Prefactor

Redefining Digital Trust

How MCP is Securing the Next Generation of AI Agents

A structural handover is underway. Systems once designed solely for humans are fast evolving to become ecosystems of collaboration between humans and NHI's. Much of 2025 has been about testing the potential of agents, and we're now moving fast towards managing them at scale as they begin to act, decide and build on our behalf.

The quiet shift happening underneath is the transformation of Identity and Access Management (IAM) to Agent Identity and Access Management (AIAM), with MCP emerging rapidly as the anchor protocol, and 2025 marking the point where agent identity outpaces human user identity. But whilst MCP defines how agents communicate across ecosystems – it doesn't define how they're verified, authorised or audited. It solves interoperability, not identity. The next evolution in identity management is MCP-native authentication, ultimately delivering audit-grade trust for mixed human-agent ecosystems and Prefactor is pioneering this evolution.

So what is driving the evolution to a new agent and identity access management category? There is an increasingly large share of today's work that is not being done by humans – Salesforce's New Agentic AI Index reports a 119% surge in enterprise agent adoption in the first half of 2025, with 79% of senior executives surveyed deploying AI agents operationally[1]. Of 300 executives that responded to a 2025 PWC survey, 79% indicated that agentic AI is already being adopted in their business[2].

SaaS platforms in particular are experiencing the pace of transformation faster than others, due to an exponential rise in the number of autonomous AI agents as active users. Increasingly, these agents aren't built by the SaaS provider, but are hosted externally and operate on behalf of users – accessing the platform directly as their representatives, executing workflows and interacting with data.

1. 2025 SFDC Agentic Enterprise Index, found at: https://www.salesforce.com/news/stories/agentic-enterprise-index-insights-h1-2025 2. 2025, PWC 2025 AI Agent Survey found at:https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-agent-survey.html

Every SaaS app will soon have potentially thousands of agents undertaking complex work – making decisions and interacting with multiple systems. **AI agents implementing and autonomous workflows are fast becoming first-class citizens with identities of their own.**

It's this intense growth that's causing the emergence of AIAM, an acknowledgement that AI systems no longer merely assist users – they act independently, collaborate, negotiate and orchestrate workflows at machine speed. Agents are multiplying faster than legacy identity systems can keep up, operating without verifiable identity, audit or control and we're seeing a friction point appear – where organisations are stuck between the need to innovate, and the need for governance. According to Matt Doughty, CEO of Prefactor, enterprise security is entering a period of rapid transformation. "This is the next great scale challenge in enterprise security – traditional IAM wasn't designed for this scale or speed. Solving it requires rethinking identity and control for an agent-driven world."

Why Traditional Identity & Access Management Fails for Platforms

Traditional Identity and Access Management systems were designed through the lens of humans as the primary digital actors – static users with predictable identities, slowly changing roles and workflows based on manual provisioning and reviews. These systems focus on controlling access through fixed credentials, roles and compliance checks that operate on relatively slow, periodic cycles. They rely on MFA, CAPTCHAs, and human-dependent workflows that agents can't complete – forcing engineering teams to bypass identity controls and hardcode credentials instead. The result: unmanaged, ungoverned, invisible risk – a security and operational liability that will cost organisations millions in remediation, compliance and lost trust. But SaaS platforms today increasingly expose their ecosystems to autonomous AI agents that operate at machine speed, and act across heterogeneous trust boundaries. Put succinctly by OpenID : 'An agent's identity must be portable and verifiable to a third party that has no visibility to its host environment'.[3]

Agents require a fundamentally different identity model, due to:

> **Autonomous, non-deterministic** agent behavior: AI agents take autonomous actions based on real-time decisions, learning from context and adapting dynamically – flexibility that defies the static permission models common in IAM.

> **Vulnerability to context poisoning and prompt injection:** Malicious inputs disguised as legitimate prompts can manipulate generative AI systems into leaking sensitive data, spreading malware and stealing data.

> **Agent behaviour can be unpredictable:** Introducing the risk of unauthorised access and data exposure. Their dynamic nature requires fine-grained and short-term permissions based on context.

> **Cross-domain asynchronous operation:** Many SaaS platforms integrate with multiple third-party services and federate access across domains. Agents operating asynchronously across these boundaries inherently require identity frameworks that are portable and inherently built to work within delegated permission frameworks established by OAuth2.1.

> **User impersonation and accountability gaps:** Without true delegated authority and clear 'on-behalf-of' flows,visibility becomes a problem – platforms risk accountability failures and regulatory issues. Audit trails are critical, but non-existent if using IAM platforms.

> **Rate limiting and abuse prevention:** Agents act fast, operating at machine speed and scale, and potentially executing thousands of actions in seconds. Existing IAM systems may not be capable of handling this speed and scale, potentially degrading service performance and causing unnecessary (and unpredictable) costs.

3. 2025, OpenID, 'Identity Management For Agentic AI', found at: https://openid.net/wp-content/uploads/2025/10/Identity-Management-for-Agentic-AI.pdf

The Risk and Cost of Repurposing Legacy Identity Models

To understand why this redefinition of identity management is moving so rapidly, it's worth zooming out to view the broader market landscape. IBM's 2025 Cost of Data Breach Report begins quantifying the AI-driven risk landscape – and the stats are sobering. Many organisations are sacrificing security and governance in their rush to adopt AI – of those who suffered AI-related breaches, a shocking 97% lacked proper AI access controls and 63% have no formal AI governance policies[4].

These stats begin to paint a picture of traditional identity models failing to contain risks in agent-driven scenarios. SaaS platforms hosting AI agents face risks that extend beyond legacy data breaches into systemic trust erosion, platform instability.

One of the most impactful breaches of 2025, involved Salesforce and its Salesloft-owned AI chatbot integration, Drift, which other companies embed on their websites. While this was a traditional API integration, the ease with which data was extracted hint at the size of the problem when agents become more widespread.

> **"Many organisations are sacrificing security and governance in their rush to adopt AI – of those who suffered AI-related breaches, a shocking 97% lacked proper AI access controls and 63% have no formal AI governance policies.[4]"**
>
> **IBM's 2025 Cost of Data Breach Report**

Between August 8 and August 18, a attackers compromised OAuth tokens held by Drift, a chatbot that customers embed on their website. They gained unauthorised access to hundreds of Salesforce customer environments, extracting sensitive data including contact information, support cases and credentials for other cloud services.

The breach didn't stop at Salesforce's systems though – attackers were able to move across platforms including Slack, Google Workspace, Amazon S3 and Microsoft Azure, underscoring the breadth of impact an agent identity failure can enable.[5,6]

Despite Salesforce's swift remediation actions – which included revoking all affected OAuth tokens and disabling Drift integrations – the operational disruption, reputational damage and compliance fallout reveal a harsh reality. Without the ability to control access in a more fine-grained way with human intervention for certain activities and dynamic authorisation based on behaviour monitoring, these sort of breaches will become more common. It exposes the limitations of relying on human-oriented IAM patterns to secure autonomous agent ecosystems. Agents can be connected to multiple systems that may not be aware of each other, and the scope for data extraction can be significant.

AIAM is already evolving rapidly to solve these problems, starting with a fundamental paradigm shift. Prefactor's platform **acknowledges agents as first-class citizens**, needing an identity that is native, dynamic and built into the protocols where agents operate – the MCP layer – not grafted onto outdated frameworks unable to keep pace.

From IAM to AIAM: A Category Shift

The trajectory of AIAM mirrors historical moments of platform transformation. Just as mobile redefined user experience and cloud recast infrastructure, AI agents necessitate a wholesale replatforming of trust. IAM depends on static provisioning cycles, slow approval workflows and manual audits. In an AI-first organisation, such latency collapses collaboration and makes it challenging for organisations to truly capitalise on the transformation-

4. 2025 IBM Cost of Data Breach Report, found at: https://www.ibm.com/think/x-force/2025-cost-of-a-data-breach-navigating-ai ***5.*** *2025, Google Cloud, 'Widespread data theft targets Salesforce instances via SalesLoft Drift': https://cloud.google.com/blog/topics/threat-intelligence/data-theft-salesforce-instances-via-salesloft-drift* ***6.*** *2025, CXToday, 'The SalesLoft Drift Chatbot Is Set to Go Offline After Customers Suffer Big Breaches, found at: https://www.cxtoday.com/customer-analytics-intelligence/the-salesloft-drift-chatbot-is-set-to-go-offline-after-customers-suffer-big-breaches/#:~:text=Drift%20is%20part%20of%20SalesLoft's,data%20of%20SalesLoft's%20customers'%20customers.*

al value of agents. AIAM remedies these gaps by treating every entity – agents, APIs, and humans – as **first-class identities** with programmable trust, enabling real-time access negotiation, continuous policy enforcement and automated lifecycle management tuned for agentic ecosystems. It shifts identity management to the agent layer itself, embedding dynamic, contextual and continuous trust mechanisms into the MCP layer.

Prefactor has operationalised this concept by building directly on MCP, effectively authenticating and governing every entity in the conversation: the agent, the API and the end user. By embedding identity in the same protocol layer where agents act, the identity and authentication mechanisms are built directly into the communication protocol (MCP) that agents use to perform tasks and access resources.

"Prefactor's approach is grounded in our belief that thinking MCP is a nice-to-have integration layer is a mistake. Instead, it's an identity problem at scale. Every agent, every API, every human needs to be authenticated and controlled on the same footing." Simon Russell, CTO Prefactor

Prefactor is agent-centric by design – with a verifiable, governed and isolated identity for every single agent – every autonomous system in your environment is identifiable, accountable and operating within known boundaries. This is the foundation of trust and accountability in autonomous systems and must become the new standard. Each agent has first-class agent identities with rotation, attestation and lifecycle controls. Governance is baked in with auditable, standards-based control over every agent action. Multi-tenant isolation and scoped authorisation safeguard data privacy, and delegated authorisation and least-privilege policies prevent over-exposure. Innovation is accelerated – not hindered – through native integration with existing OAuth / OIDC flows (accelerate deployment), delegated human-to-agent access (maintain least-privilege trust at scale), and CI / CD-driven policy updates (enabling rapid iteration without downtime).

By elevating agents to true first-class citizens in the identity ecosystem, teams have the freedom to innovate and deploy new AI capabilities at full velocity without sacrificing control or compliance.

Prefactor's approach is grounded in our belief that thinking MCP is a "nice-to-have" integration layer is a mistake. Instead, it's an identity problem at scale. Every agent, every API, every human needs to be authenticated and controlled on the same footing.

Simon Russell, CTO Prefactor

It equips SaaS platforms to extend digital trust, accountability and oversight to autonomous software actors with the same rigour applied to human users, ensuring every action is attributable and subject to the same standards of transparency and control that underpins modern digital platforms. This enables the unified governance of agents, APIs and human users within a comprehensive and cohesive trust framework.

The future SaaS platform won't have a handful of AI assistants – it will orchestrate thousands of specialised agents that communicate across systems. Each agent carries distinct credentials, operational contexts and policy requirements. Managing this multiplicity with legacy identity tools leads to uncontrolled access sprawl and governance gaps. AIAM platforms like Prefactor remove the friction of siloed identities, creating a shared identity fabric where agents collaborate securely, contextually and autonomously.

"Early masters of MCP-native identity position themselves for significant advantage", predicts Matt Doughty. "The winners will be those who **can iterate and innovate at relentless speed.** Those who fail to adopt will spend their resources managing breaches instead of driving innovation".

Auth's Evolution: What Comes Next

Prefactor's mapping of MCP's evolution spotlights three accelerating forces reshaping how agent identity functions at scale. This same terrain will be included and mapped in our upcoming Agentic Index Annual Report, a market-first study examining adoption, penetration, maturity and emerging standards that define how organisations are learning to work with agents.

1 **Supertools and Deeper Metadata:** The MCP ecosystem currently describes tools in flat lists without hierarchy. Prefactor anticipates supertools – meta-structures grouping related functions with shared context and permissions. This evolution mirrors AIAM's approach of aggregated, role-bound capabilities, where agents gain access to modular tool suites governed by contextual identity.

2 **Seamless Enterprise Auth:** Prefactor predicts federated MCP authentication across enterprise boundaries, where users and agents traverse ecosystems without repeated credentialing. Borrowing from SSO, this allows cross-domain trust under unified governance. For AIAM, that means a world where agent identities move securely between organisations, maintaining policy fidelity at every step.

3 **Autonomous Agent Credentials:** Today, MCP auth revolves around human-tethered identity – a structure that will soon break. Despite MCP being adopted as the universal standard for agent communication with external tools, enterprises adopting MCP have found it lacking in basic enterprise governance features, such as enforced authentication, authorisation and audit[7]. Agents must own identity outright. By granting agents independently verifiable credentials, organisations enable truly autonomous operations, such as scheduled reporting, supply chain negotiation, or monitoring tasks performed without user proxies. AIAM will define this shift, where agents authenticate as themselves, with clear provenance and bounded autonomy.

Gartner predicts that by 2028, 33% of enterprise software applications will include agentic AI, up from less than 1% in 2024, and will be making 15% of day-to-day work decisions autonomously[8]. As enterprises adopt agentic systems incredibly fast, the question is no longer whether they can manage trust at scale, but whether they can reimagine identity itself. Prefactor recognised that simply stretching human-centric identity tools to fit agents would never solve the new challenges of autonomy, velocity and scale.

AIAM needs to be a unique, foundational building block that enables agents to participate securely and transparently in digital ecosystems, right alongside humans. Every org must treat agent, API and human identities equally, or risk exposing itself to risks. AIAM is a foundational architectural enabler that allows product and infrastructure teams to capitalise on the transformative benefits of agentic AI without compromising security or control.

7. 2025, Redhat, 'Model Context Protocol (MCP): Understanding security risks and controls', found at: https://www.redhat.com/en/blog/model-context-protocol-mcp-understanding-security-risks-and-controls ***8.*** *2024, Coshow T, 'Intelligent Agents in AI Really Can Work Alone. Here's How', Gartner, found at: https://www.gartner.com/en/articles/intelligent-agent-in-a*