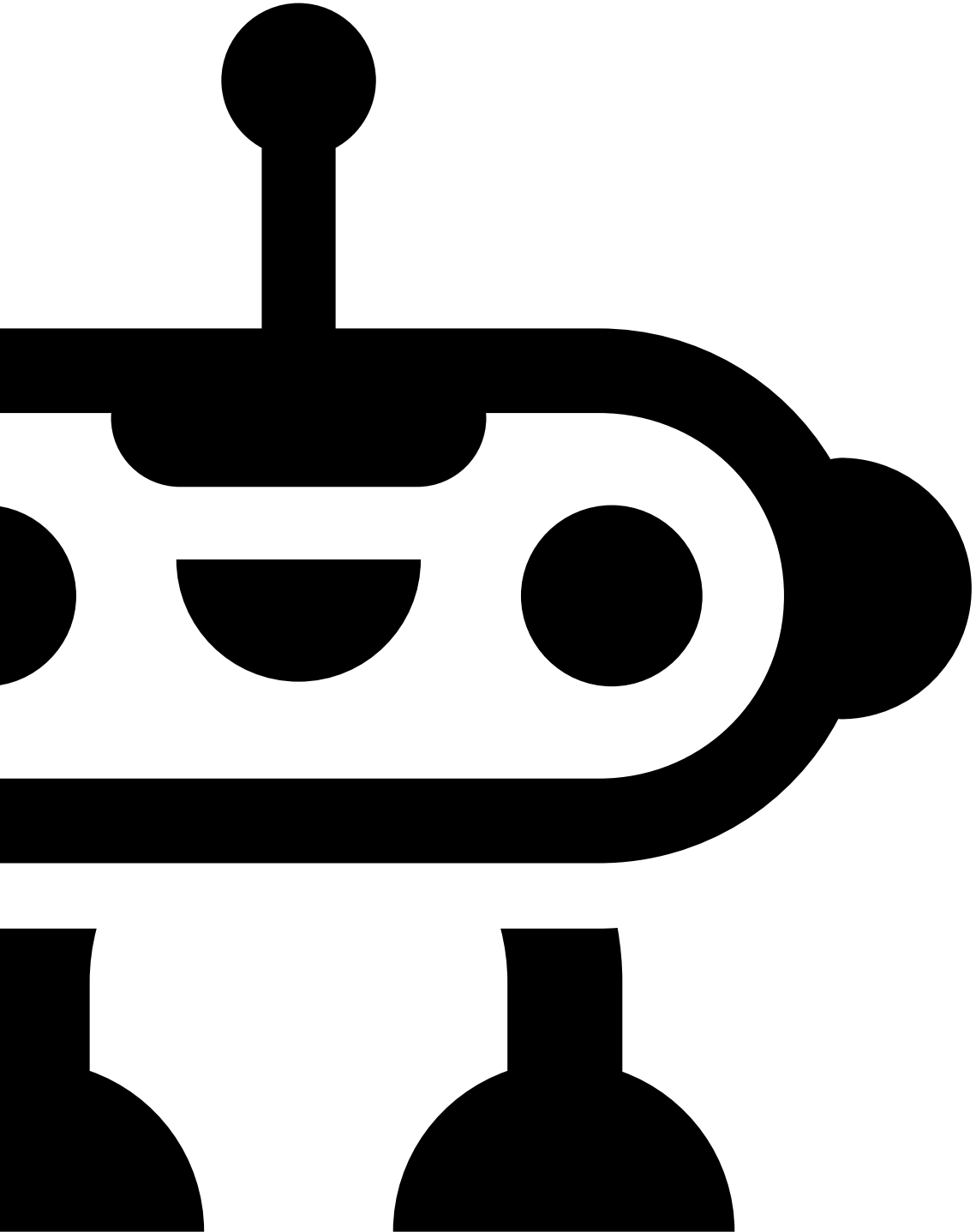
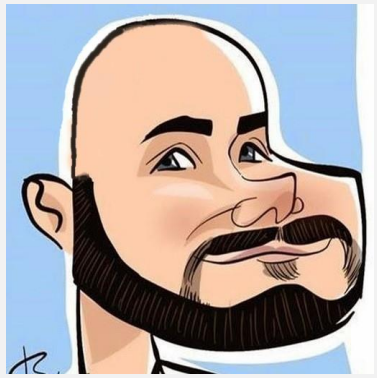


The Boring Attack That Will Actually Get You

obot

Today’s Presenter – Me!



Craig Jellick

VP of Engineering

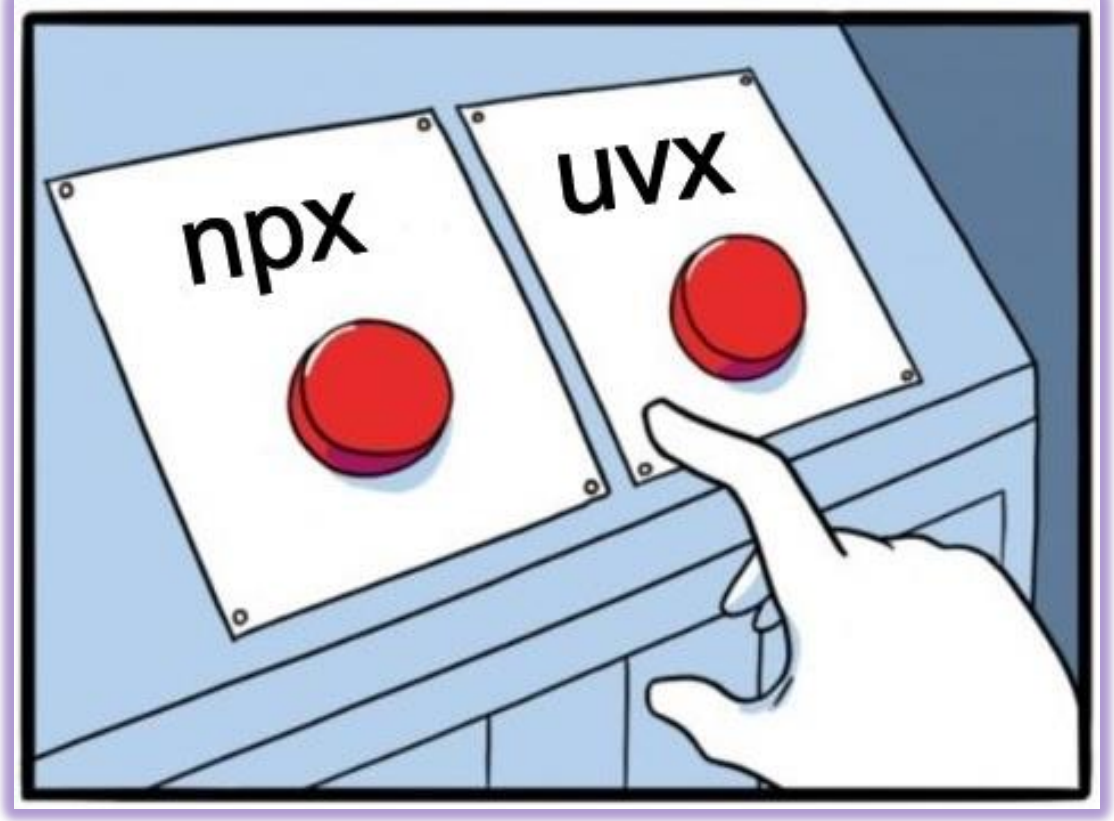
obot

The Problem

We spend a lot of time talking about novel AI & MCP centric attacks, but...

YOU NEED TO BE MUCH MORE WORRIED ABOUT SUPPLY CHAIN SECURITY

obot

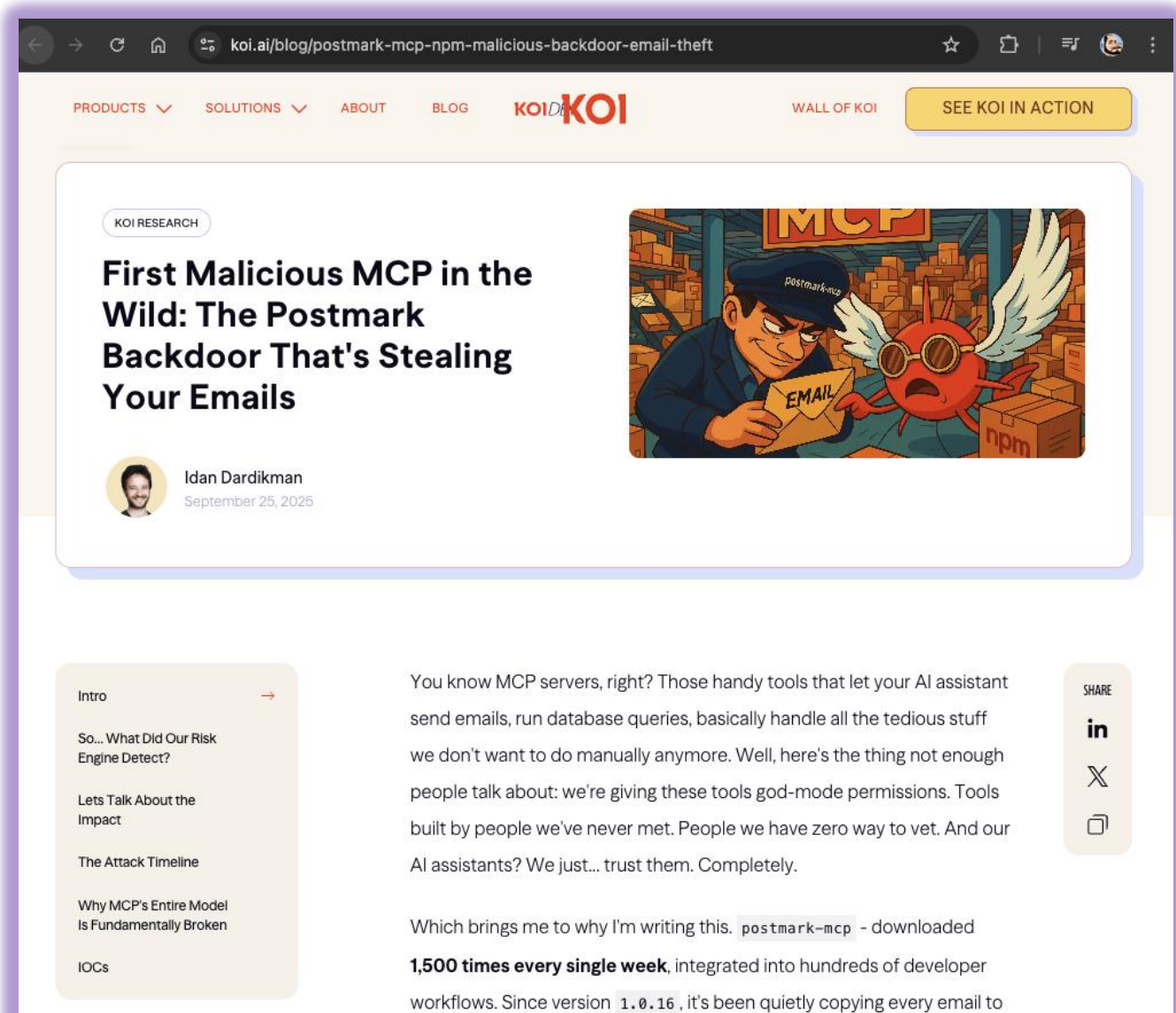


This is how I want you to feel

Case Study 1: postmark-mcp impersonation

https://www.koi.ai/blog/postmark-mcp-npm-malicious-backdoor-email-theft

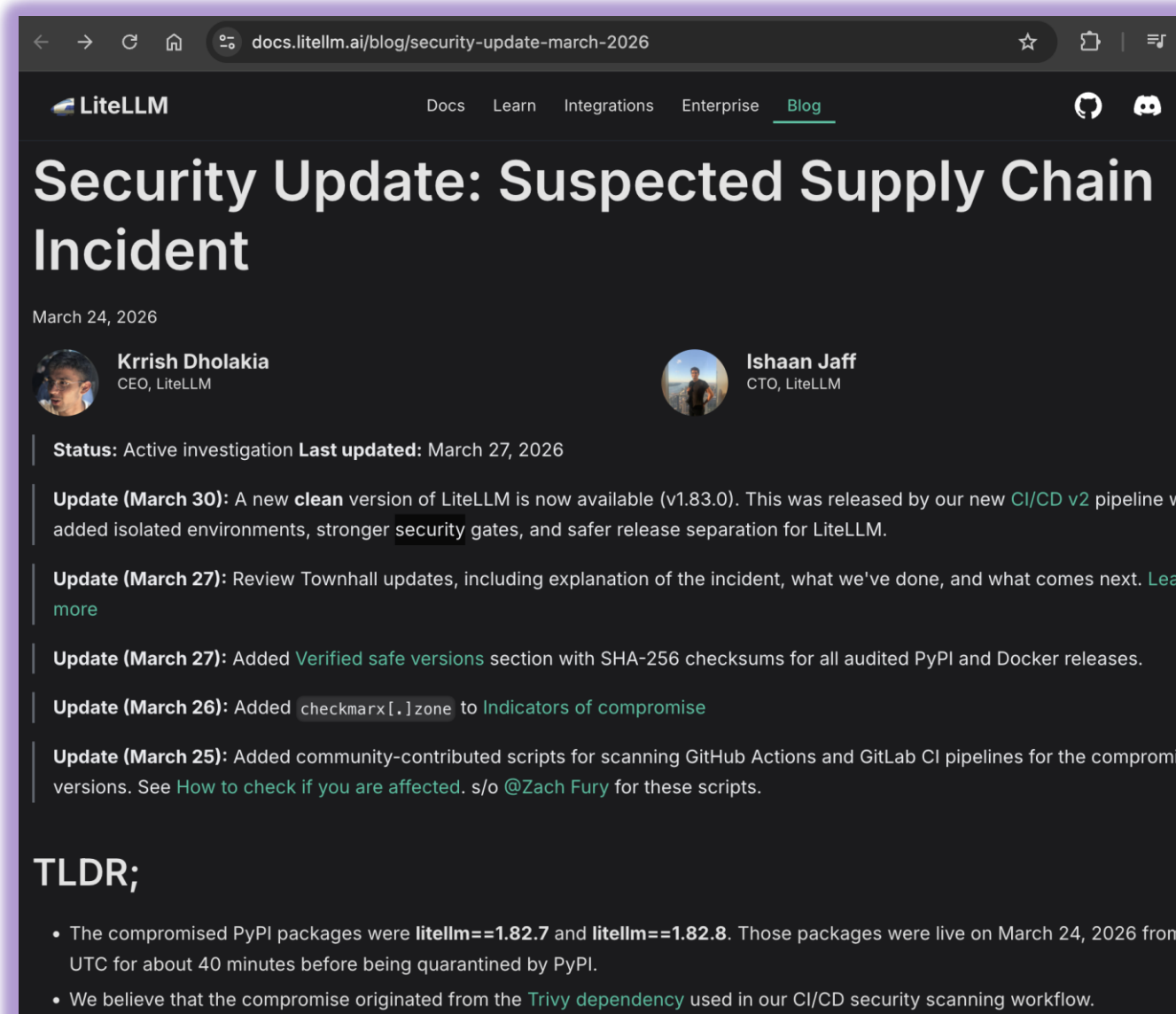
- 25 SEP 2025
- Attacker legitimately owned postmark-mcp on npm.
- Code matched the official GitHub repo & worked flawlessly for 15 versions
- Single line added to exfiltrate emails via BCC



Case Study 2a: LiteLLM Supply Chain Attack

https://docs.litellm.ai/blog/security-update-march-2026

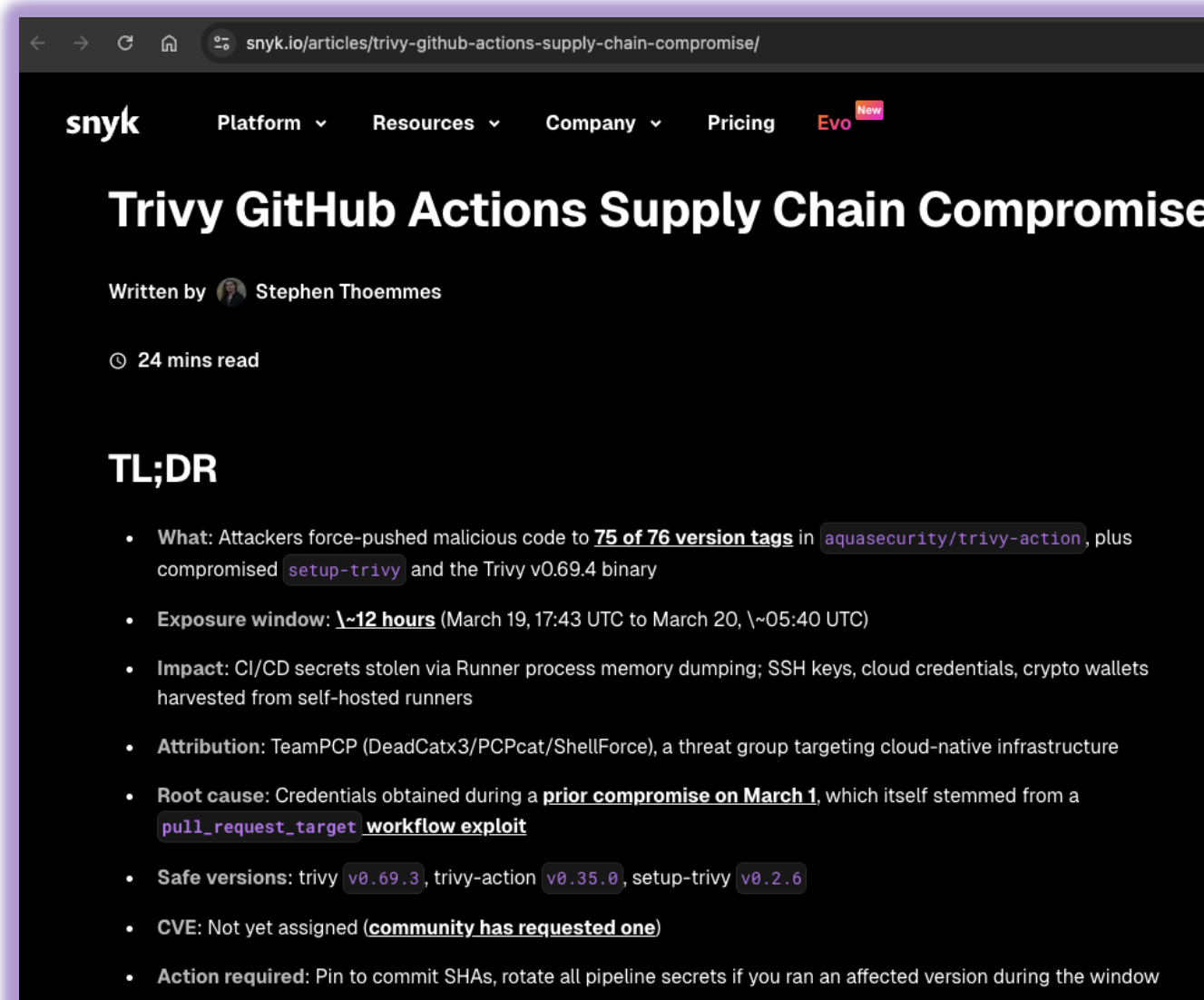
- 25 MAR 2026
- Discovered when Callum McMahon at FutureSearch was testing a **Cursor MCP plugin that pulled in litellm as a transitive dependency**
- Malicious code scanned for and stole keys, tokens, & passwords from host machine
- Litellm was compromised via a Trivy exploit...

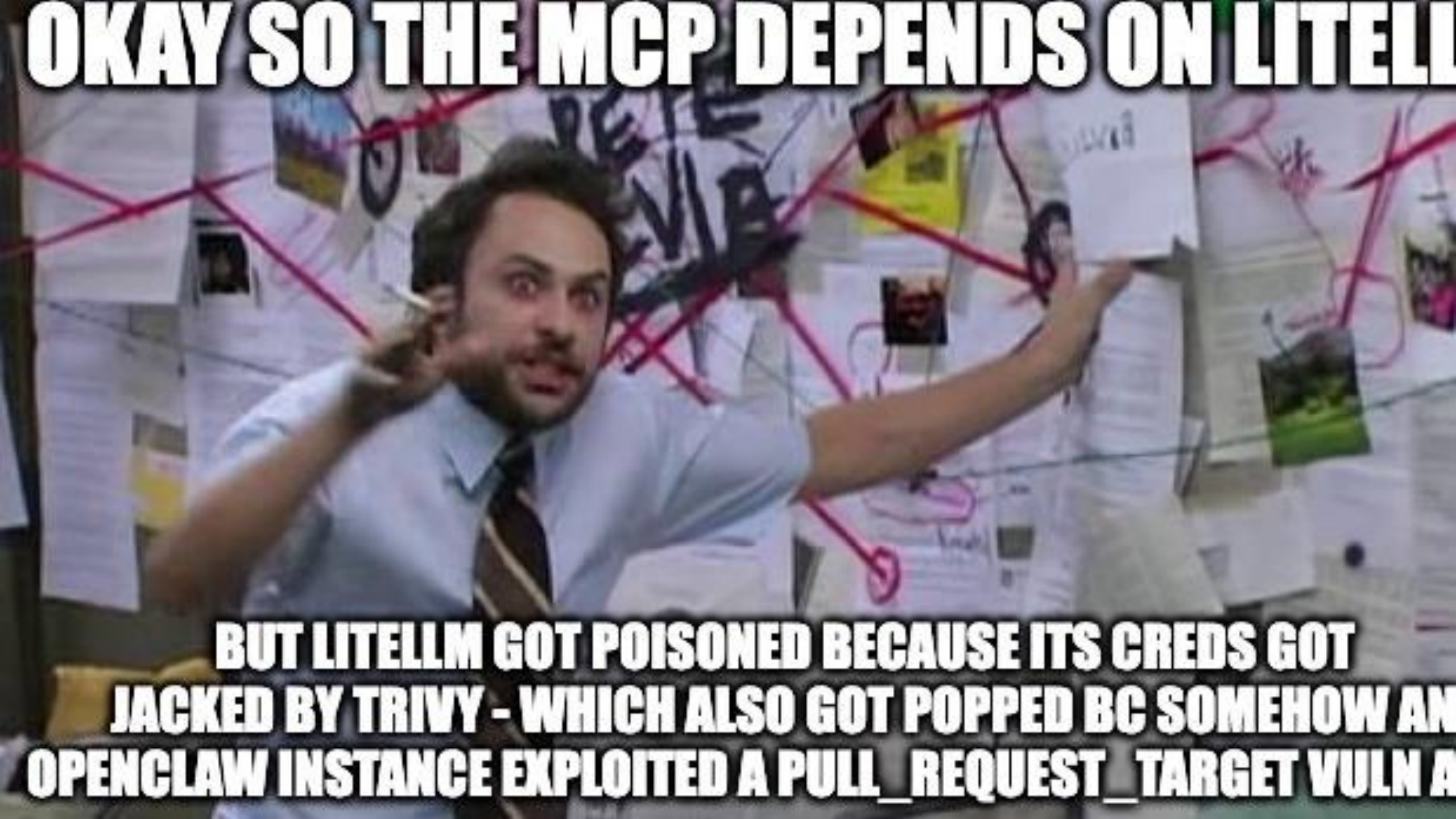


Case Study 2b: Trivy GitHub Action Breach

https://snyk.io/articles/trivy-github-actions-supply-chain-compromise/

- 19 MAR 2026
- Attacker rewrote Trivy GitHub Actions with creds obtained from an EARLIER attack - *hackerbot-claw* scanning for *pull_request_target*
- Projects that used Trivy GH Action to scan dependencies had GH secrets stolen

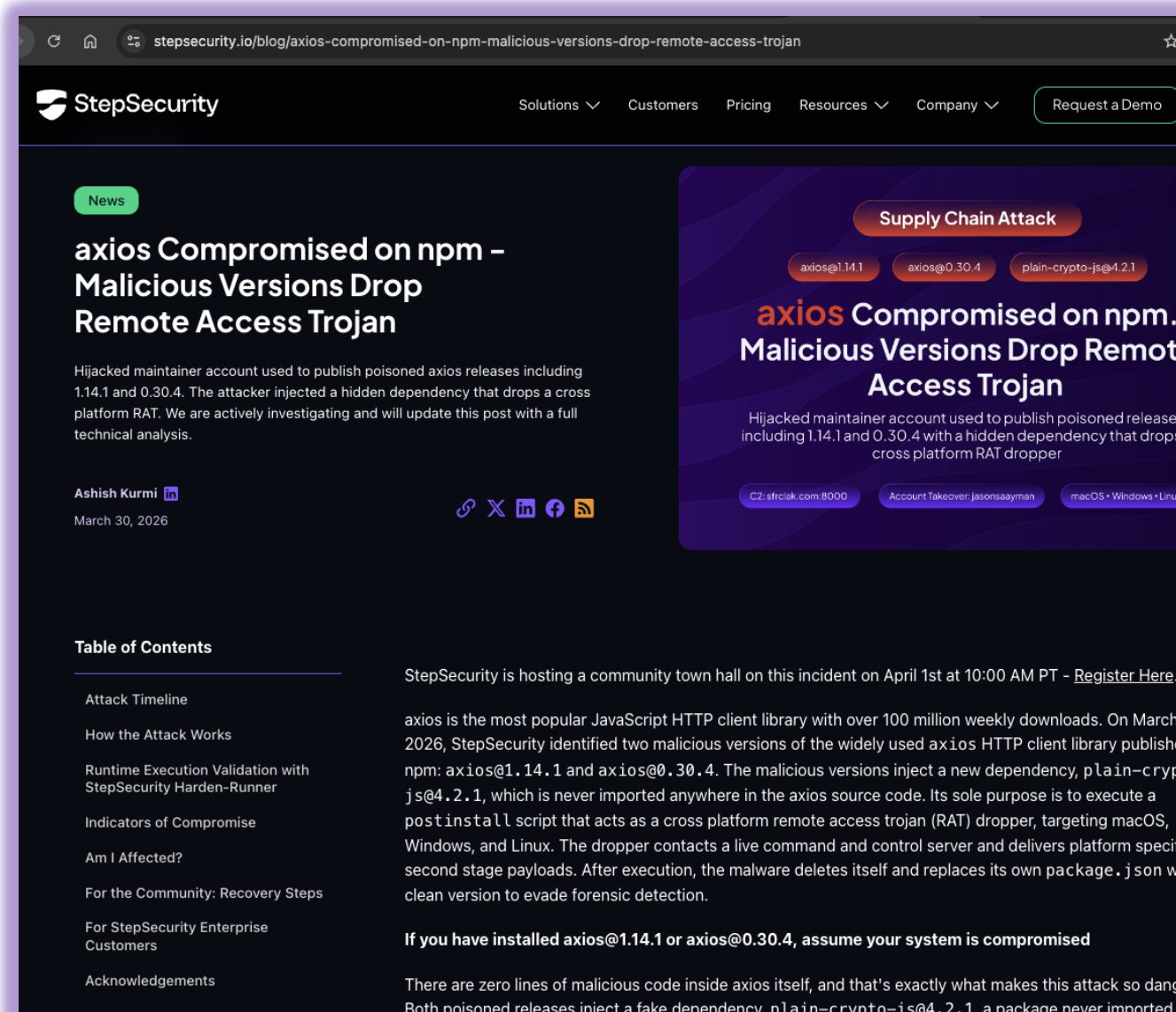



OKAY SO THE MCP DEPENDS ON LITELL
BUT LITELLM GOT POISONED BECAUSE ITS CREDS GOT
JACKED BY TRIVY - WHICH ALSO GOT POPPED BC SOMEHOW AN
OPENCLAW INSTANCE EXPLOITED A PULL_REQUEST_TARGET VULN A

Case Study 3: Axios RAT npm compromise

https://www.stepsecurity.io/blog/axios-compromised-on-npm-malicious-versions-drop-remote-access-trojan

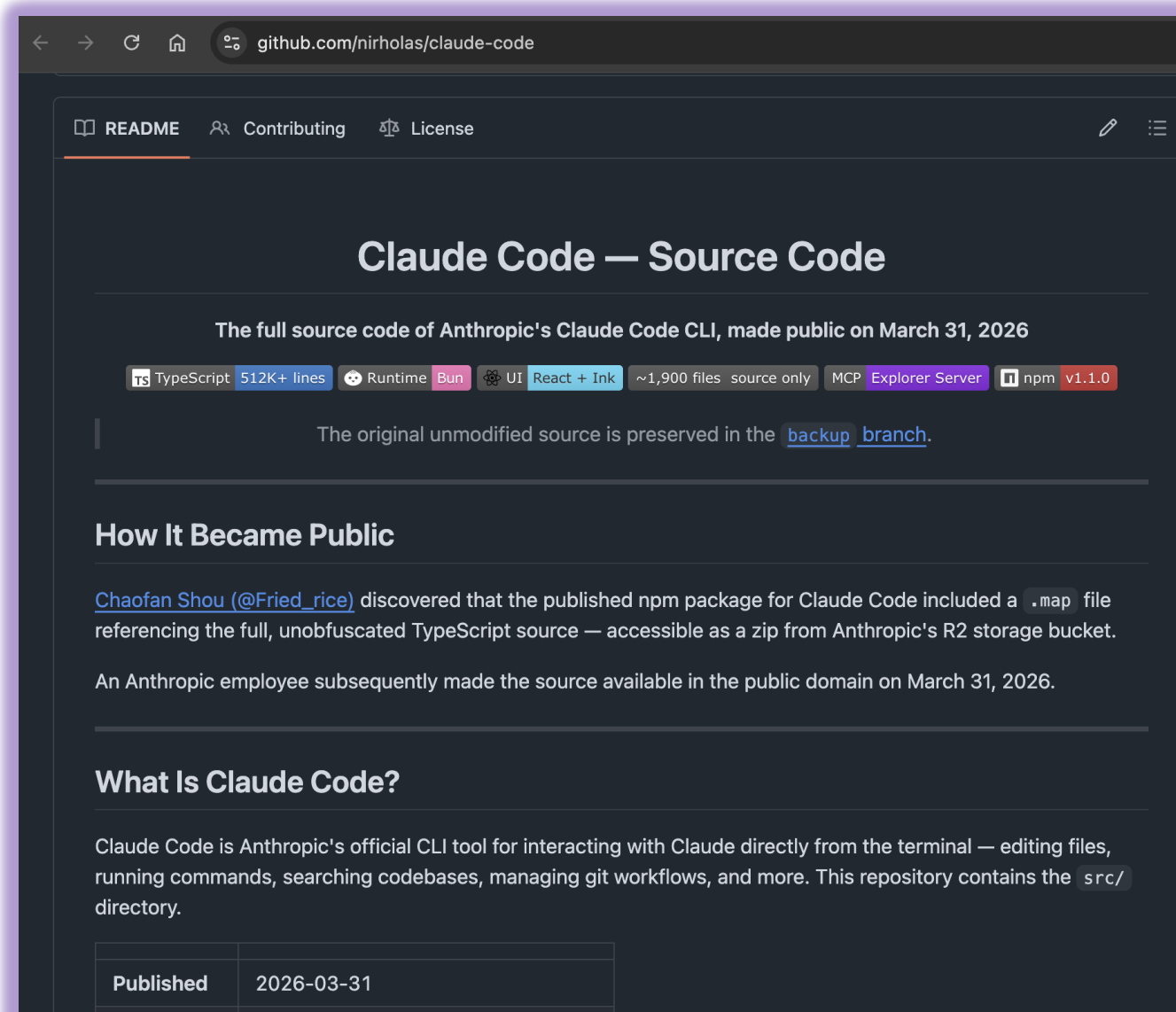
- 30 MAR 2026
- HUGE impact to JS/TS ecosystem
- Maintainer's account compromised
- Double-obfuscated and self-erasing hidden dep
- Platform-specific payloads for command execution



Bonus Study: Claude Code Source Code Leak

https://github.com/nirholas/claude-code

- 31 MAR 2026
- Not a security exploit
- Not the keys to the kingdom
- BUT…what if?



What Can We Do?

- Treat each MCP server as an untrusted, independent security domain
- Review server source code and tool definitions before installation
- Verify package integrity with checksums or code signing
- Only install MCP servers from trusted, verified sources
- Use tools to automatically analyze and monitor installed servers for malicious behavior or changes
- Carefully verify package names before installation
- Scan MCP server dependencies for known vulnerabilities
- Conduct regular security audits
- Sandbox and Isolate MCP Servers (no local MPC servers)
- Log all MCP tool invocations with full parameters, user context, and timestamps
- Leverage secure compute and network infrastructure

Obot Overview

AI Clients and Tools

ChatGPT
Copilot
Claude
glean
CURSOR
goose
Windsurf
n8n
LangChain

Obot Agent
Chat | Workflow | Skills

Obot Gateway

OAuth Controller

- Callback URL Handling
- Token Refresh
- Secure Token Storage

MCP Security & Ops

- Access Control
- Content Filtering
- Audit Logs
- Secret Management
- Scaling & Operations
- Skills Registry
- MCP Registry & Hosting

Obot Catalog

External

Internal

Custom

Skills

Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

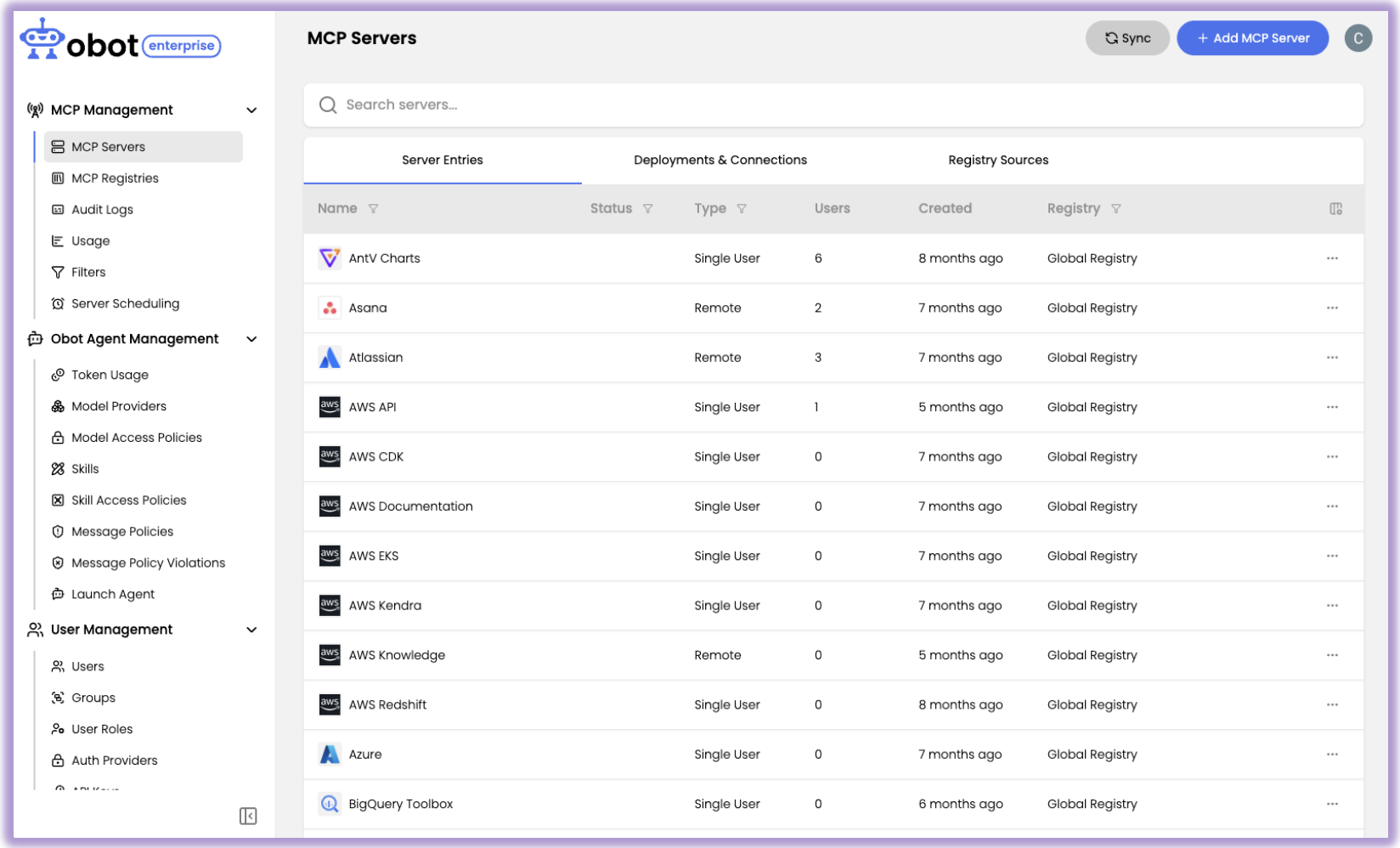
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

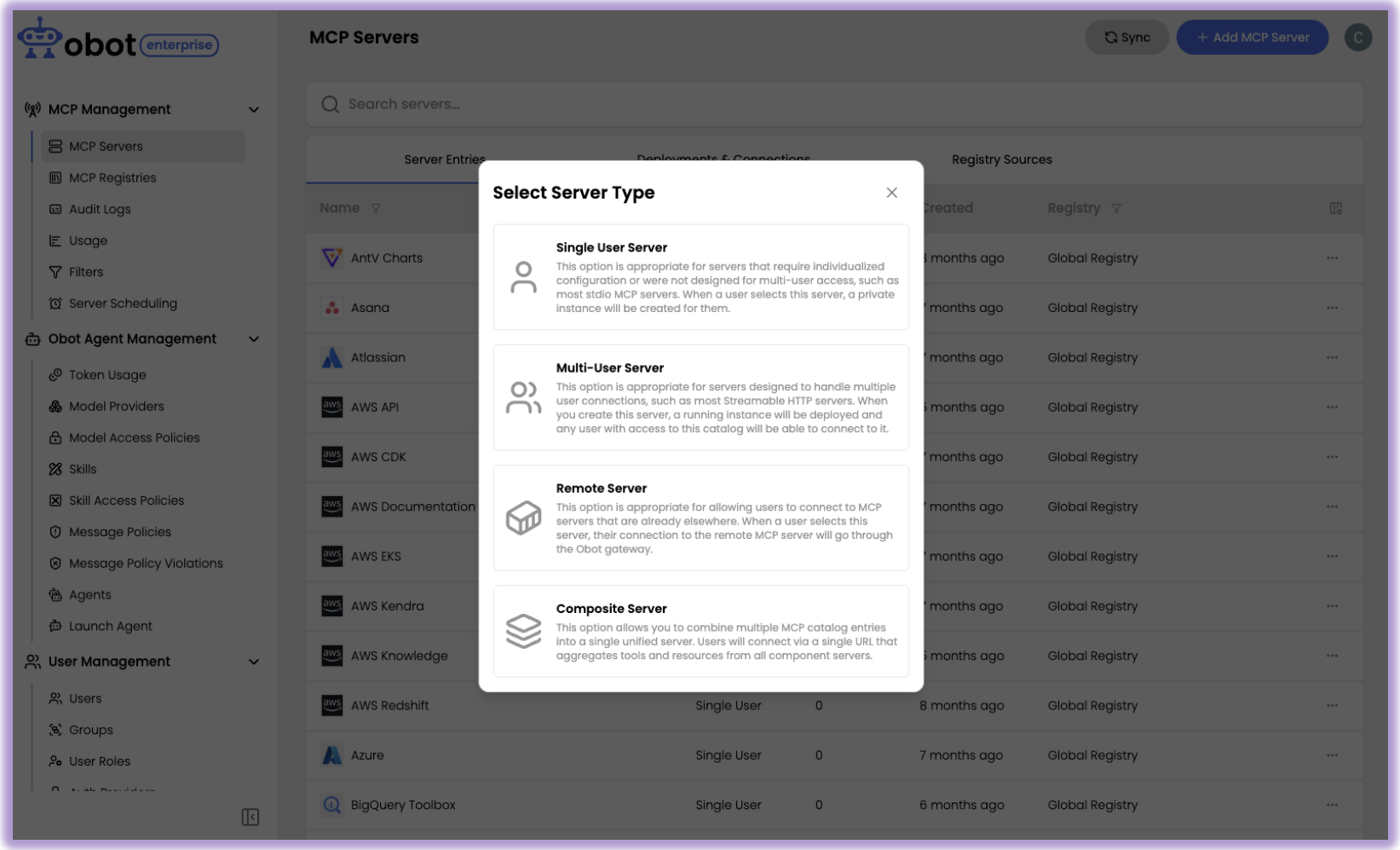
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

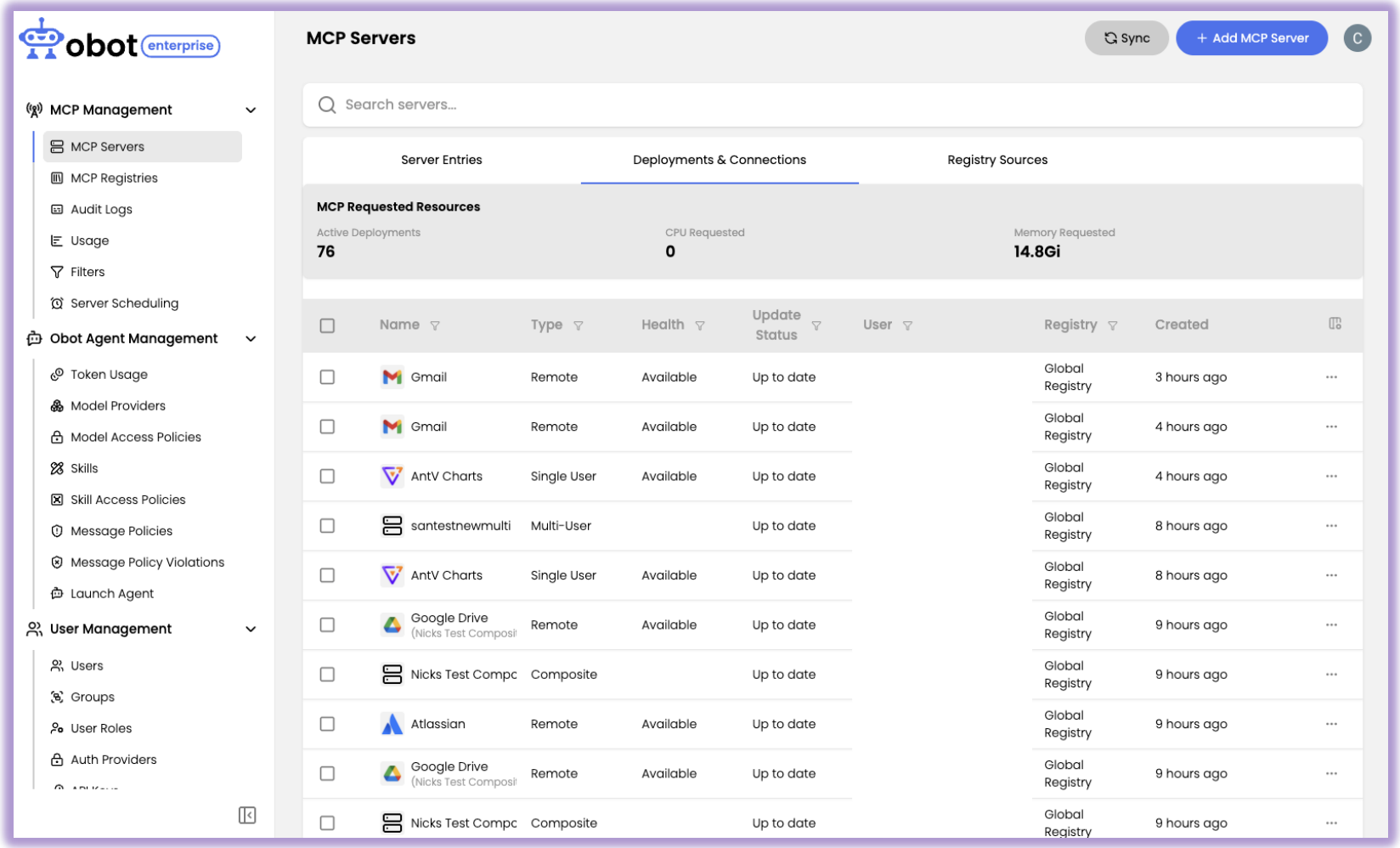
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

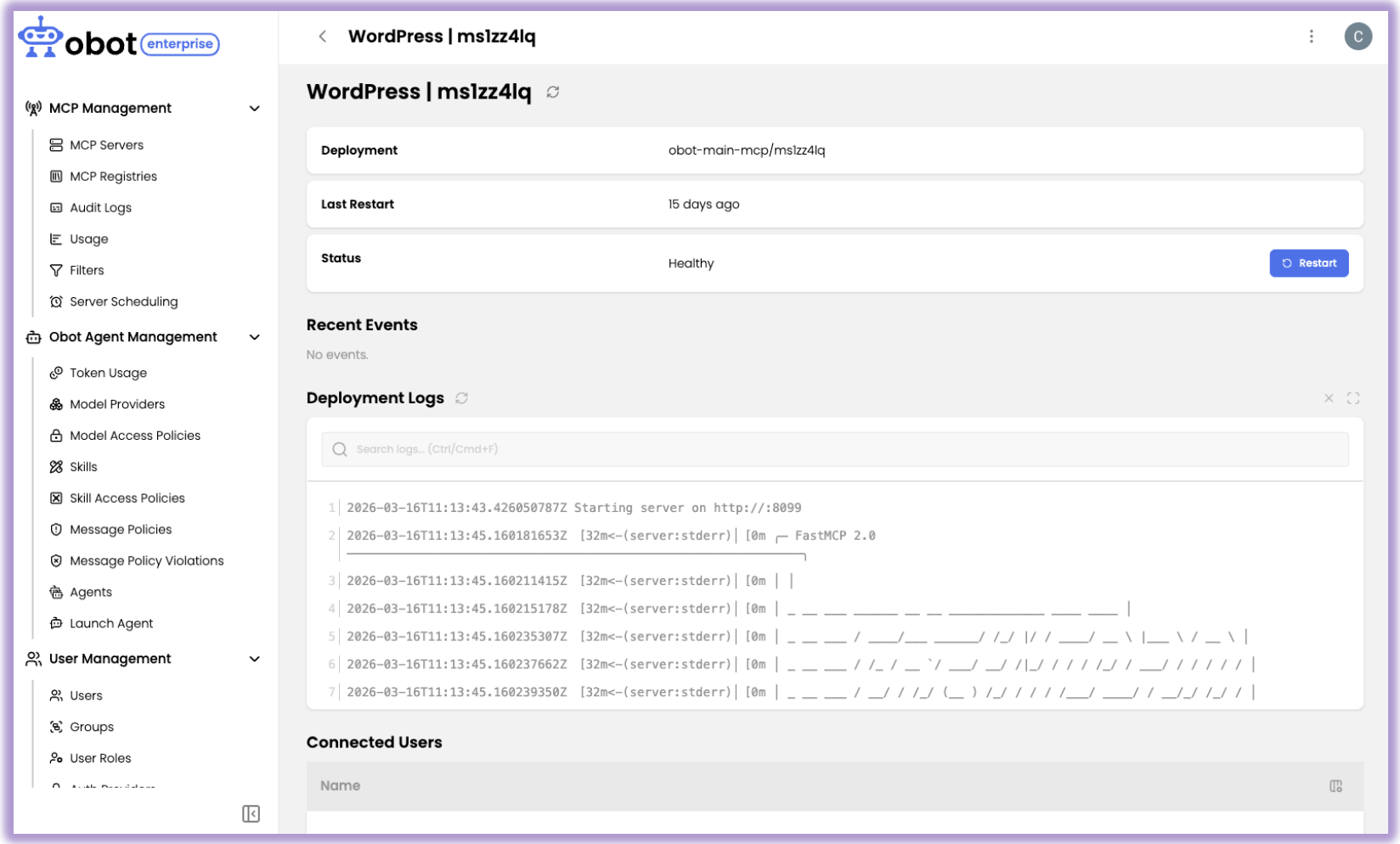
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

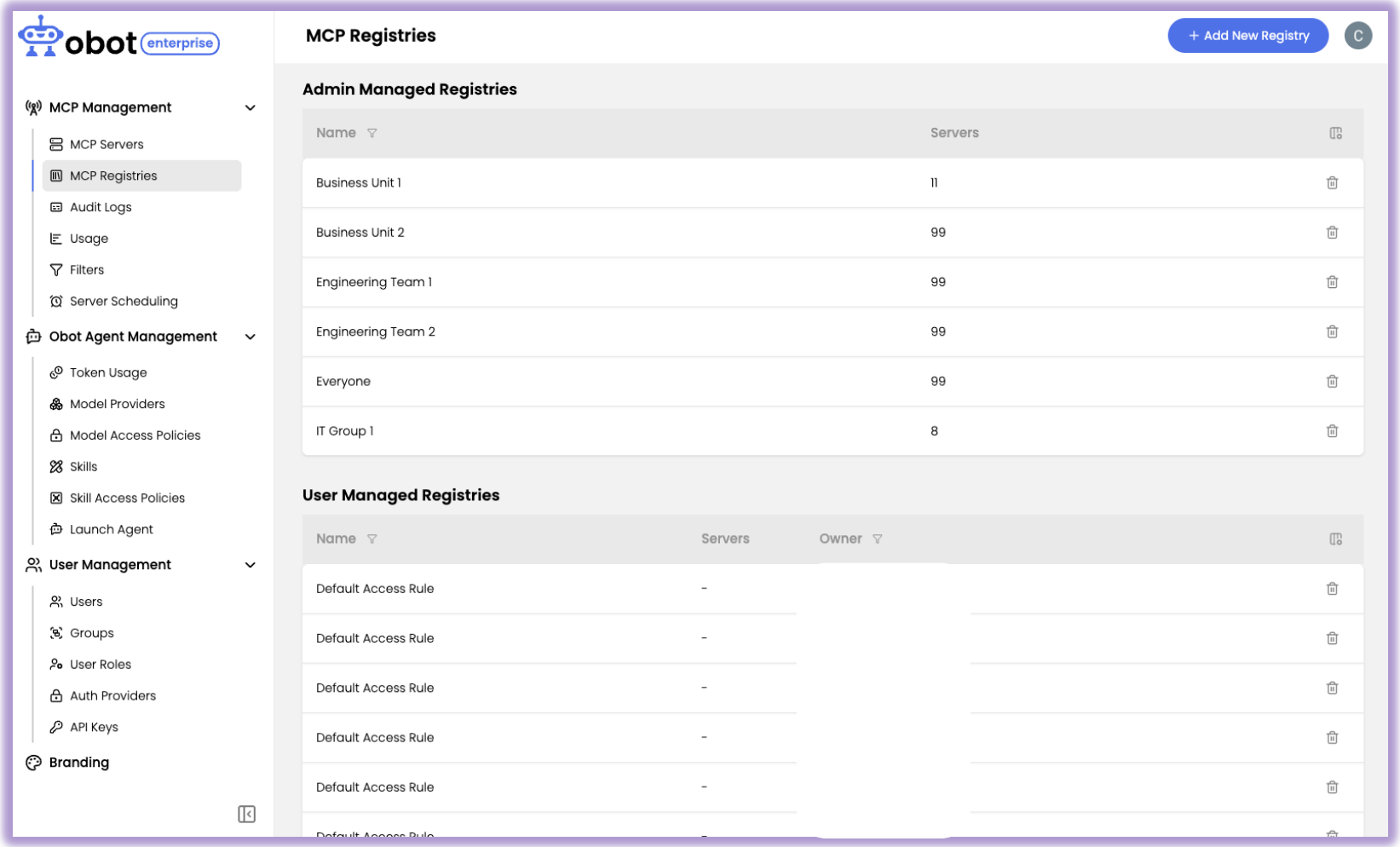
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

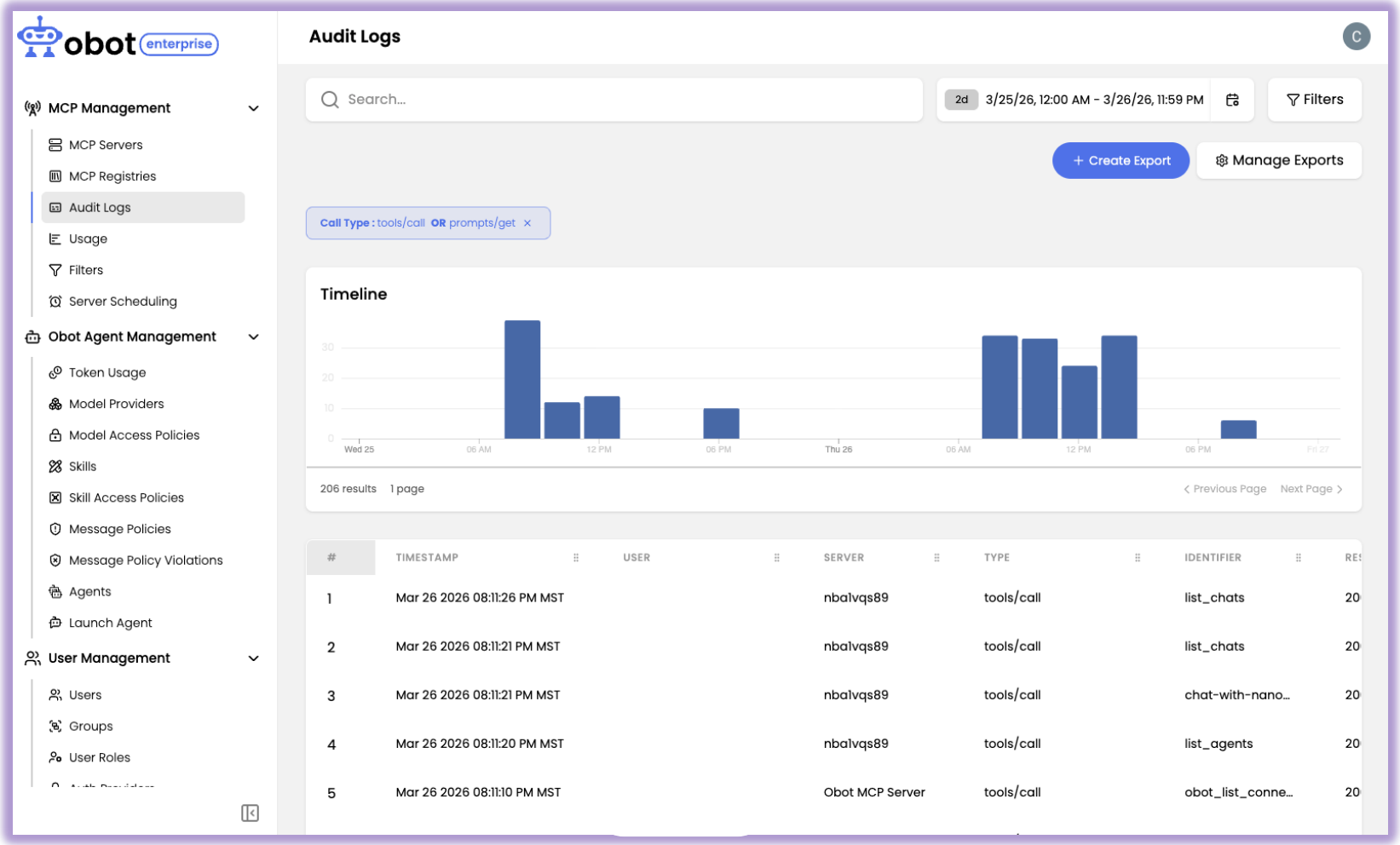
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

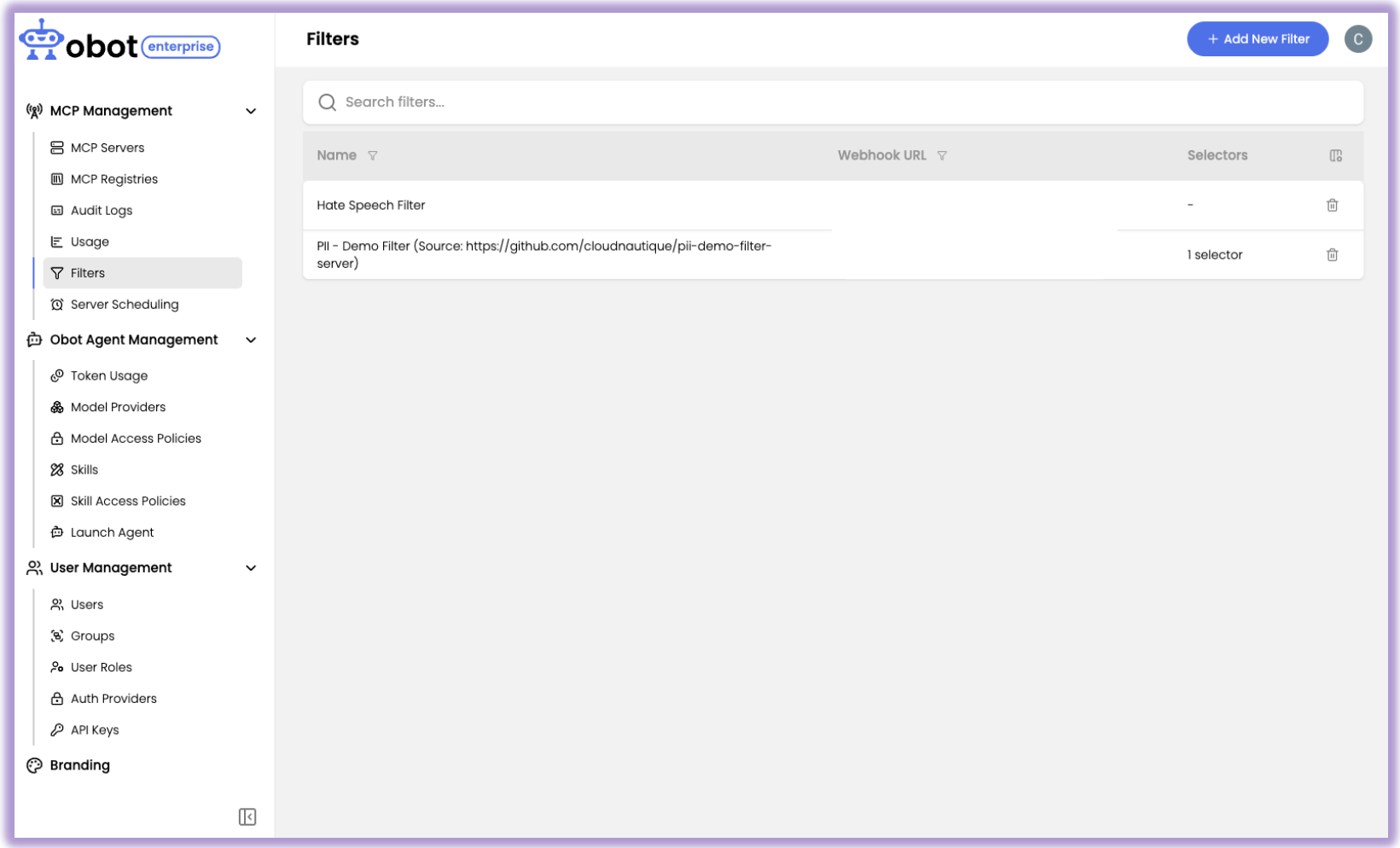
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

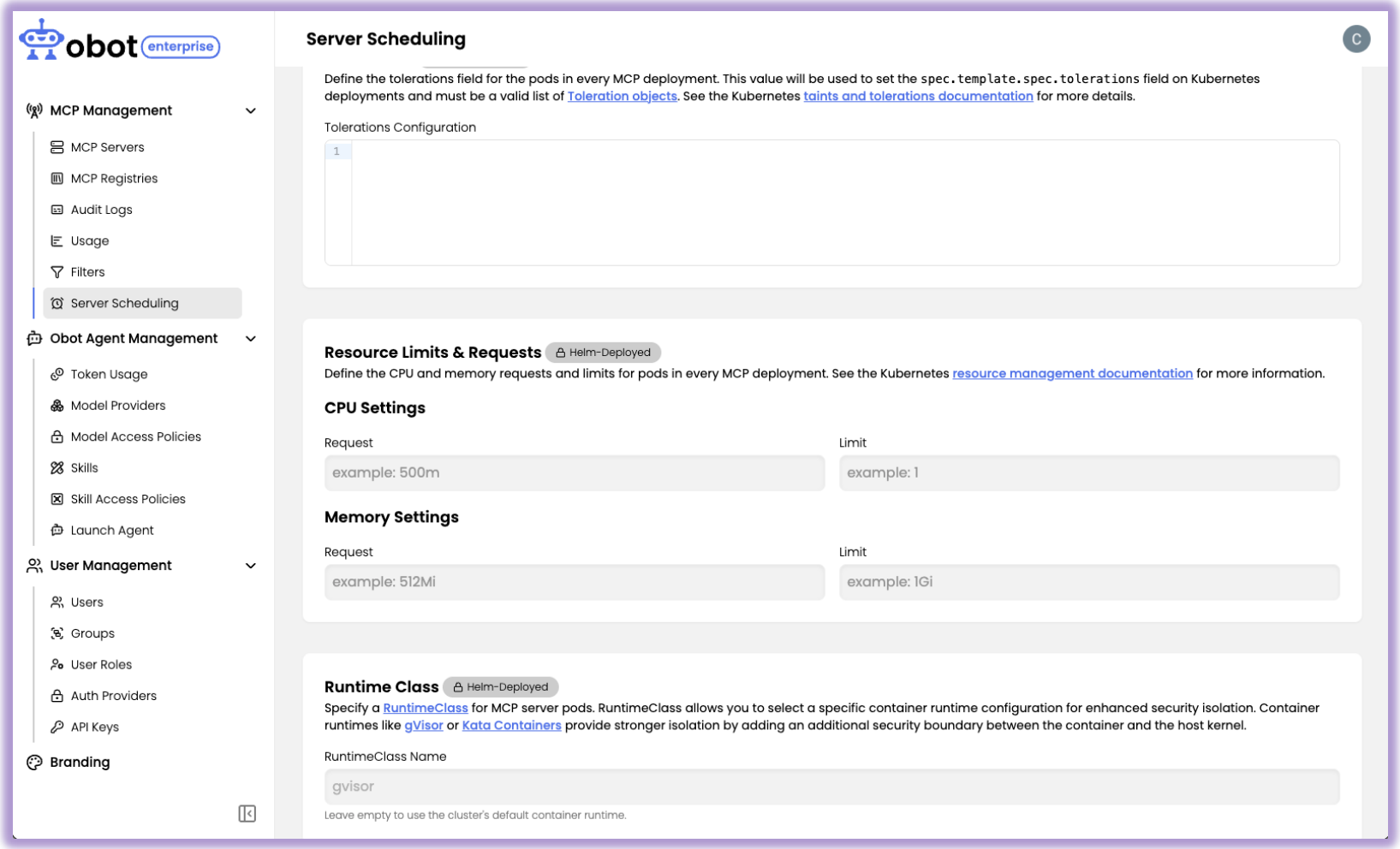
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

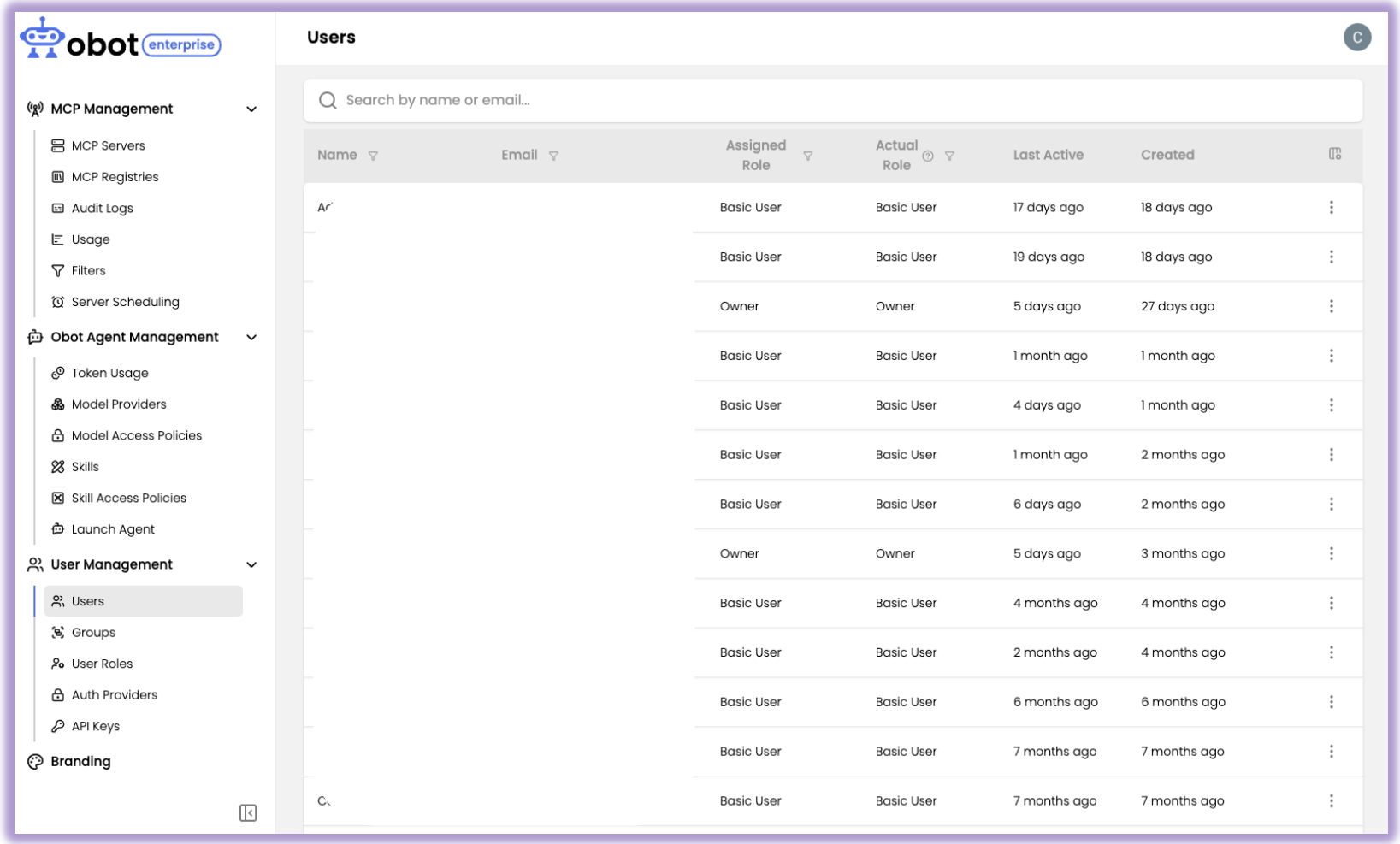
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

Infrastructure-level security

User Management and Role Based Access Control



Obot Gateway

MCP Server Curation & Creation

Deep Operational Visibility

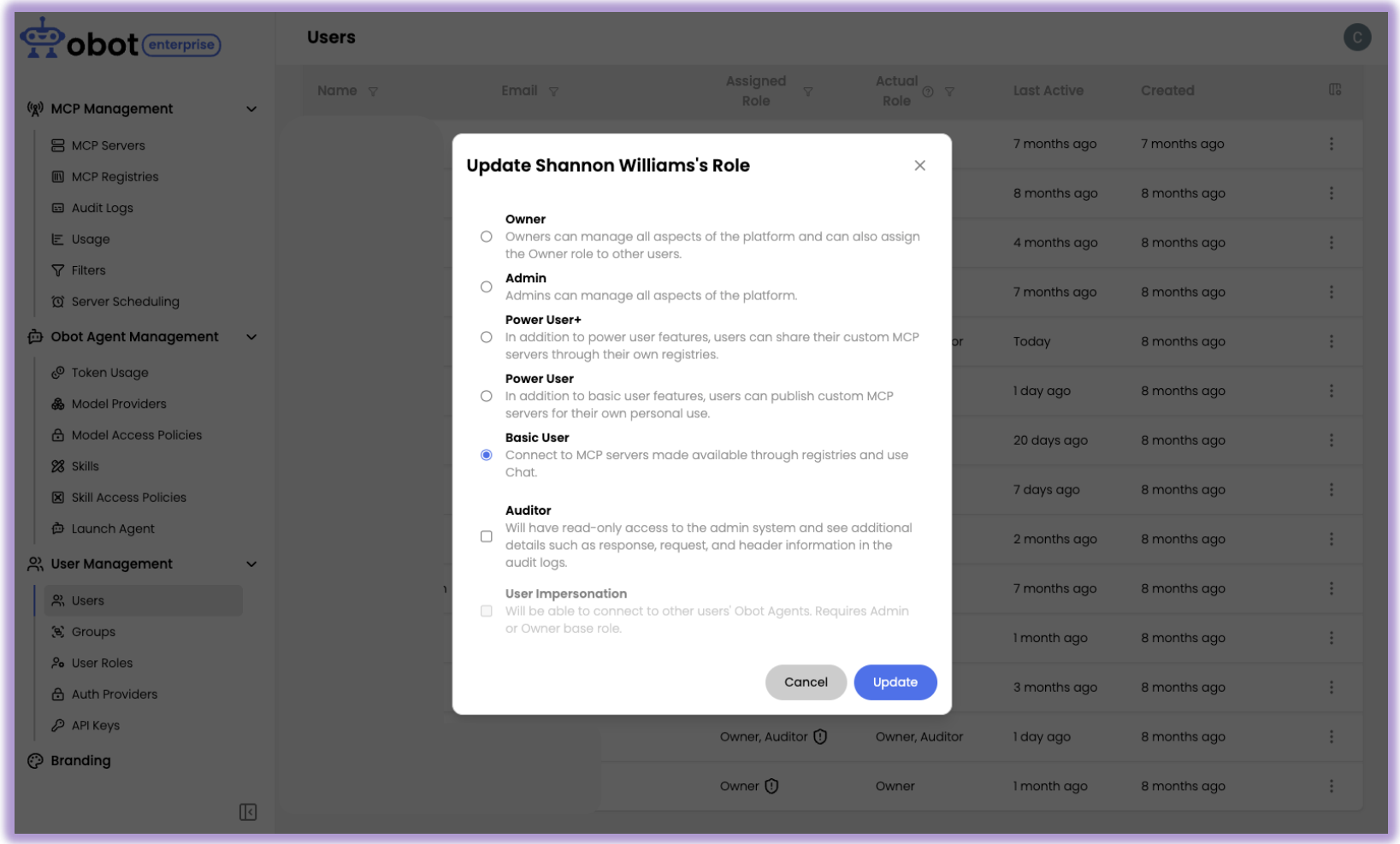
MCP Registry Access Control

Audit Logs and Usage Statistics

Runtime Filters and Hooks

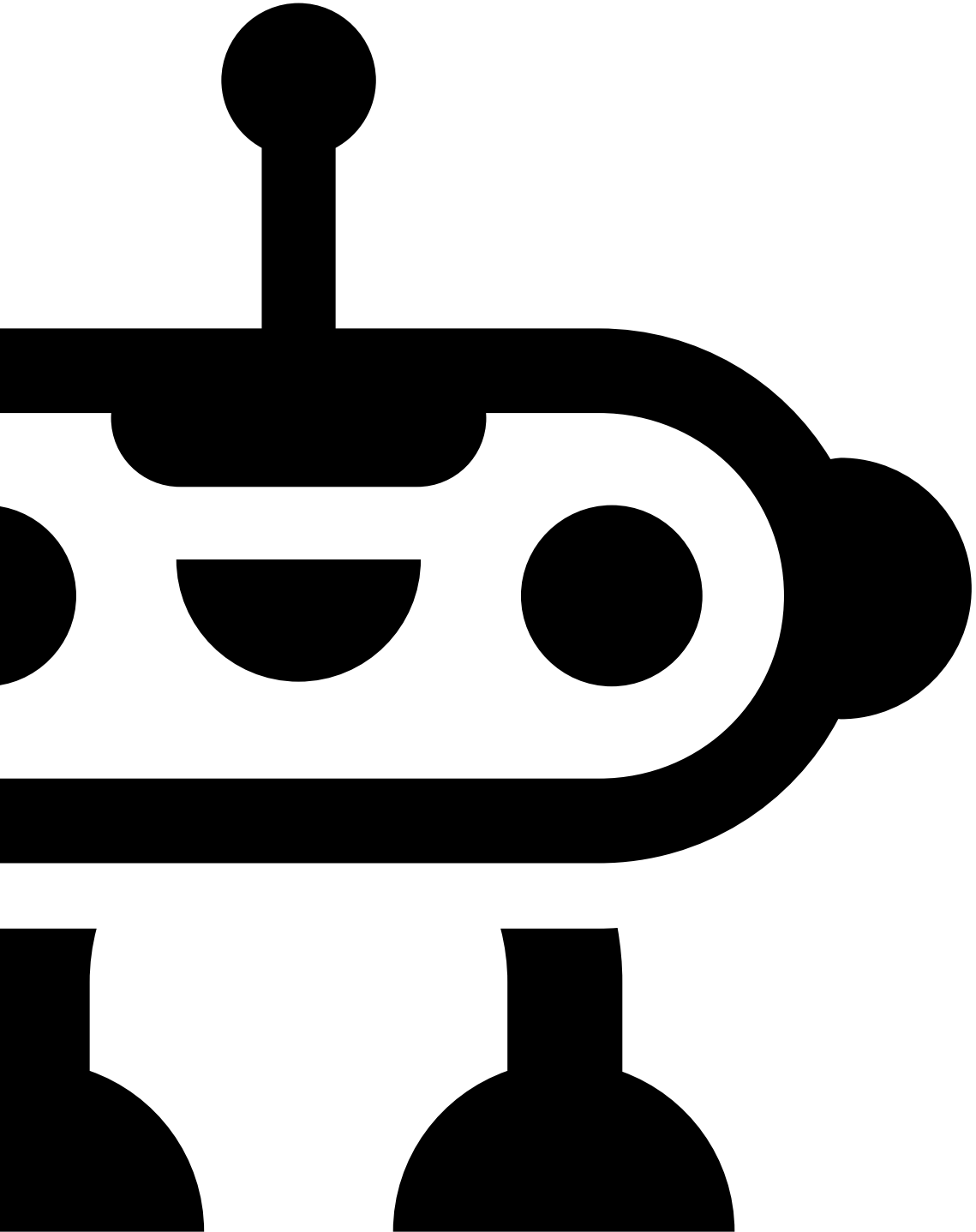
Infrastructure-level security

User Management and Role Based Access Control



Citations and Resources

- https://www.koi.ai/blog/postmark-mcp-npm-malicious-backdoor-email-theft
- https://www.wiz.io/blog/trivy-compromised-teampcp-supply-chain-attack
- https://www.trendmicro.com/en_us/research/26/c/inside-litellm-supply-chain-compromise.html
- https://www.aquasec.com/blog/trivy-supply-chain-attack-what-you-need-to-know/
- https://github.com/aquasecurity/trivy/discussions/10425
- https://www.reversinglabs.com/blog/rl-identifies-malware-ml-model-hosted-on-hugging-face
- https://snyk.io/blog/poisoned-security-scanner-backdooring-litellm/
- https://snyk.io/blog/axios-npm-package-compromised-supply-chain-attack-delivers-cross-platform/
- https://github.com/nirholas/claude-code
- https://cheatsheetseries.owasp.org/cheatsheets/MCP_Security_Cheat_Sheet.html



Thank You!

https://obot.ai

obot