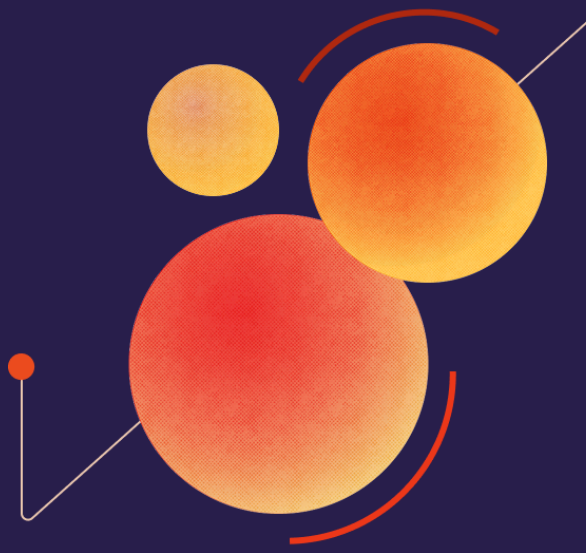


A $300B Investment Firm *Secures Claude Access with Aembit*

Aembit is the identity control plane securing this firm's Claude deployment across all connected MCP servers — full identity, access control, and audit logging in place for every employee, with financial analysts and executives as the first critical use cases.

aembit

Gartner COOL VENDOR 2022

CSO Cybersecurity startups to watch for in 2023

2024 WINNER CYBER SECURITY EXCELLENCE AWARDS

RSAC Innovation Sandbox 2024 FINALIST

2024 SINET16 INNOVATOR AWARD

Industry	Financial Services – Investment Management.
Company Size	700+ employees.
Region	North America.
Regulatory Environment	SOX, PCI DSS; rigorous internal security standards.
Datasets	Microsoft 365 MCP server, Factset, Kensho, Atlassian, Chronograph.
AI stack	Claude (claude.ai web), Okta (human IdP), Enterprise MCP servers, CrowdStrike AIDR, Aembit.
Use Case	Enterprise Agentic AI deployment.
Challenge	No authenticated agent identity, no audit trail distinguishing and attributing agent actions versus human actions across MCP-connected services, no centralized management of agent access and identity, agent access is too persistent and too broad.
Solution	Aembit IAM for Agentic AI.
Deployment Timeline	Two weeks.
Key Results	500+ users 100% of agent actions logged and auditable 100% of enterprise AI agents secured by Aembit.

Why the Customer Needed Aembit

This organization is using Claude to deploy personalized AI assistants across its entire workforce, giving agents access to financial datasets for research, email and calendars for productivity, and sensitive company information through SharePoint and Microsoft 365.

The security team identified three critical gaps their regulated environment couldn't accept:

No agent identity, only user identity

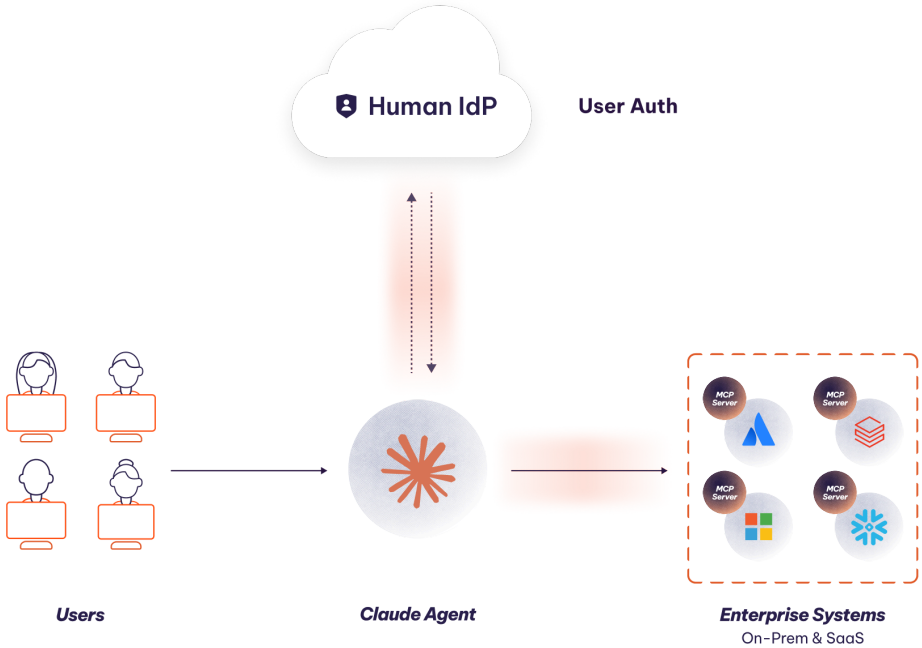
When Claude accessed an MCP server on an analyst's behalf, it did so using the analyst's full identity and access rights — with no distinct agent identity, no audit attribution, and no coverage for services outside Okta's governance.

Long-lived credentials stored in the wrong places

MCP authentication required storing secrets, tokens, and API keys directly in Claude or in MCP server configurations — static, unrotated, and sitting in places never designed to be credential stores.

No centralized access control or auditability

Policy enforcement for agent behavior was piecemeal, based on human identity, and lacked visibility into whether the human or agent took action.



What Aembit Does

Blended human-agent identity

Aembit creates a unique blended identity based on the user and the agent enabling a unified framework and single, consistent security policy for managing both human and non-human identities.

Run-time policy enforcement

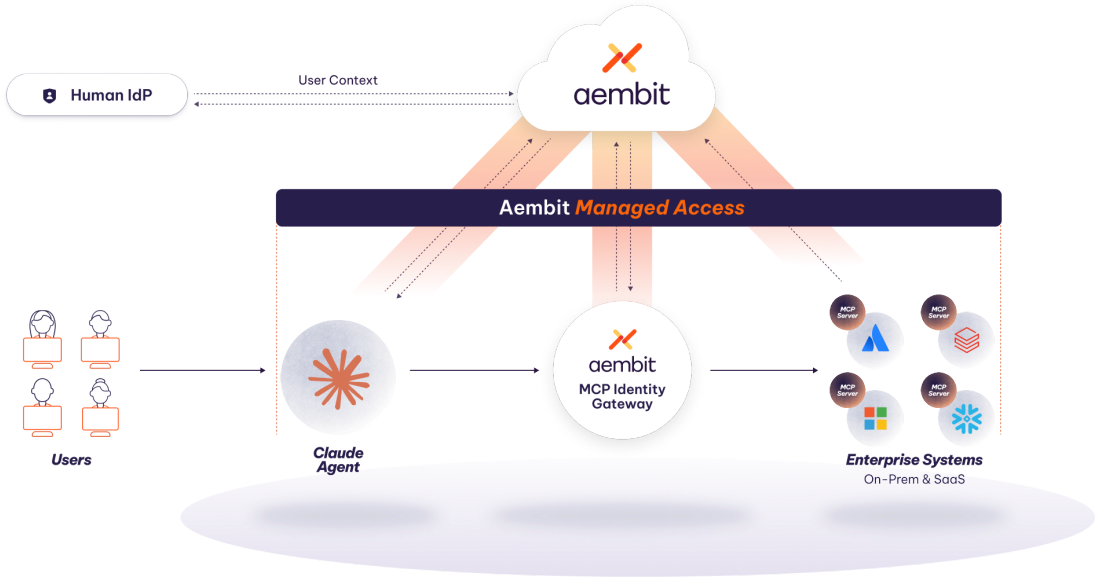
Granular least-privilege access enforced per agent, per MCP server, within the firm's existing Okta identity foundation.

Just-in-time, secretless access

Long-lived stored credentials eliminated; replaced with short-lived, policy-scoped tokens issued per session. No static secrets.

Auditability based on detailed access logs

Every agent action is logged and attributed with full access logs forwarded to CrowdStrike SIEM, giving the security team the same visibility into agent behavior as human user activity.



How It Works

Aembit is deployed as a SaaS service. Its cloud control plane offers policy enforcement and logging. Its **MCP Authorization Service** handles standards-compliant OAuth with PKCE and issues blended identity tokens, and an **MCP Gateway** that enforces access policy with token exchange in real time and logs every tool invocation. Human identity is anchored to the firm's existing Okta instance. Deployment takes two weeks, with less than six hours of security team time required to stand up the solution.

The Challenge

This organization manages pension fund investments on behalf of hundreds of thousands of beneficiaries. The firm's AI innovation team knew AI was going to influence every facet of their business; eventually everyone would have a personal assistant analyzing data, building business cases, and even doing day-to-day work like composing emails for them. For AI agents to be effective, they needed access to the full Microsoft Graph across the organization: executives' emails, calendars, SharePoint documents, and every other data surface baked into decades of organizational history. The security team's position was straightforward: deploying AI agents across the entire analyst and executive workforce without authenticated agent identity, without access policy, and without audit logging represented a significant risk.

The organization had spent a decade learning through hard experience what enterprise security controls needed to look like for human users. Retrofitting those controls after the fact, across an AI deployment that had already scaled, would be significantly harder than building them in from the start. Internal pressure from business stakeholders made it clear that AI adoption would proceed, with or without security guardrails in place. The security team's response could not be to shut it down. It was to find a way to say yes – safely.

Organizations deploying AI are finding that AI adoption moves faster than security teams can track, with the blast radius of each new integration larger than the last. It is the defining characteristic of enterprise agentic AI deployments right now. AI access management is a company-wide security issue, and authentication, access control, audit logging, and least-privilege access aren't bureaucratic checkboxes. They are the foundation that makes everything else defensible – and they are the central benefits that Aembit's solution is built to deliver.

> **AI access management is a company-wide security issue, and authentication, access control, audit logging, and least-privilege access aren’t bureaucratic checkboxes.**

The Status Quo

To understand the gaps this security team needed to fill to deploy AI safely, it helps to understand how MCP (Model Context Protocol) authentication actually works. The MCP was designed to make it fast to connect AI agents to tools and data sources and was adopted extremely rapidly, with no formal standards process. But in the interest of speed and ease of adoption, early MCP implementations discarded or bypassed many of the security controls that had been established for HTTP and OAuth over the preceding decade.

The first generation of MCP servers was entirely unauthenticated; any client could connect. Subsequent versions added API key authentication, but API keys don't associate requests with individual users. When OAuth was introduced, implementations frequently omitted critical security features that the broader security community had come to treat as non-negotiable, like PKCE (Proof Key for Code Exchange) – a mechanism that prevents authorization code interception attacks – and constrained dynamic client registration, which requires OAuth clients to be explicitly onboarded by an administrator rather than self-registering without credentials.

The Risks

As a result, when AI agents like Claude connect to an MCP server using the patterns most MCP implementations support, it does so in ways that would trigger an immediate security review if a human-built application had proposed the same approach.

The specific failure modes the firm's security team identified fell into three categories:

No agent identity, only user identity

When Claude accessed an MCP server on a user's behalf, it did so using the user's identity and full access rights – not a distinct agent identity. There was no way to distinguish in an audit log whether the user had directly accessed a resource or whether Claude had done it autonomously. For a regulated financial institution, attribution is not optional: the firm needed to know not just what was accessed, but who or what accessed it. Not only was this approach unable to distinguish between the agent and the user, it didn't extend to services that Okta, their human IdP, didn't govern access to.

Long-lived credentials stored in the wrong places

The standard approach to MCP authentication involves storing credentials (secrets, tokens, API keys) directly in Claude or in the MCP server configuration. Long-lived secrets with no rotation policy, sitting in places that were not designed to be credential stores was a non-starter in a regulated environment with strict controls over how secrets are managed.

No centralized access control or auditability

There was no mechanism to say "this user's agent can access Microsoft365 but not Atlassian." Access was dependent on the user and policies were applied inconsistently and required piecemeal updates across the enterprise. For a firm deploying agents across the entire analyst workforce, with access to sensitive financial data and internal Microsoft infrastructure, a misconfigured or compromised agent could be a disaster.

The firm's security team documented these risks explicitly, producing a detailed mapping of eleven specific risk categories they identified in MCP-based agentic AI deployments. Aembit was able to mitigate or eliminate the risks of ten of those, listed below:

Risk	How Aembit Helps
Token Theft and Credential Leakage	Aembit **eliminates static credentials entirely**. Agents and workloads receive short-lived, ephemeral tokens generated at runtime — never stored, never exposed to users or AI agents — so there are no long-lived secrets to steal or leak.
Excessive Permissions and Over-Scoping	Credential Providers **issue minimally scoped tokens** bound to specific target services. Access scope is enforced by policy at the point of issuance instead of being left to individual developers or agents to configure correctly.
Unverified Third-Party Endpoints (Server Spoofing)	Aembit's **centralized control plane** governs which MCP servers are permitted. No user or agent can introduce an unverified endpoint outside of that policy boundary, reducing the attack surface for spoofed or rogue servers.
Confused Deputy Attacks in OAuth Proxy Flows	Aembit **enforces PKCE** on all OAuth flows and binds Credential Providers to specific server workloads with explicit scope configuration. MCP servers that don't meet OAuth 2.1 minimums cannot authenticate. Misconfigured consent flows are blocked by design.
Prompt Injection and Tool Poisoning	Aembit **integrates with third-party AI security tooling** to detect and block injection attempts at the layer where credentials and tool access are brokered, adding a security control point that sits between the agent and the resource.
Command Execution and Code Injection	Aembit enables defense-in-depth around the credential and access layer through **integrations with AI security tools**, pairing identity controls with runtime threat detection to contain execution-based attacks.

Replay Attacks and Session Hijacking	**Short-lived ephemeral tokens** drastically shrink the window for replay exploitation. Aembit's **audit logs** provide timestamped, attributable records of token issuance and usage, enabling rapid detection of anomalous refresh token activity.
Inadequate Observability and Auditing	Aembit provides a **centralized audit trail** that attributes every access event to both the human user and the AI agent involved, closing the observability gap that arises when non-human identities act on behalf of humans. **Logs export directly to SIEM** platforms.
Token Passthrough Anti-Pattern	When deployed between an MCP server and a resource server, Aembit **intercepts and exchanges credentials at runtime** rather than allowing raw tokens to pass through.
Malicious or Compromised MCP Servers	Aembit's roadmap includes granular **MCP Tool Filtering** to restrict which tools an MCP server can expose, limiting the blast radius of a compromised or malicious server and reducing exposure to tool poisoning attacks.

The Solution: Aembit IAM for Agentic AI

Human identity provider solutions answer the question: “who is this user?” Identity and access management solutions for workloads answer the question: “what is this software?” But AI agents have unique needs: an underprivileged agent cannot work effectively for a portfolio analyst, while an overprivileged one could access far more of the firm's financial data and M365 infrastructure than any single workflow requires.

To handle user-driven AI agents, Aembit answers both questions, giving the agents the autonomy they need to work on behalf of a user while clearly attributing and logging the action to the agent.

Blended Identity and Secretless Access

To do this, Aembit creates a “blended identity”: when Claude makes a request through the Aembit-secured path, it is issued an identity credential that encodes both the employee's Okta identity and the agent's identity. This blended token makes it possible to answer definitively whether a given action was taken by the analyst directly or by Claude acting on their behalf, providing a unified framework for managing both human and non-human identities under a single, consistent security policy.

At the heart of this is Aembit's standards-compliant OAuth implementation, deployed through its cloud-hosted Authorization Service. Where the MCP OAuth implementations used by the firm's target servers cut corners — omitting PKCE, permitting unconstrained dynamic client registration, relying on long-lived static secrets — Aembit enforces the full complement of enterprise security controls while eliminating static credentials entirely.

JIT Credentials and Token Exchange

Rather than issuing persistent API keys that sit in configuration files waiting to be discovered, Aembit generates short-lived, ephemeral credentials at the moment they are needed and discards them immediately after use. No secrets are stored in Claude and no credentials are handed to the agent to manage.

This secretless approach extends to token exchange: when Claude needs to authenticate to a downstream service — whether Atlassian, Factset, Kensho, or Microsoft Graph — Aembit's Authorization Service brokers that exchange on its behalf, validating the requesting agent's blended identity, applying the relevant access policy, and issuing a minimally scoped token targeted to the specific service being called. The agent never holds a reusable credential, and receives only a JIT token valid for a narrow window that cannot be meaningfully replayed or repurposed even if intercepted.

Enforcement, Management, and Logging

Enforcement is handled in real time by Aembit's MCP Gateway, which sits inline between Claude and the MCP servers it connects to. When Claude makes a tool invocation request, the Gateway intercepts it and checks with the Aembit Authorization Service: “allow or deny?” Based on policy, Aembit either forwards the request with the appropriate downstream credentials or blocks it before it reaches the target server, before any tool is invoked, and before any data is exposed.

This policy enforcement model is managed entirely from Aembit's cloud control plane, giving the firm's security team a single point of governance across all agent-to-service

interactions — whether Claude is querying a financial data platform or accessing an analyst's Microsoft 365 calendar and email. Policies are defined once, applied consistently, and updated centrally for every Claude agent across the workforce.

Aembit creates a complete, attributable audit record of every request: which employee initiated the session, which Claude agent acted on their behalf, which tool was invoked, what parameters were passed, and what the policy outcome was. Those logs are shipped directly to CrowdStrike, closing the observability gap that had left the security team unable to distinguish analyst actions from agent actions across the firm's infrastructure — and giving them the same real-time, searchable visibility into Claude's behavior that they had long had for their human workforce.

Aembit provides:

- **Secretless authentication** — no static credentials stored, managed, or exposed.
- **Blended identity** — unified treatment of human and agent identities within the same access and audit framework.
- **Runtime JIT credentials** — ephemeral, single-use tokens generated at the moment of need and discarded immediately after.
- **Token exchange** — a governed, policy-enforced brokering layer that replaces the insecure token passthrough pattern.
- **Policy enforcement** — real-time allow/deny decisions at the traffic boundary, before any tool is invoked or data is reached.
- **Centralized management** — a single control plane governing all agent-to-service access policies across the environment.
- **Audit with attribution** — complete, human-agent-attributed logs for full operational visibility.

Setup, Configuration, and Deployment

Initially, the organization's security team invested a few hours a week on configuration with Aembit and helping employees onboard, but this time investment dropped after launch. Users needed to authenticate once with Aembit, just like they already did with Okta. With Aembit, developers continued to work with their agents and MCP servers without needing to worry about hardcoding secrets or credential rotation – nothing in their workload changed except security got stronger.

While self-hosting the Gateway is available for organizations with strict data residency requirements, the firm opted for Aembit-managed hosting.

The current deployment architecture for the firm is:

- **Identity provider:** Okta – all human authentication is tied back to the firm's existing Okta instance, ensuring that agent access policies are anchored to the same human identity foundation the firm already manages.
- **AI agent:** Claude via claude.ai web and Claude desktop.
- **MCP Gateway:** Deployed and managed by Aembit; sits inline between Claude and all connected MCP servers.
- **MCP servers in scope:** Microsoft 365 (primary, highest priority), Factset, Kensho.
- **SIEM integration:** CrowdStrike – all logs forwarded for security monitoring and alerting.
- **Policy management:** Currently configured via the Aembit UI by the firm's security engineers; Terraform-based configuration management is planned for steady-state production.

Within three weeks of launch, 100% of employees leveraging Claude agents were secured by Aembit – without an official rollout. With Aembit deployed, the firm's security team moved from a position of "we can't say yes" to "we can say yes, and here's exactly what yes means."

Before and *After*

Scenario	Before Aembit	After Aembit
End User	Agent identity and access based only on the user's identity and rights.	Users authenticate once with Aembit to create their agent's blended identity.
Developer	"Best effort" authentication with a mix of Oauth and secrets based on use case.	No-code auth, with no secrets management needed.
Financial Research Agent (Factset, Kensho)	Analyst's agent accessed financial data platforms using the analyst's full identity and rights. No distinction in logs between analyst-initiated and agent-initiated access.	Agent carries a blended identity credential (analyst + agent) enabling the security team to attribute every data access event to a specific agent acting on a specific user's behalf.
M365 Productivity Agent (Email, Calendar, SharePoint)	Agent would have operated with the analyst's full Microsoft identity, with access to the entire M365 surface area. No mechanism to restrict scope below the application level.	Agent access anchored to the analyst's Okta identity. Microsoft application-level permissions define the outer boundary. Aembit policy enforces least-privilege access within that boundary.

Internal Proprietary MCP Server	Firm's own MCP server implementation faced authentication gaps. Standard MCP OAuth implementations lacked PKCE and constrained client registration, failing the firm's internal security standards.	Aembit Authorization Server provides standards-compliant OAuth authentication for the firm's internal MCP server, meeting enterprise security requirements the standard MCP spec did not.
Security and Audit Teams	No way to distinguish agent-initiated actions from human-initiated actions in logs. No real-time visibility into what agents were doing or attempting.	Full per-agent audit trail forwarded to CrowdStrike SIEM: agent identity, tool invoked, parameters passed, policy decision, timestamp. Same operational visibility as human user activity.

The Benefits

The firm went from a security team that had been saying "no" – not out of obstruction, but because no tool existed to let them say yes responsibly – to a production deployment of agentic AI across its analyst workforce, with full identity, access control, and audit coverage in place from day one.

Centralized Access Control

- Across all MCP-connected agents and services, managed from a single Aembitpolicy control plane.

Elimination of long-lived Credentials

- Stored in Claude or MCP server configurations; replaced with short-lived, policy-scoped tokens issued per-session by Aembit.

Full Attribution

- For every agent action – the blended identity model means audit logs can definitively distinguish human-initiated access from agent-initiated access.

Real-time SIEM integration

- With CrowdStrike, giving the security team operational visibility into agent behavior on the same platform they use for human user monitoring.

What's Next

The firm's engagement with Aembit is expanding in three directions.

Expanding the AI footprint to cover more agents and MCP servers

The organization's AI deployment is only going to keep growing. The foundation they've laid with Aembit's IAM for Agentic AI is poised to expand as they bring on additional agents and MCP servers.

Expanding AI security into agent-to-agent communication

Within the next two months, the organization will need to begin addressing agent-to-agent communication scenarios in which one AI agent invokes another as part of a workflow, creating an identity and access control challenge that goes beyond the human-to-agent model the current solution addresses. This is the next frontier of agentic AI security, and it is already on the horizon for this customer.

Expanding IAM to workloads

Aembit's IAM for Agentic AI solution is built on the foundations of solving a larger problem of IAM for workloads. As the partnership between the organization and Aembit progresses, the same technology that gives AI agents identity-based access to sensitive resources can expand to provide real-time access security to secure CI/CD pipelines, eliminate manual credential management for DevOps teams, and simplify compliance across the organization with secretless, just-in-time credentials for sensitive applications.

Secure Claude With Aembit's *IAM for Agentic AI*

The free 'Starter' tier is the fastest path to understanding how Aembit can secure your organization's Claude deployment. With enough capacity to run agents across development and QA before you scale to production, you can get started today – no sales call required.

Sign up for the free 'Starter' tier | **Talk to an engineer** | **Read the AI Guide docs**

This case study has been anonymized at the customer's request. Technical details and scenario descriptions have been verified with the customer. Quantitative metrics and additional implementation details confirmed prior to publication.

aembit

Gartner COOL VENDOR 2022 · CSO Cybersecurity startups to watch for in 2023 · 2024 WINNER CYBER SECURITY EXCELLENCE AWARDS · RSAC Innovation Sandbox 2024 FINALIST · 2024 SINET16 INNOVATOR AWARD