downstream enterprise environments, bypassing perimeter, endpoint, and credential-based security controls entirely, since no credentials are being stolen or phished.

This model closely parallels the 2025 Scattered Lapsus$ Hunters campaign, in which attackers compromised the Salesloft Drift marketing automation platform and leveraged OAuth tokens stored there to access more than 700 downstream organizations — including Google, Cloudflare, Rubrik, and Elastic [6]. The pattern is repeatable precisely because the economics favor the attacker: a single compromise of a mid-sized AI SaaS vendor can yield OAuth tokens for enterprise customers that would each individually represent a far more challenging target to breach directly.

The "double" aspect of the term is important. Context.ai's own customers were exposed in the initial breach; Vercel's customers were then exposed as a second-order consequence of Vercel's internal exposure. Each layer of OAuth delegation adds a tier of potential blast radius to a successful compromise upstream in the chain.

Shadow AI and the Procurement Bypass

The Context.ai AI Office Suite was not a Vercel-approved, procurement-reviewed vendor. The authorization that enabled the attack was established unilaterally by one employee through a self-service signup, using a corporate identity, without IT review or security assessment [1]. This is not anomalous behavior — it is the norm. Reco AI's 2025 State of Shadow AI Report documented that enterprises now manage an average of 490 SaaS applications per organization, with fewer than half procured through authorized channels — a scale that makes comprehensive OAuth grant visibility a significant operational challenge [7]. Formal AI governance policies specifically addressing third-party tool adoption remain far less common than the pace of AI tool adoption itself would warrant.

The result is an OAuth sprawl problem that is structurally difficult to observe. When employees connect AI tools to their enterprise identities without disclosure, the resulting OAuth grants do not appear in procurement records, vendor risk assessments, or access reviews. Push Security research found that enterprises average 17 unique AI application integrations per organization in Microsoft and Google environments [1], and the CSA AI Agent Governance survey (January 2026, N=418) found that 82% of organizations discovered previously unknown AI agents operating in their environments over the preceding year [8].

These figures suggest that most enterprise environments currently contain OAuth ghost tokens for AI applications that security teams have never reviewed, cannot enumerate without dedicated tooling, and do not include in threat modeling. The attack surface represented by these tokens is real, already established, and growing with every new AI tool adoption.

Scope Inflation and the Blast Radius of "Allow All"

A persistent technical vulnerability compounds these structural issues: OAuth scope grants to AI tools are routinely broader than the minimal access those tools require. When an employee grants an AI application the full `drive` scope — as many productivity AI tools request — that grant provides read and write access to everything the user can access in Drive, including shared team drives, collaborative documents, and files shared from other divisions [1]. Applications requesting narrower scopes such as `drive.readonly` or `drive.file` present a meaningfully reduced blast radius, which is why scope review is a concrete remediation step. When the granted scope includes Gmail or calendar access, the same principle applies across even more sensitive data surfaces. The attacker who inherits that token inherits the entire scope.

Enterprise identity platforms provide administrative controls to restrict OAuth consent — Google Workspace administrators can require admin approval for third-party OAuth grants, and Microsoft 365 offers equivalent tenant-level policies. However, applying these controls retroactively requires

identifying and auditing grants that may have been established years earlier by employees who have long since changed roles or departed. Organizations that have not actively managed OAuth consent policies face the task of auditing an inherited, largely unknown set of access relationships before they can enforce meaningful restrictions going forward.

Recommendations

Immediate Actions

Organizations should treat the Context.ai–Vercel incident as a prompt for an urgent OAuth inventory exercise. The first step is enumerating all third-party OAuth grants in Google Workspace and Microsoft 365 tenants. Both platforms expose this data through native admin consoles: Google Workspace administrators can access the list of authorized third-party apps at the admin console under Security > Access and data controls > API controls; Microsoft 365 administrators can enumerate OAuth grants through the Entra ID portal under Enterprise applications. Organizations that have deployed the Context.ai AI Office Suite should revoke the OAuth client identified in the incident (client ID: `110671459871-30f1spbu0hptbs60cb4vsmv79i7bbvqj.apps.googleusercontent.com`) [5] and remove the associated Chrome extension (extension ID: `omddlmnhcofjbnbflmjginpjjblphbgk`, as corroborated in secondary incident reporting) — organizations should confirm both identifiers against official incident notifications from Context.ai or Vercel [3] before revoking.

Any credentials — API keys, tokens, database credentials — that may have been accessible to the compromised employee accounts identified in incident notifications should be rotated regardless of whether specific evidence of access exists. The attacker's demonstrated willingness to exploit even indirect OAuth access — navigating Vercel's internal API surfaces with a token that required no further credential bypass [3] — warrants a conservative posture on the scope of what should be treated as potentially exposed.

Short-Term Mitigations

Security teams should implement default-deny OAuth consent policies as a near-term priority. In Google Workspace, this involves setting the "Allow users to authorize apps" control to require admin review for applications that are not on an approved list. In Microsoft 365, this translates to configuring user consent

settings in Entra ID to disallow user consent for applications or to require admin approval for applications requesting permissions beyond a baseline set. These controls break the self-service pathway by which employees can silently establish new OAuth trust relationships without organizational review.

Alongside technical controls, organizations should establish a lightweight employee AI tool disclosure process. Employees are adopting AI tools rapidly, and prohibition without an accessible alternative path creates pressure toward covert adoption. A low-friction process that allows employees to request evaluation of new AI tools — with a defined review timeline and a standing approved list — provides both a governance mechanism and a cultural signal that shadow adoption is not necessary.

Existing OAuth grants should be reviewed for overly broad scopes. Not all third-party AI applications require the full scope they request; many will function adequately with read-only access to specific resources rather than full read-write access across an entire drive or mailbox. Where scope reduction is not possible within a vendor's authorization model, the risk posture of that grant should be explicitly documented and monitored.

Strategic Considerations

The fundamental issue exposed by ghost token attacks is that OAuth authorization *grants* — the long-lived consents that allow applications to obtain new tokens indefinitely — are issued once and almost never revisited. Unlike access tokens, which platforms expire automatically, authorization grants persist until an administrator or user explicitly revokes them. Enterprise environments have built no equivalent of access review or expiration for this layer of the identity stack. Addressing this structurally requires treating OAuth token lifecycle as a formal identity governance discipline, with the same rigor applied to human account access reviews.

This means, at minimum, adding OAuth grants to the scope of periodic access reviews (typically quarterly or semi-annually) and establishing defined criteria for revocation: employee departure, role change, vendor deprecation, or the lapse of a specified period without authenticated use. Non-human identities — AI agent credentials, OAuth grants, API keys, service account tokens — should be inventoried and managed under the same governance frameworks as human identities, rather than being treated as implementation details owned by individual application teams.

Procurement processes should also be updated to account for the distinct risk profile of AI tools that request OAuth access to enterprise identity systems. A tool that requests read-only access to a specific project folder presents a different risk posture than one that requests access to an employee's full mailbox, calendar, and Google Drive. Third-party risk assessments for AI tools should include review of the OAuth scopes requested, the security practices of the vendor with respect to token storage and access control, and the existence of a public breach notification and incident response history.

Finally, organizations should evaluate SaaS security posture management (SSPM) tooling that provides continuous visibility into OAuth grant inventories across the enterprise's SaaS footprint. Native admin console visibility into OAuth grants is a start, but it requires manual effort and lacks alerting capabilities. SSPM platforms purpose-built for OAuth visibility can provide continuous monitoring, alerting on new high-risk grants, and automated decommissioning workflows for deprecated integrations — capabilities that make the ghost token problem tractable at enterprise scale. Organizations evaluating SSPM vendors should note that much of the market research data cited in this space originates from vendors with a direct commercial interest in the category; independent validation against internal OAuth audits is advisable before drawing conclusions about organizational exposure scale.

CSA Resource Alignment

The security challenges illustrated by the Context.ai–Vercel incident map directly to several CSA frameworks and publications that provide structured guidance for enterprise response.

MAESTRO (Multi-Agent Environment, Security, Threat, Risk, and Outcome) provides a seven-layer threat modeling framework for agentic AI systems [9]. The ghost token attack surface described in this note is particularly relevant to MAESTRO's Agent Ecosystem layer (L7), which addresses trust relationships between agents and external systems, and the Deployment and Infrastructure layer (L4), which covers the runtime environments in which agent credentials and OAuth tokens are stored and managed. Organizations applying MAESTRO to their AI agent deployments should explicitly include OAuth grant inventories and token lifecycle risks in their L4 and L7 threat models.

The **AI Controls Matrix (AICM)** and its associated implementation guidelines for Application Providers and Orchestrated Service Providers directly address third-party access control, supply chain risk, and non-human identity management. The AICM's identity and access management control domains provide a structured basis for formalizing OAuth token governance policies and third-party AI vendor assessment criteria.

AI Organizational Responsibilities: AI Tools and Applications addresses the supply chain and third-party risk dimensions of AI tool adoption in depth, including guidance on vendor assessment processes, software bill of materials for AI systems, and the distinct responsibilities that organizations bear when they deploy AI tools that integrate with corporate identity systems. The framework's treatment of shadow IT risk in AI adoption is directly applicable to the OAuth sprawl problem described in this note.

The CSA **AI Agent Governance survey (2026)** [8] provides empirical benchmarking data on the prevalence of shadow AI agent discovery, the maturity of agent lifecycle governance practices, and the frequency of AI agent security incidents — data that organizations can use to calibrate the urgency of their response to OAuth ghost token exposure.

Finally, CSA's **Zero Trust guidance** is foundational to the architectural posture that addresses ghost token risk at its root: the principle that no identity — human or non-human, internal or delegated — should be granted persistent implicit trust. Zero Trust applied to OAuth means treating every authorization grant as requiring ongoing verification of continued legitimate need, not as a one-time decision that remains valid indefinitely.

References

[1] Push Security. "Unpacking the Vercel Breach: Shadow AI and OAuth Sprawl." Push Security Blog, April 2026.

[2] Cloud Security Alliance. "AI Security: When Authorization Outlives Intent." CSA Blog, February 25, 2026.

[3] Vercel. "Vercel April 2026 Security Incident." Vercel Knowledge Base, April 2026.

[4] The Hacker News. "Vercel Breach Tied to Context AI Hack Exposes Limited Customer Credentials." The Hacker News, April 2026.

[5] Wiz. "Context.ai OAuth Token Compromise." Wiz Blog, April 2026.

[6] Reco AI. "AI & Cloud Security Breaches: 2025 Year in Review." Reco AI Blog, 2026.

[7] Reco AI. "2025 State of Shadow AI Report." Reco AI, 2025.

[8] Cloud Security Alliance. "Autonomous but Not Controlled: AI Agent Incidents Now Common in Enterprises." CSA Research, April 21, 2026.

[9] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA Blog, February 6, 2025.

