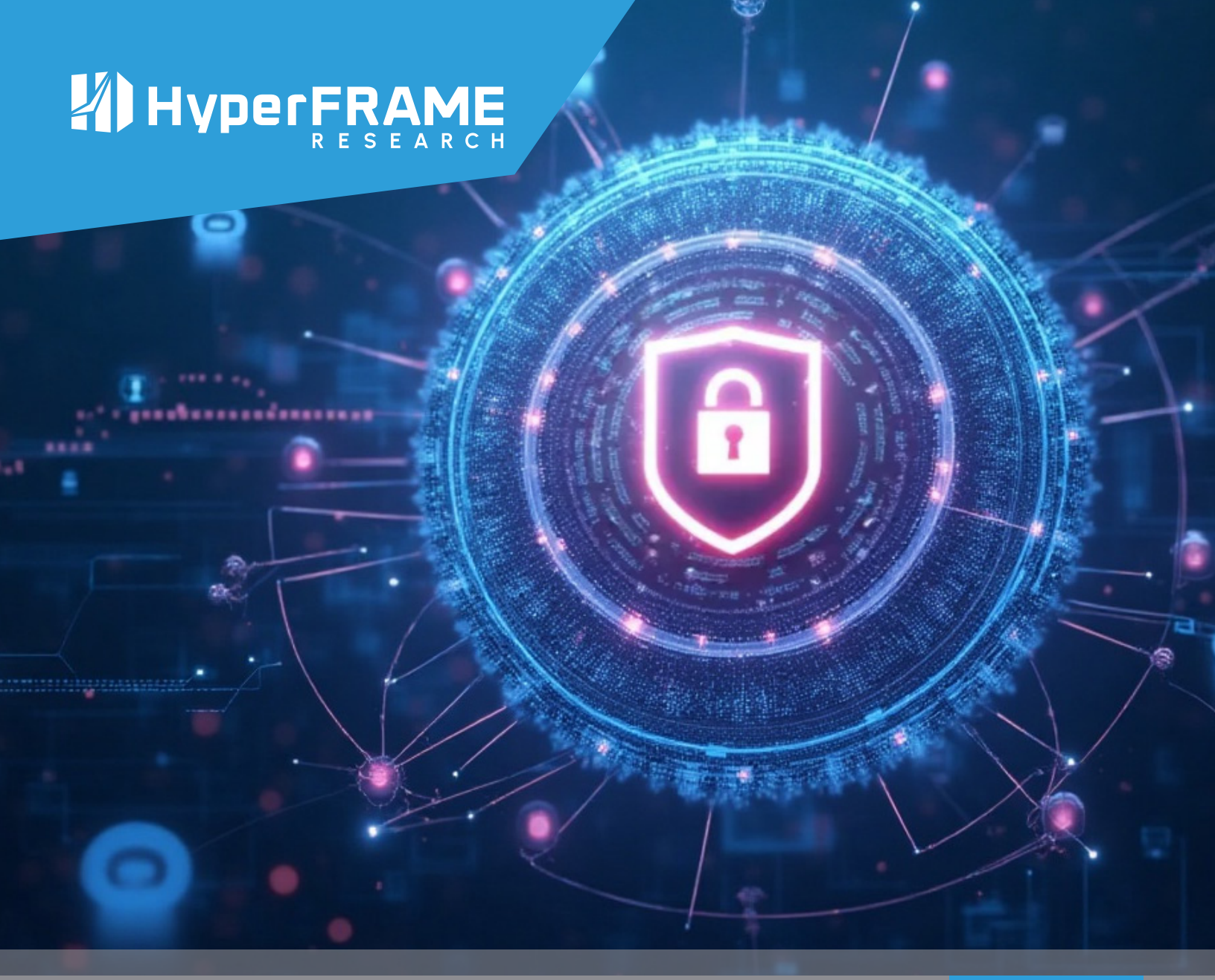
HyperFRAME RESEARCH



WHITEPAPER

Securing the Model Context Protocol: Access, Authorization, and Audit for Enterprise AI

The rise of Model Context Protocol and enterprise LLM deployment demands a fresh look at AI security.

Author:
Stephanie Walter
Analyst In Residence - AI Tech Stack

MAY 2025

Executive Summary

Large language models (LLMs) are reshaping how enterprises interact with internal systems and data. AI models are increasingly being embedded into core business workflows as organizations move from AI experimentation to production. As part of the overall rearchitecting of the AI stack, a key component has emerged, namely the Model Context Protocol (MCP), which provides a standardized way for LLMs to interact with enterprise systems. The flexibility delivered through MCP has accelerated innovation while also introducing complex access and data security risks.

As the AI stack has been deployed, inherent security risks have become manifest. However, instead of building a new security model specific to AI from scratch, organizations should consider security frameworks such as Infrastructure Identity that enable AI security within the same framework that governs humans and machines within infrastructure. Infrastructure Identity combines cryptographic identity, Zero Trust, short-lived privileges, governance, and audit into one security posture. This extension strategy helps to ensure that AI systems remain governed by the same access and compliance policies already in place for both human and machine users. Put simply, this approach is designed to support existing policies and reduce security drift.

This white paper explores the security challenges that MCP introduces, including architectural decisions around LLM deployment, governance of data access, and prevention of model manipulation. It also outlines enterprise-ready approaches to securing MCP using principles like task-based authorization, short-lived credentials, and workload observability. Finally, this paper examines how solutions such as Teleport allow enterprises to apply consistent, policy-driven controls across both traditional and AI-powered infrastructure.

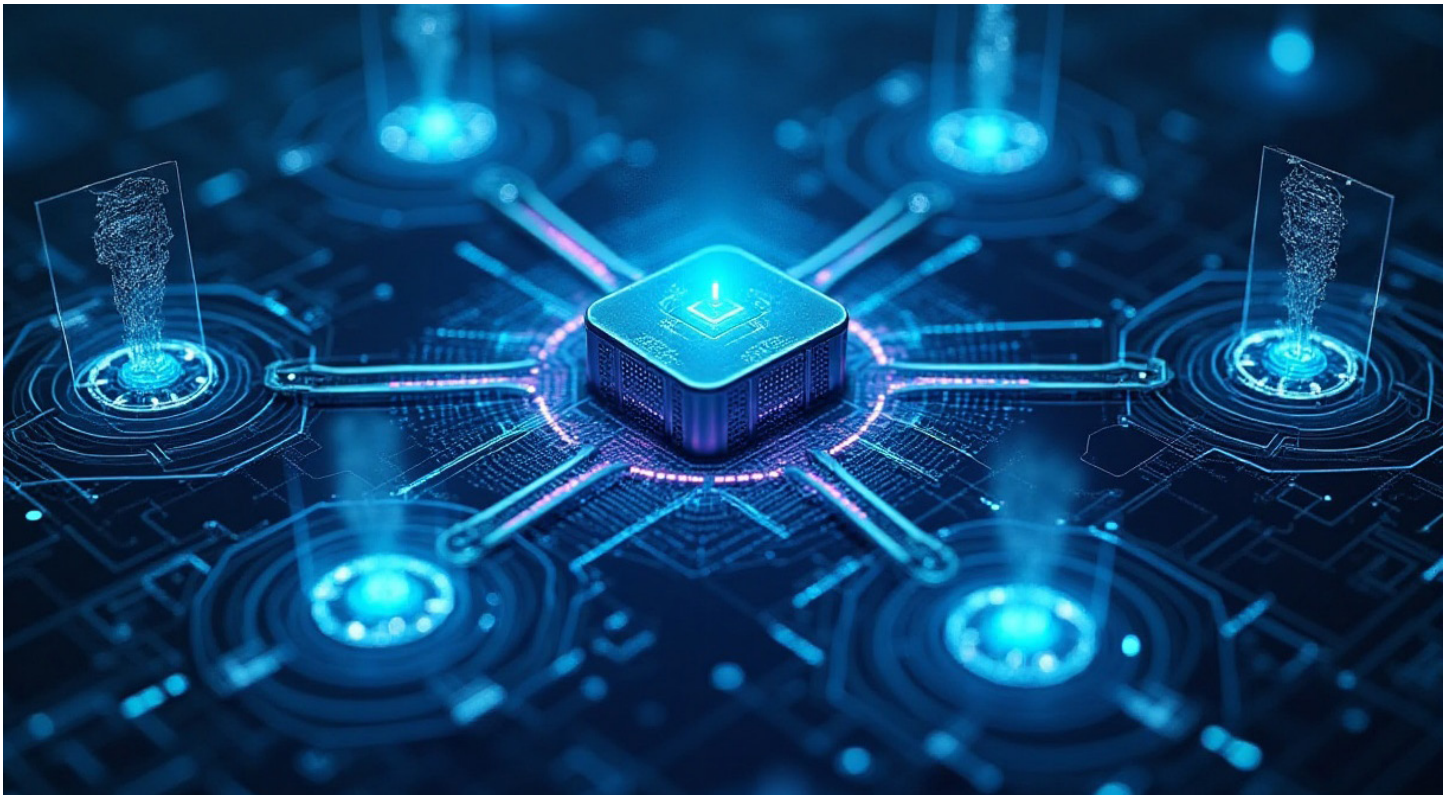
MCP in Enterprise LLM Architectures

Until recently, most LLMs have been acting on data within a closed system. MCP is an open standard designed to allow LLMs to interact with external systems in structured, reliable, and repeatable ways. Unlike conventional APIs that rely on rigid, predefined requests, MCP enables real-time, context-aware exchanges that support agentic workflows. It uses JSON-RPC 2.0 over various transports, including HTTP, SSE, and stdio, and sits in the control plane between LLMs and internal systems.

MCP interactions are defined between two roles: MCP clients and MCP servers. MCP clients are typically LLMs or AI agents. MCP servers, the internal services, tools, or data sources, register specific methods or actions that clients can invoke, similar to how APIs expose endpoints. This architecture provides consistency while giving the service side control over what is callable and how.

The enterprise infrastructure architecture is evolving as enterprises adopt LLMs. Increasingly many organizations are deploying a constellation of smaller, task-specific LLMs or even Small Language Models (SLMs) rather than a single centralized model. This trend is driven by the need for lower cost, improved performance, and better alignment with business needs. However, this shift increases the complexity of managing AI and requires organizations to carefully think through their LLM deployment architecture. Early adopters are integrating the smaller models with internal systems via MCP, an emerging trend, which creates new risks across access control, authorization, and policy enforcement. Until fairly recently, security has been undefined within MCP. The OAuth protocol, which governs secure authorization for delegated access, plays an important role in protecting web and service interactions. Notably, OAuth was only recently added to the MCP specification on March 26, 2025. Treating LLMs as first-class workload identities with distinct policies and scopes is critical.

MCP acts as a mediator between the LLM and the tool or data source it is attempting to use. This layer allows the enterprise to standardize how LLMs invoke tools, make queries, or retrieve structured data. MCP benefits enterprises by providing the standardization of AI-tool interactions, facilitation of agentic AI systems, and AI integration with existing applications. Common MCP use cases include AI-driven support assistants, internal knowledge queries, and software development assistants. MCP is gaining adoption from major AI infrastructure players such as Anthropic, OpenAI, and Cloudflare. However, MCP security practices are uneven and often immature across enterprises.



MCP Security Challenges

Securing MCP requires recognizing that the risks it introduces span three distinct categories: deployment architecture risks, authorization and governance gaps, and LLM manipulation threats.

1. Deployment Architecture Risks

- Overprivileged access: Many MCP servers request broad, persistent access to systems or data.
- Expanded attack surface: Centralized tokens and loosely configured endpoints increase the likelihood of data sprawl and shadow access. This is exacerbated by publishing MCP servers openly to the internet using bearer token authentication, effectively creating backdoors at scale.
- Supply chain exposure: MCP servers may use third-party or open-source components with limited security validation.

2. Authorization and Governance Gaps

- Static credentials: Long-lived secrets remain in circulation far too often.
- Inconsistent policy enforcement: Access control for LLMs is often implemented ad hoc and sometimes not at all.
- Limited visibility: Many environments lack sufficient logging or insight into AI-driven resource usage. Numerous organizations have no visibility or even know how many MCP servers are in their environment. Compounding this is a lack of discoverability and higher-level visibility, creating challenges in understanding which teams use MCP the most and identifying anomalies in its usage. These issues also arise in classic workflow involving human users.

3. LLM Manipulation Threats

- Prompt injection: Malicious input can alter model behavior or trigger unauthorized actions. LLMs may invoke the same action twice, get stuck in a loop, or misapply parameters. Organizations need access requests for elevated privileges for LLMs and agents.
- Context leakage: Poor session management or segmentation allows data to bleed across users or tools. Many MCP servers today don't leverage user identity, obscuring who has requested or created the action.
- Organizations should prioritize foundational controls such as access governance, role-based authorization, and audit visibility, particularly in the context of integrating LLMs into existing infrastructures. These controls form the basis for securing MCP workflows as they mature.

Consequences of Insecure MCP Deployments

Enterprises that fail to secure their MCP deployments risk introducing blind spots into their broader infrastructure security strategy. Because MCP acts as a real-time broker between AI agents and enterprise systems, any lack of access control, auditability, or policy enforcement can lead to unexpected exposure.

Common risks include data exfiltration, inappropriate access to sensitive systems, and inconsistent authorization paths across LLM-driven tools, which can lead to increased security debt, compliance violations, and even reputational damage. These risks are not hypothetical; they represent practical consequences when LLMs are allowed to operate without guardrails.

JPMorgan Chase's CISO, Patrick Opet, recently noted that prioritizing speed over security in AI deployments creates risk not only for individual firms but for customer ecosystems at large. He stated that the "pursuit of market share at the expense of security exposes entire customer ecosystems to significant risk." He advocates for "secure-and-resilient-by-default architectures" and "provable controls" to ensure robust security postures.

Governance gaps in emerging interfaces like MCP can undermine trust, create lateral movement vectors, and break audit continuity. As enterprises increase investment in AI, MCP must be treated as part of the production infrastructure and not as an experiment. That starts with acknowledging what can go wrong.

Recommendations for Securing MCP and LLMs

Enterprises should approach MCP security by considering AI agents, human users, and machine identities within a unified security framework. This framework should extend existing access, authorization, and audit policies to AI workloads, rather than requiring the development of a standalone security model. A separate, siloed security approach to AI only increases operational complexity and introduces the potential for policy drift.

- **Use Infrastructure Identity, Not Ad Hoc Controls**
  Treat LLMs and MCP servers as first-class identities. Extend identity-based access control, authorization, and audit to AI workloads rather than creating a parallel security framework.
- **Eliminate Static Credentials**
  Avoid long-lived tokens and passwords. Where possible, use short-lived credentials and mutual authentication to reduce exposure.
- **Implement Task-Based Authorization**
  AI agents often act on behalf of users or departments. Policies should reflect what an agent is allowed to do, not just who initiated the request. Role scoping by task improves governance. Just-In-Time (JIT) access also plays a crucial role; most MCP servers should grant read-only access by default, with JIT access via the LLM capable of elevating privileges when necessary.
- **Apply Fine-grained Access Control**
  Use RBAC and ABAC to tightly control what each agent can access. Define access boundaries at the data source, table, or method level.
- **Introduce Data Governance into AI Workflows**
  Segment data and apply classification labels. Ensure that LLMs only access data aligned with their role or scope. Promote data stewardship practices across teams.
- **Make Logging and Auditing Mandatory**
  MCP interactions should be logged with identity context and metadata. Export logs to supporting platforms for incident response and compliance. Monitoring and anomaly detection are also key.
- **Continuously Test and Validate MCP Integrations**
  Conduct routine code reviews, dependency checks, and vulnerability scans on MCP connectors. Treat them with the same care as production microservices.
- **Educate Developers and Security Teams Together**
  Create shared accountability by embedding security education into AI development cycles. Provide guardrails without blocking experimentation.

By aligning MCP security with broader infrastructure practices, organizations can ensure consistent and scalable protections as AI adoption grows.

Teleport and MCP Security in the Enterprise

One vendor offering credible solutions for securing MCP workflows is Teleport, a platform that applies Infrastructure Identity principles across traditional and AI systems. Teleport is especially relevant for enterprises looking to avoid fragmented security models as they scale LLM usage.

Based on HyperFRAME Research's analysis, Teleport allows enterprises to bring MCP workflows under the same access, authorization, and audit policies that already govern traditional infrastructure. Rather than introduce a parallel security architecture for LLMs, organizations can extend their Infrastructure Identity approach to include MCP clients and servers.

Teleport is designed to treat these AI-driven components as first-class identities. This means applying mutual Transport Layer Security (TLS) authentication between LLMs and protected resources, enforcing granular role-based access control (RBAC) and attribute-based access control (ABAC) policies, and ensuring that every interaction is logged with identity context and full audit metadata. These audit logs capture both successful and denied access attempts, including the identity of the initiating user or agent and the requested action. This ensures complete visibility for security and compliance teams.

Teleport also supports short-lived, just-in-time credentials. This eliminates the need to persist secrets across multiple toolchains or LLM sessions while reducing the risk of token sprawl. For organizations already relying on centralized identity providers and security information and event management (SIEM) platforms, Teleport integrates natively, enabling consistent policy management and observability.

Effective governance of AI assistants interacting with internal systems via MCP needs a comprehensive control strategy. This strategy should address:

1. **User Access to AI Assistants:** Determine which individuals or teams are authorized to interact with specific AI assistants.
2. **AI Assistant Access to MCP Servers:** Define which MCP servers each AI assistant can communicate with and ensure that assistants only access services pertinent to their function.
3. **MCP Server Access to Data:** Specify the exact data sets or resources that each MCP server can access or edit.

This approach ensures that at every stage, access is meticulously controlled and audited. For instance, enterprises can:

- Restrict which internal teams can access a particular MCP server exposed to an LLM-based support assistant.
- Allow LLMs to issue scoped queries to protected databases with traceable, auditable access.
- Intercept and validate MCP requests as they pass through identity-aware proxies.

By supporting MCP as a first-class integration, Teleport helps ensure that LLM workloads follow the same policy and compliance rules as other workloads in the environment. This unified approach is not merely about reducing fragmentation or accelerating deployment; it's essential for effective governance. Without considering AI agents alongside human and machine users, organizations risk inconsistent enforcement, policy gaps, and audit failures. Treating AI as an isolated entity undermines the ability to maintain control, trust, and compliance.

This reduces fragmentation, improves governance, and accelerates the path from pilot to production.

Looking Ahead

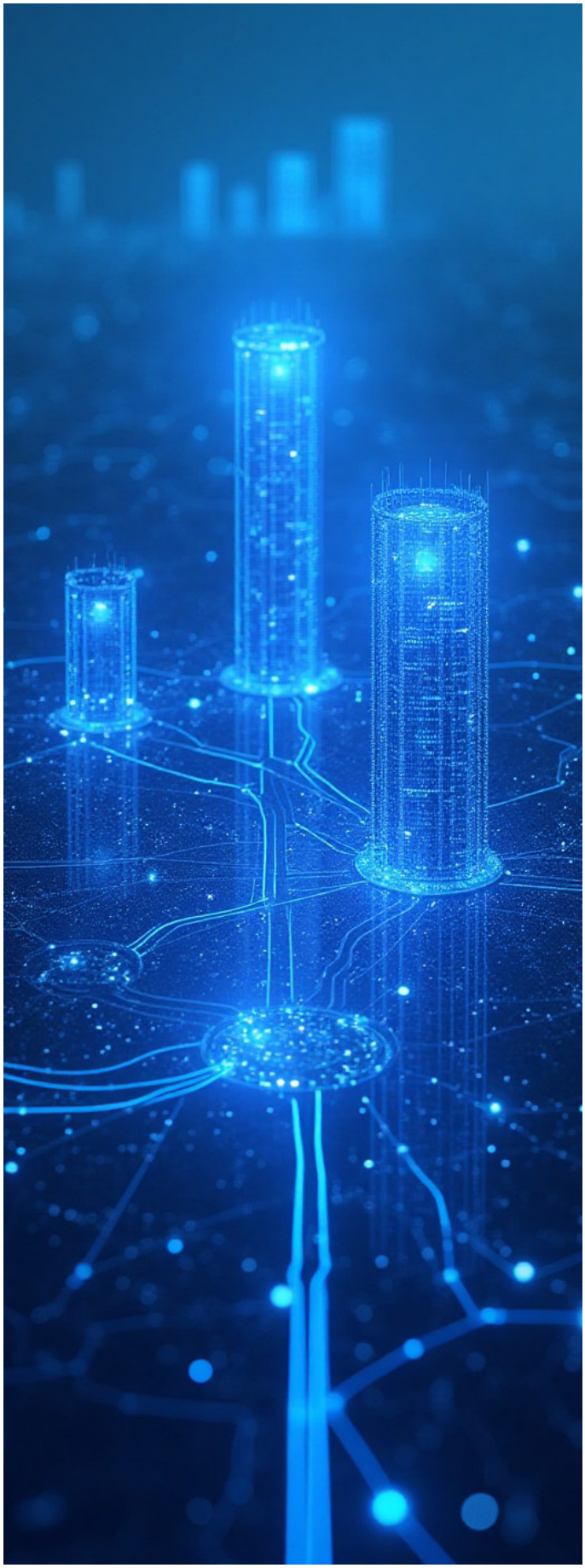
The Model Context Protocol is enabling a new wave of LLM and SLM driven innovation inside the enterprise. By providing a standard interface between models and systems, MCP simplifies integration. However, it also creates new surface areas for authorized access, policy violations, and governance failures.

Many enterprises are still in the early stages of understanding what it means to operationalize AI security. Without consistent identity, access, and audit controls, MCP deployments risk becoming blind spots within otherwise mature infrastructure.

Rather than managing AI as an exception, organizations should bring AI into the fold of their existing security strategy, like Teleport and its support for Infrastructure Identity. Treating MCP clients and servers as first-class entities makes it possible to apply the same controls used for human and machine access, which avoids the need for fragmented policies or siloed toolchains.

Securing MCP is not only about protecting sensitive data or ensuring compliance. It is about building trust in the systems that will increasingly power core business functions. AI will not remain a test environment for long. Enterprises that invest in securing MCP now will be better positioned to scale safely, innovate, and unlock the full value of AI across the stack.

Given this wider context and the pressing security challenges that result, HyperFRAME Research strongly recommends that enterprises evaluate their needs relating to model security and access. In the context of this evaluation, HyperFRAME Research also recommends that enterprises consider approaching vendors such as Teleport that offer comprehensive solutions designed to build on existing frameworks, but that also secure LLMs and MCP deployments.



ABOUT HYPERFRAME RESEARCH:

HyperFRAME Research delivers indepth research and insights across the global technology landscape, spanning everything from hyperscale public cloud to the mainframe and everything in between. We offer strategic advisory services, custom research reports, tailored consulting engagements, digital events, go to market planning, message testing, and lead generation programs.

Our industry analysts specialize in rigorous qualitative and quantitative assessments of technology solutions, business challenges, market forces, and end user demands across industry sectors. HyperFRAME Research collaborates closely with your Analyst Relations, Product, and Marketing teams to build and amplify your thought leadership, positioning your expertise to enhance brand and product recognition. Through content that engages readers, viewers, and listeners alike, we ensure your voice resonates across channels.

CONTACT HYPERFRAME RESEARCH:

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

Email Address:

steven.dickens@hyperframeresearch.com

Telephone Number:

+1 845 505 1678

X: - @StevenDickens3

LinkedIn: Steven Dickens

BlueSky: Steven Dickens

CONTRIBUTORS

Stephanie Walter
Analyst In Residence - AI Tech Stack

INQUIRIES

Contact us if you would like to discuss this report and HyperFRAME Research will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "HyperFRAME Research." Non-press and non-analysts must receive prior written permission by HyperFRAME Research for any citations

LICENSING

DISCLOSURES

HyperFRAME Research provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.


HyperFRAME
RESEARCH