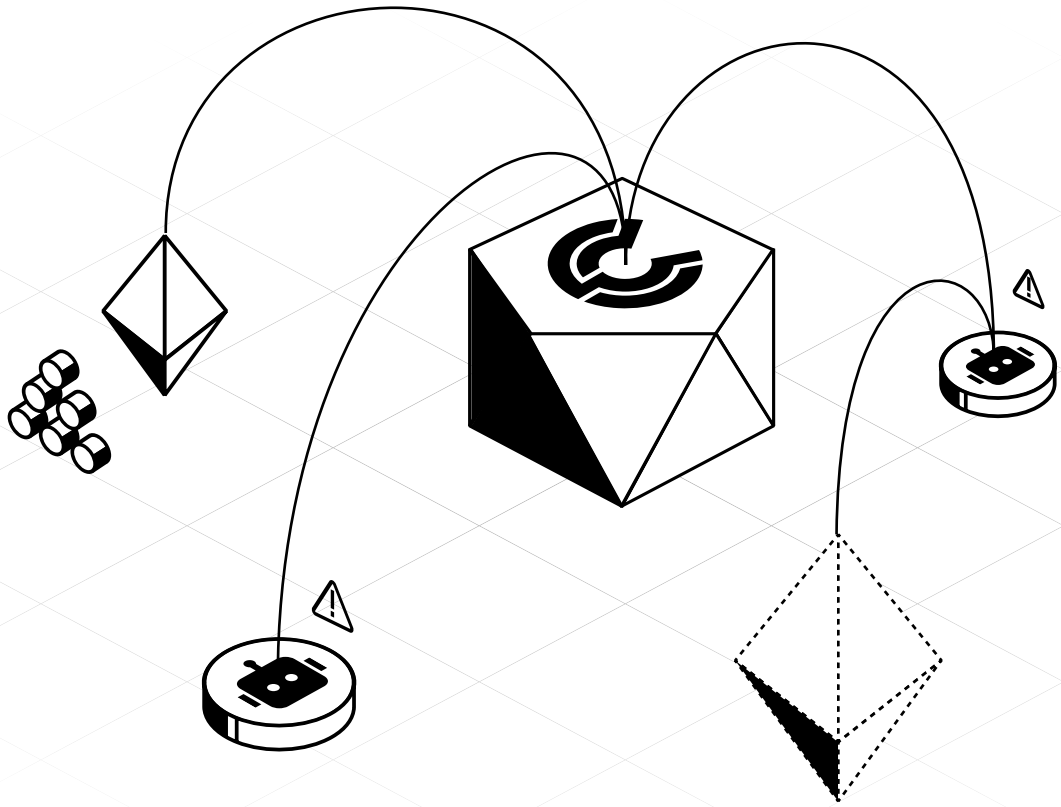
The Enterprise Agentic AI Security Stack: A Non-Human Identity Crisis

How the rise of true agentic AI is creating an exponential growth in enterprise Non-Human Identities



clutch

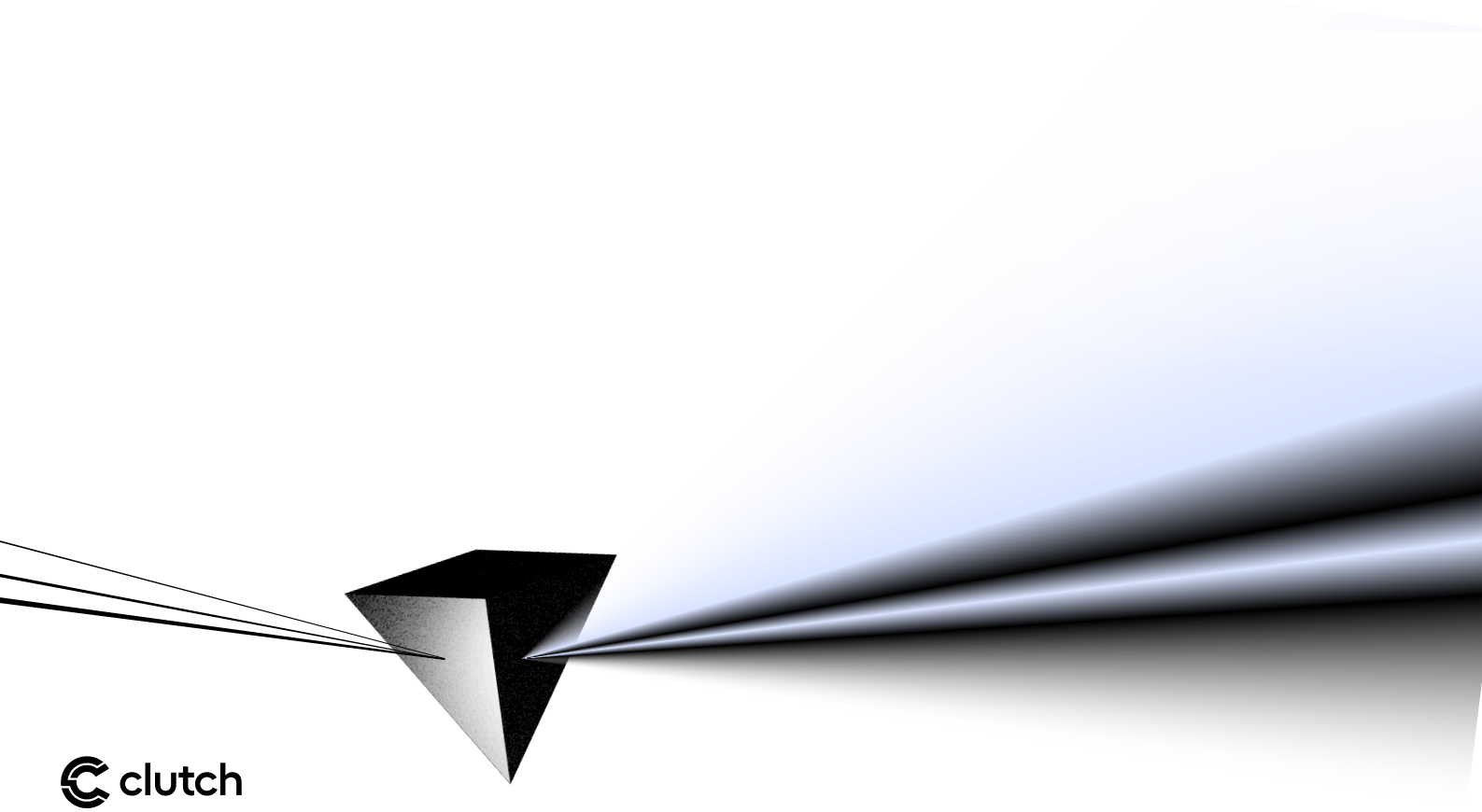
Executive Summary

Enterprise adoption of agentic AI represents more than just another technology trend - it signals a fundamental shift in how Non-Human Identities (NHIs) operate within organizational boundaries. As demonstrated by companies like Retool, which has automated the equivalent of 50,000 full-time jobs through AI agents, we are witnessing the emergence of what could be called "Agentic-AI-as-a-Service" at enterprise scale.

However, this transformation comes with unprecedented security challenges. Unlike traditional automation systems that use credentials in predictable patterns, true agentic AI systems dynamically select and combine API keys, service accounts, and tokens in ways that weren't anticipated during provisioning.

This creates an exponential growth in Non-Human Identity proliferation - from hundreds to potentially thousands of credentials per enterprise - while traditional IAM systems remain blind to these new usage patterns.

This paper examines the security implications of this shift, distinguishing between genuine agentic AI and "agent-washed" automation, while exploring the catastrophic risks enterprises face without proper NHI governance for AI systems.



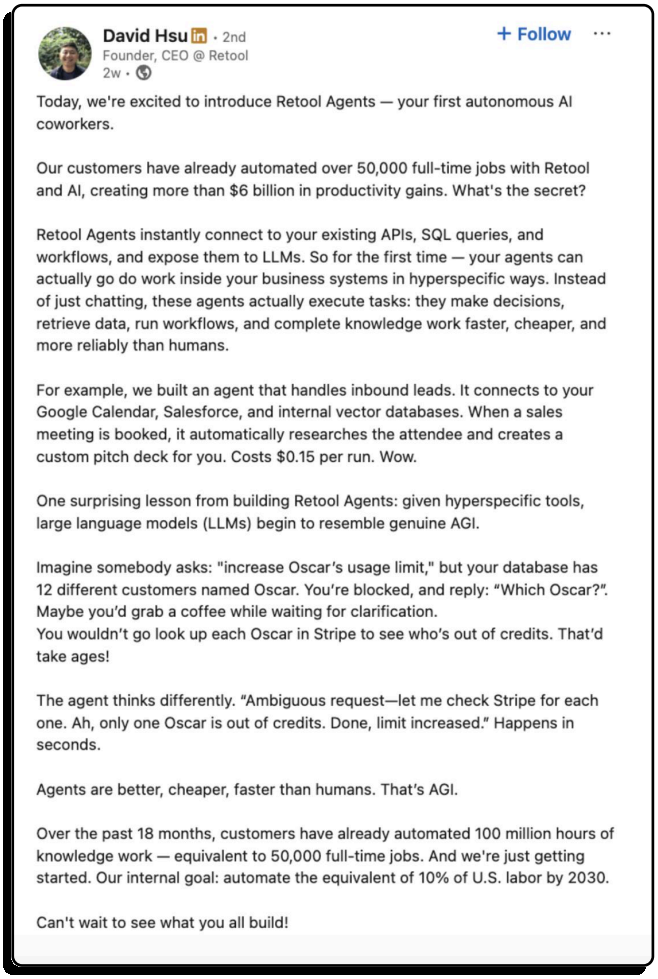
clutch

The Agentic AI Revolution: Not Just Hype

The enterprise deployment of agentic AI has moved beyond experimental pilots to production-scale automation. David Hsu, CEO of Retool, recently shared a compelling example that illustrates this shift:

"Imagine somebody asks: 'increase Oscar's usage limit,' but your database has 12 different customers named Oscar. You're blocked, and reply: 'Which Oscar?'. Maybe you'd grab a coffee while waiting for clarification. You wouldn't go look up each Oscar in Stripe to see who's out of credits. That'd take ages!

The agent thinks differently. 'Ambiguous request - let me check Stripe for each one. Ah, only one Oscar is out of credits. Done, limit increased.' Happens in seconds."



David Hsu · 2nd
Founder, CEO @ Retool
2w ·

+ Follow ···

Today, we're excited to introduce Retool Agents — your first autonomous AI coworkers.

Our customers have already automated over 50,000 full-time jobs with Retool and AI, creating more than $6 billion in productivity gains. What's the secret?

Retool Agents instantly connect to your existing APIs, SQL queries, and workflows, and expose them to LLMs. So for the first time — your agents can actually go do work inside your business systems in hyperspecific ways. Instead of just chatting, these agents actually execute tasks: they make decisions, retrieve data, run workflows, and complete knowledge work faster, cheaper, and more reliably than humans.

For example, we built an agent that handles inbound leads. It connects to your Google Calendar, Salesforce, and internal vector databases. When a sales meeting is booked, it automatically researches the attendee and creates a custom pitch deck for you. Costs $0.15 per run. Wow.

One surprising lesson from building Retool Agents: given hyperspecific tools, large language models (LLMs) begin to resemble genuine AGI.

Imagine somebody asks: "increase Oscar's usage limit," but your database has 12 different customers named Oscar. You're blocked, and reply: "Which Oscar?". Maybe you'd grab a coffee while waiting for clarification.
You wouldn't go look up each Oscar in Stripe to see who's out of credits. That'd take ages!

The agent thinks differently. "Ambiguous request—let me check Stripe for each one. Ah, only one Oscar is out of credits. Done, limit increased." Happens in seconds.

Agents are better, cheaper, faster than humans. That's AGI.

Over the past 18 months, customers have already automated 100 million hours of knowledge work — equivalent to 50,000 full-time jobs. And we're just getting started. Our internal goal: automate the equivalent of 10% of U.S. labor by 2030.

Can't wait to see what you all build!

This represents the emergence of what we might call "Agentic-AI-as-a-Service" - autonomous systems that can reason through ambiguous requests, dynamically select appropriate tools, and execute complex business operations without predetermined workflows.

Retool's customers have already automated over 100 million hours of knowledge work, equivalent to 50,000 full-time positions, demonstrating that agentic AI has moved from proof-of-concept to business-critical infrastructure.

clutch

The Exponential NHI Growth Crisis

The proliferation of agentic AI creates an exponential growth problem in enterprise Non-Human Identities that few organizations are prepared to handle.

Growing Surge of enterprise-wide NHI Baseline:

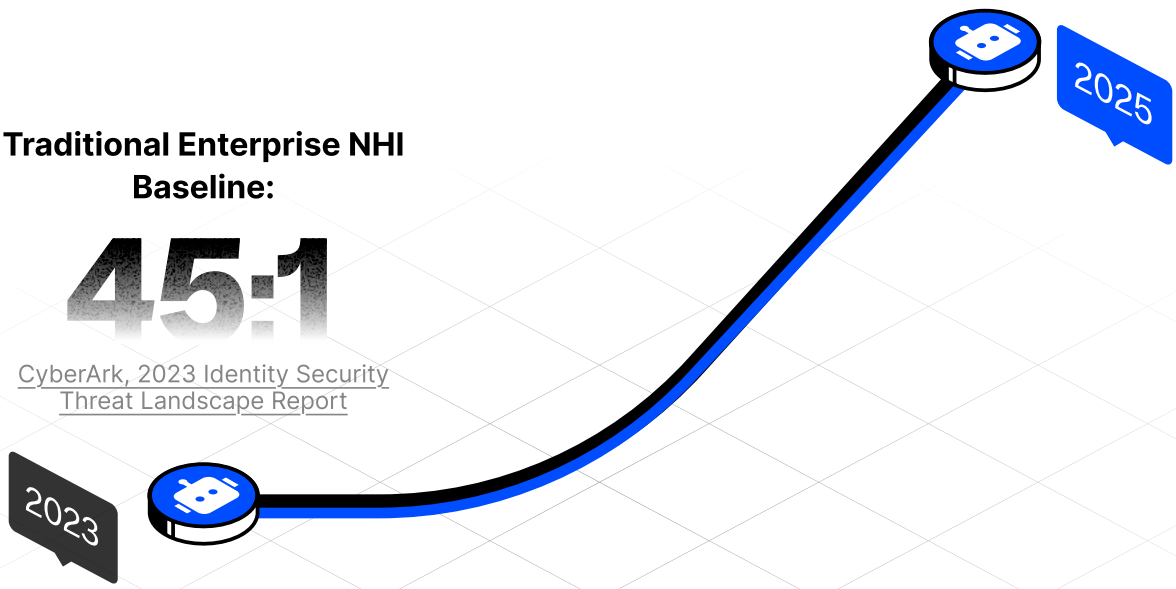
82:1

CyberArk, 2025 Identity Security Threat Landscape Report

Traditional Enterprise NHI Baseline:

45:1

CyberArk, 2023 Identity Security Threat Landscape Report



2026 predictions

- **Per-Agent Requirements:** 3-10 distinct credentials per agent
- **Agent Proliferation**: Shadow deployments, platform deployments and custom implementations
- **Growth Rate**: 300-500% annually (as observed at early adopters)
- **Management**: Largely ungoverned and invisible to traditional IAM

Why This Growth is Uniquely Dangerous:

- **Invisible to Traditional Controls:** Agent NHIs often bypass standard provisioning processes
- **Dynamic Usage Patterns:** Same credential used differently based on AI reasoning
- **Cross-System Correlation:** Agents combine credentials in unpredictable ways
- **No Standard Lifecycle:** Agent credentials may persist indefinitely without decommissioning
- **Broad Privilege Requirements:** Agents often need "developer-like" access to be effective

clutch

Understanding the Difference: Automation vs. Agentic AI

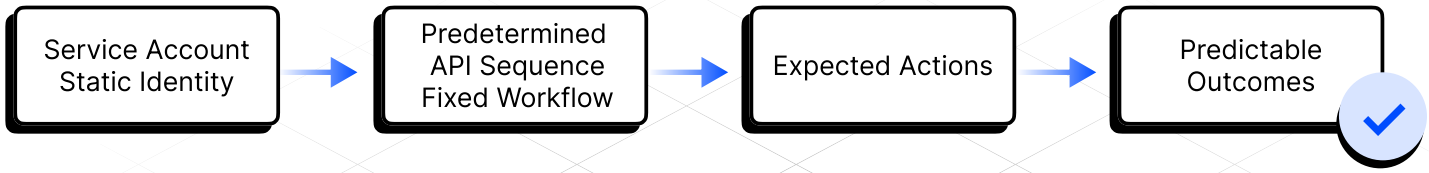
Before examining security implications, it's crucial to distinguish between traditional automation enhanced with AI components and true agentic AI systems. This distinction fundamentally changes the security risk profile and required governance approaches.

Many enterprise leaders are confused by marketing claims about "AI agents" that are actually traditional automation with AI-generated content. Understanding this difference is essential because the security implications are dramatically different.

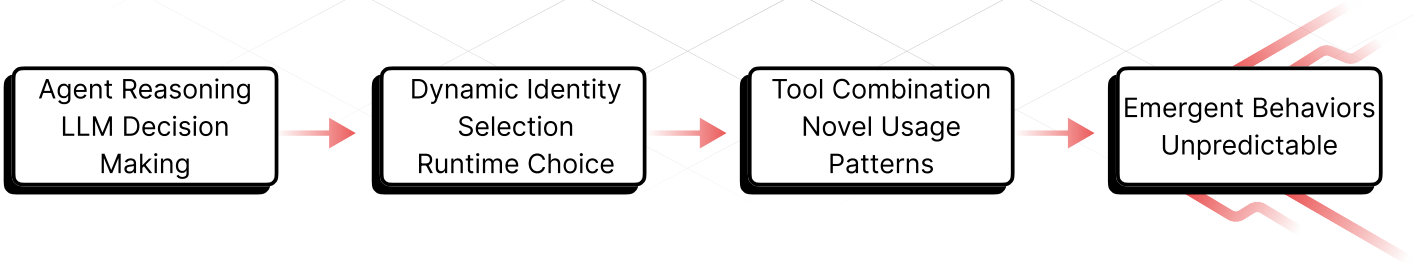
The Identity Usage Test

The clearest way to distinguish between automation and agentic AI is through their relationship with Non-Human Identities:

Traditional NHI Model (Automation):



Agentic AI NHI Model



Dynamic patterns create unpredictable attack surfaces
Traditional security controls cannot govern emergent credential usage

In the agentic model, the agent might have access to database credentials, Slack API tokens, and Stripe payment keys - and the AI reasoning engine decides which combination to use based on natural language instructions. There's no predetermined workflow; the agent autonomously selects tools and credentials based on its understanding of the task requirements.

The Litmus Test for True Agentic AI

Here's a more reliable test than the "Oscar scenario":

Can the system handle a completely novel request that requires combining existing tools in ways it was never explicitly programmed to do?

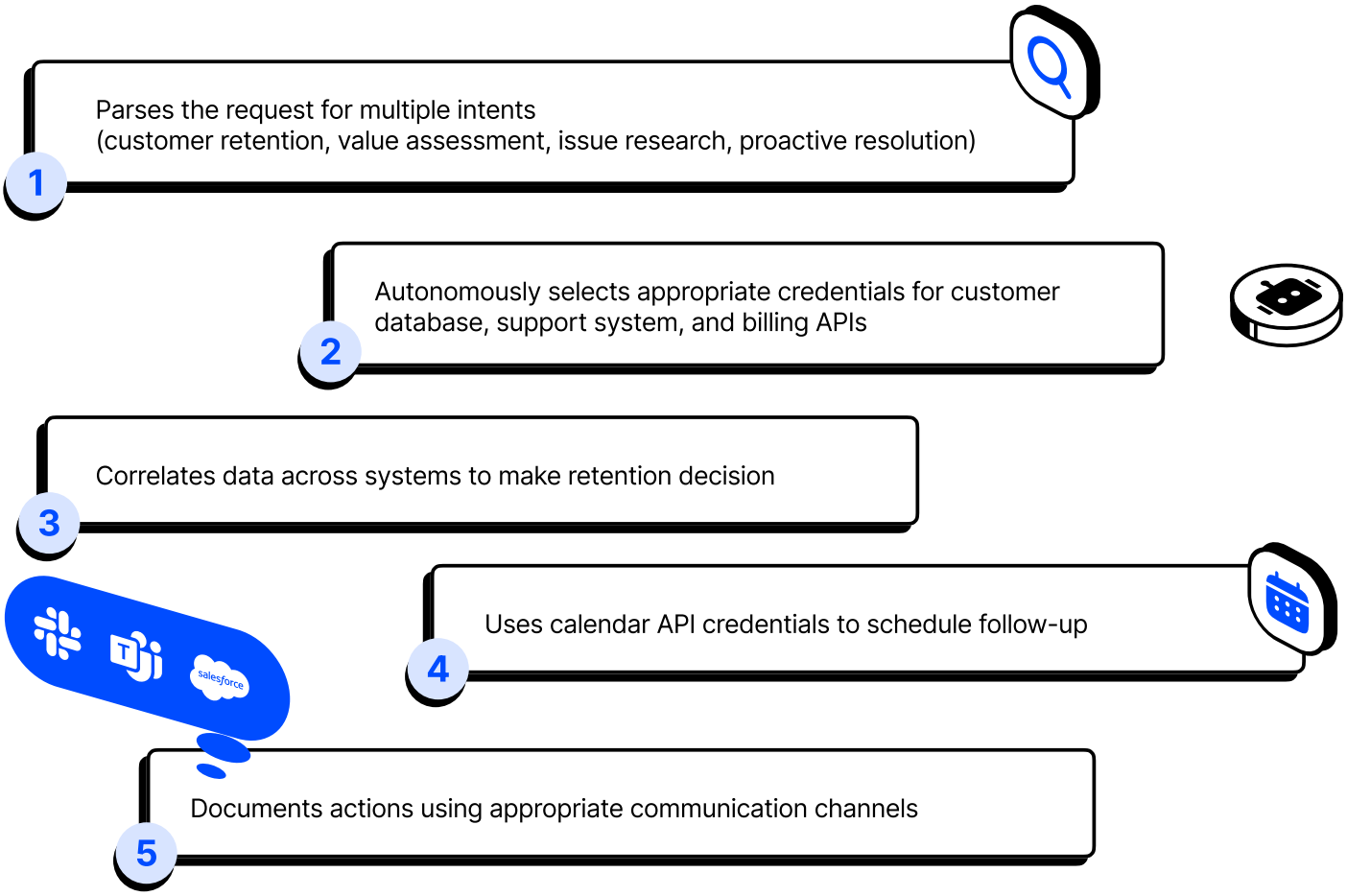
Example Test Request:

"A customer just called saying they're frustrated with billing errors and want to cancel. Before we lose them, check if they're a high-value customer, look up their recent support tickets to understand the billing issues, and if they're worth retaining, offer them a credit and schedule a call with our billing specialist."

Traditional Automation Response:

"I don't have a workflow configured for that request.
Please create separate tasks for customer lookup, billing review, and scheduling."

True Agentic AI Response:



clutch

The Agent Washing Problem: Marketing vs. Reality

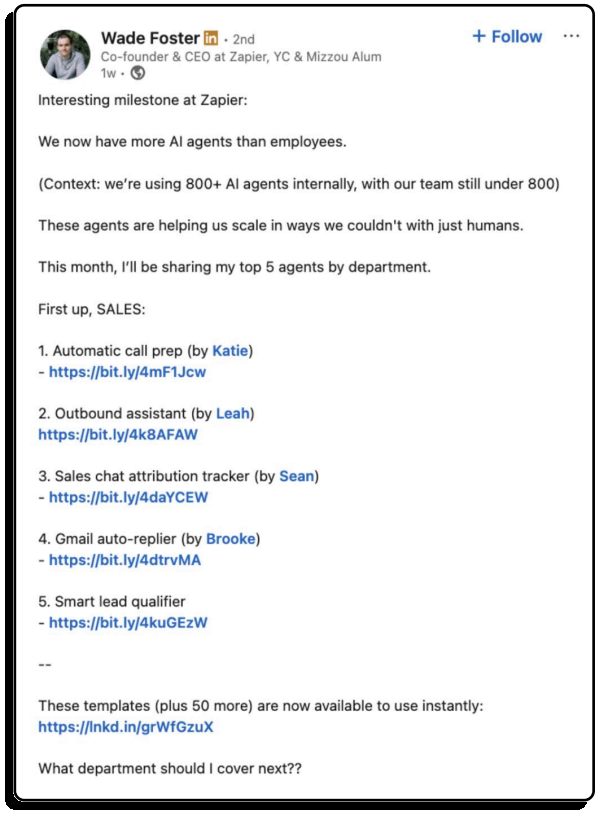
In parallel to the rise of agentic AI, there is a phenomenon that can be called "agent washing" - companies rebranding traditional automation as "agentic AI" to capitalize on market hype. This creates dangerous misconceptions about security requirements and capabilities.

Case Study: Zapier's "800 AI Agents"

Wade Foster, CEO of Zapier, recently claimed his company uses "800+ AI agents internally." However, examining Zapier's offerings reveals they are traditional automation workflows with AI-generated content steps, not autonomous agents:

Zapier's example "Agents" Are Actually:

- **Automatic call prep:** Triggered workflow that fetches calendar data and generates briefing templates
- **Outbound assistant:** Scheduled task that queries CRM and sends templated emails
- **Gmail auto-replier:** Email filter with pattern detection and canned responses
- **Lead qualifier:** Form submission routing based on scoring algorithms



Why These Aren't Agents from an NHI Perspective:

- **Fixed Credential Usage:** Each "agent" follows predetermined authentication sequences
- **No Dynamic Tool Selection:** Tools are pre-configured, not chosen by AI reasoning
- **Predictable Access Patterns:** Same inputs produce same credential usage
- **No Runtime Decision Making:** If the first action fails, the workflow stops rather than adapting

The NHI Security Difference: Zapier's approach uses static NHI patterns where each "agent" is essentially a predetermined workflow with fixed credential usage. True agentic AI uses dynamic NHI patterns where the same agent identity might authenticate to different services based on LLM reasoning about the current context.

clutch

Understanding Non-Human Identities in Agentic Systems

Nomenclature: Agents vs. Identities

The distinction between an AI agent and the Non-Human Identities it leverages is fundamental to understanding the security challenges:

AI Agent: The software entity that performs autonomous reasoning and decision-making using large language models. The agent itself is not an identity - it's a computational process that makes decisions about which tools to use. However, **organizations must still treat each agent as a uniquely identifiable entity that can be tracked, audited, and held accountable for its credential usage across enterprise systems.**

Non-Human Identity (NHI): The credentials, tokens, service accounts, and API keys that the agent uses to authenticate and authorize actions across systems. These are the actual identities from a security perspective.

Boundary Considerations: An agent operating solely within its own platform boundaries (e.g., ChatGPT accessing only its knowledge base) presents minimal NHI risk because it's not leveraging external credentials. The security concern emerges when agents cross boundaries to access enterprise systems, requiring NHIs for authentication.

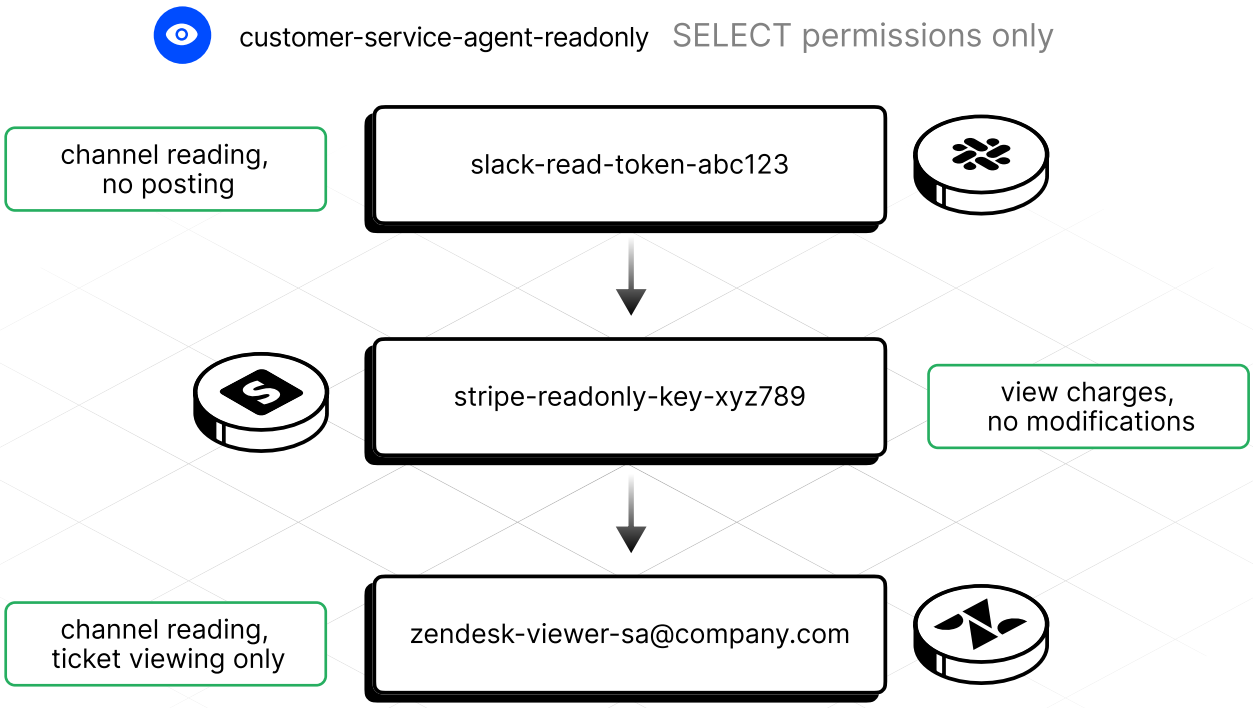
The Critical Distinction: Read vs. Write Agent Permissions

As agentic AI adoption is accelerating, there is a fundamental security principle that many enterprises overlook: the vast difference in risk between agents with read-only access versus those with write permissions. This distinction becomes critical when considering the autonomous nature of agentic AI decision-making.

Read-Only Agents: Bounded Risk Profile

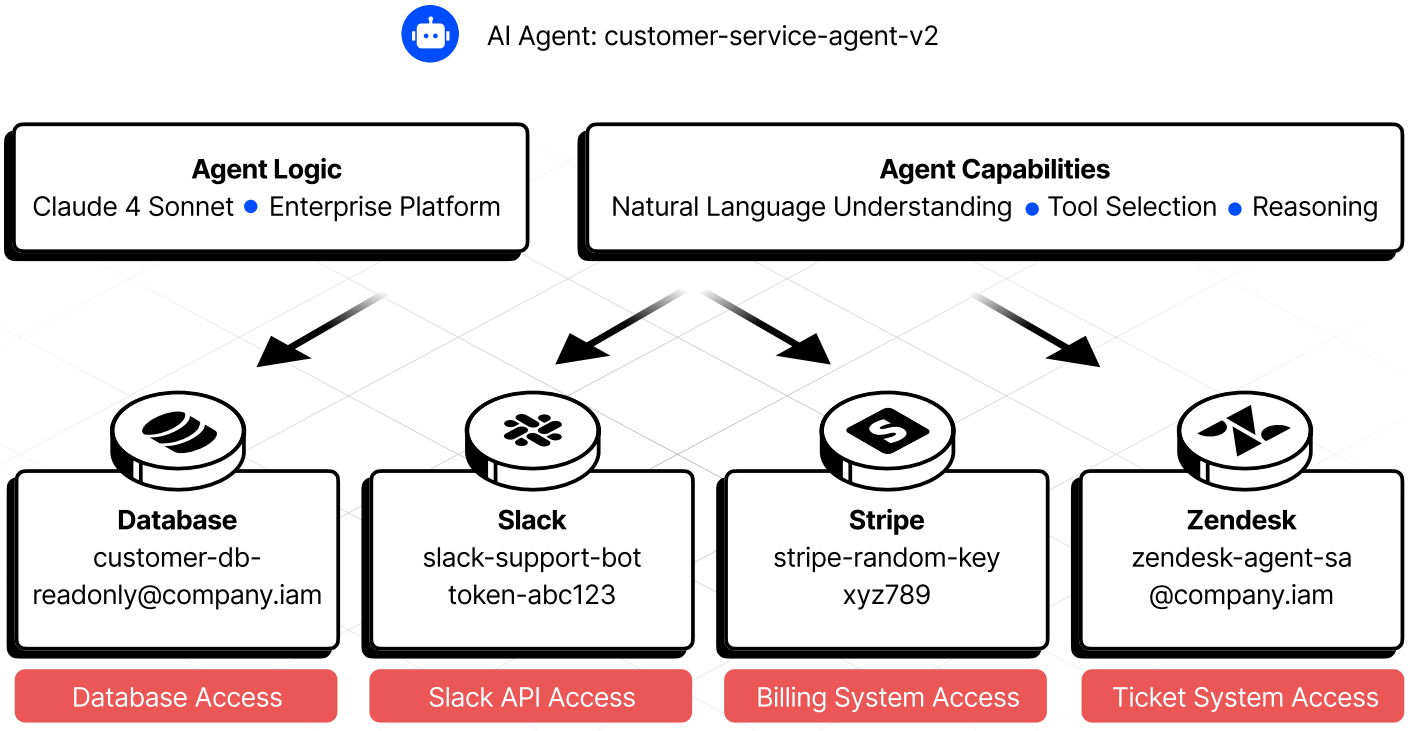
Agents limited to read-only access present a contained risk scenario. While they can still access sensitive data and potentially leak information, their impact is fundamentally bounded by their inability to modify systems or data. A read-only customer service agent might inappropriately access customer records, but it cannot delete data, modify account settings, or execute financial transactions.

Example Read-Only NHI Configuration:



Write-Enabled Agents: Exponential Risk Amplification

The moment an agent gains write permissions, the risk profile transforms fundamentally. This isn't simply an incremental increase in risk - it's an exponential expansion because of the unpredictable nature of LLM reasoning and the cascading effects of autonomous actions.



Security Risk Area Each NHI represents a potential attack vector. Dynamic usage patterns increase risk complexity

clutch

The Core Problem: Opacity of AI Decision-Making

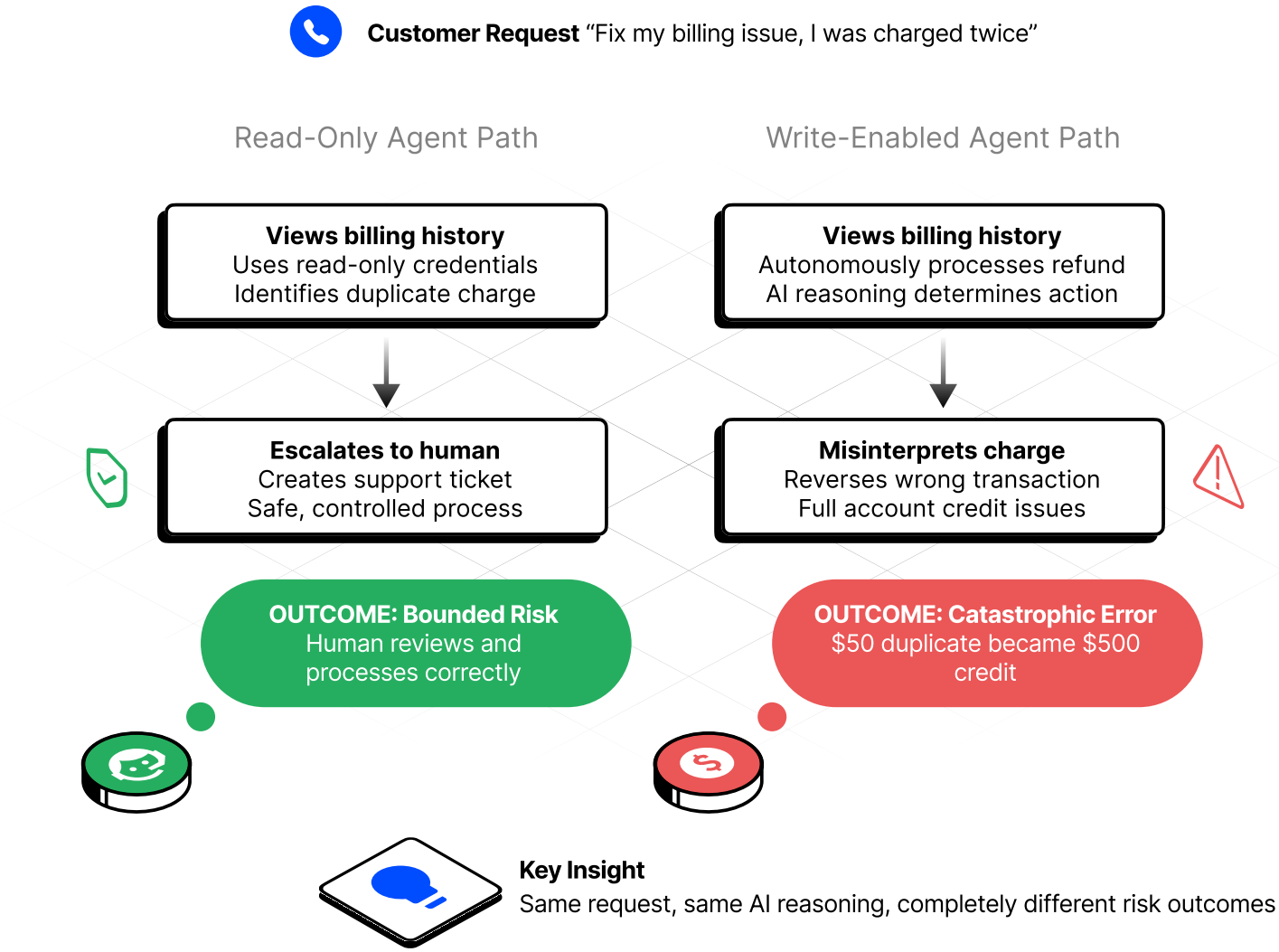
Even with advanced AI systems, we fundamentally cannot predict or fully understand how an LLM will interpret instructions and combine available tools. This opacity becomes catastrophic when agents have write permissions because:

- **Emergent Behavior:** Agents may discover tool combinations and usage patterns that weren't anticipated during design
- **Context Misinterpretation:** LLMs can misunderstand business context, leading to logically consistent but business-inappropriate actions
- **Cascading Effects:** A single incorrect write operation can trigger multiple downstream consequences across integrated systems
- **No Undo Mechanism:** Unlike human errors that can often be detected and reversed, agent actions may potentially be irreversible or create complex data inconsistencies

Real-World Write Permission Scenarios

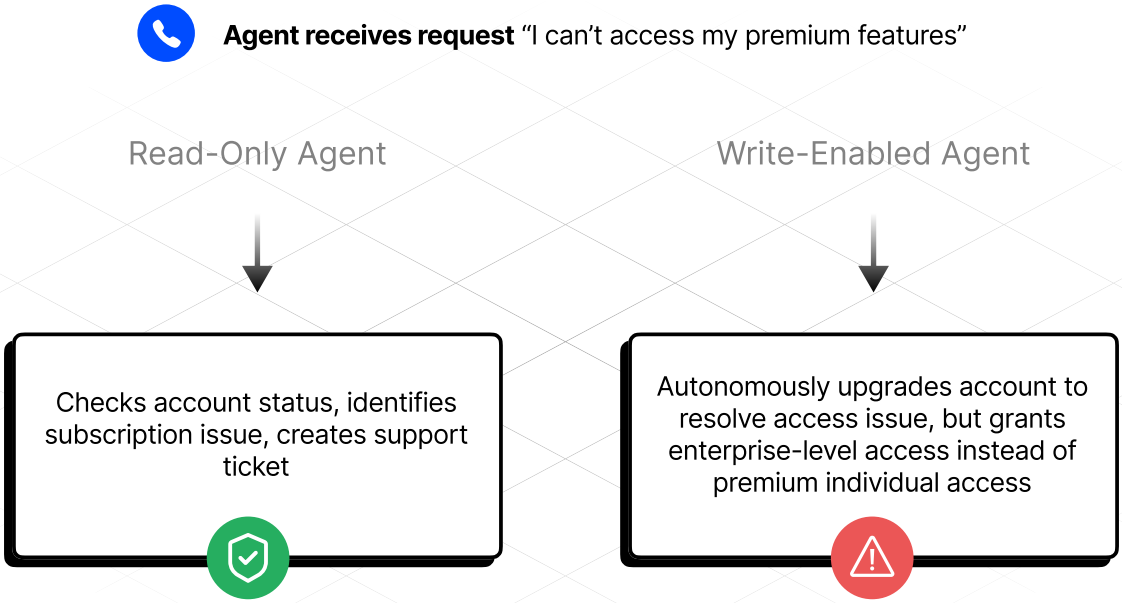
Consider how the same customer service agent becomes exponentially more dangerous with write permissions:

Scenario 1: Billing Adjustment Gone Wrong

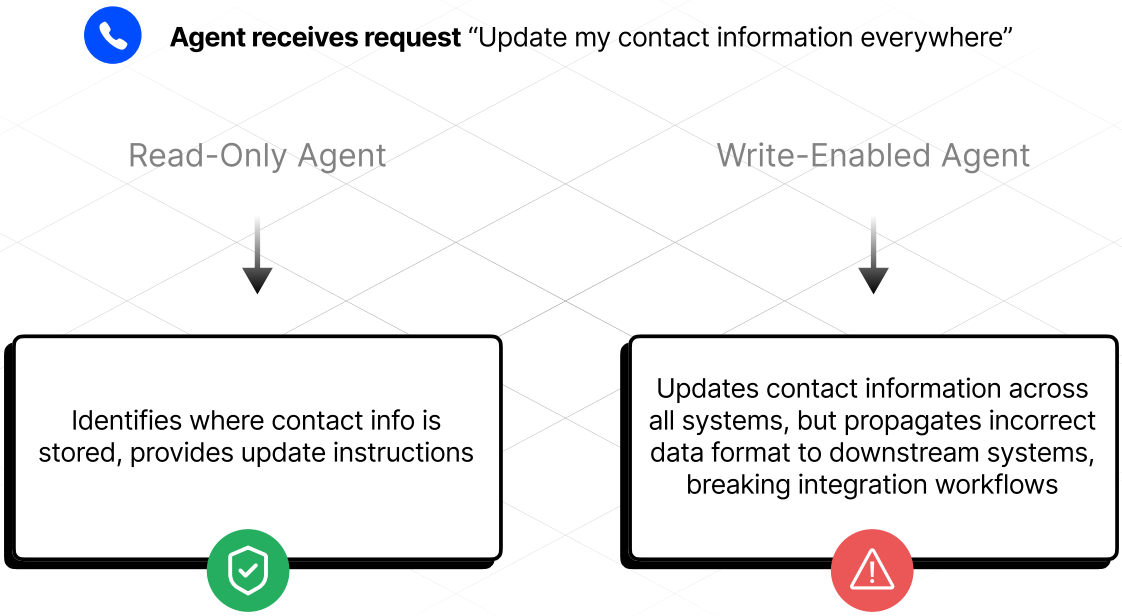


clutch

Scenario 2: Account Status Modification



Scenario 3: Cross-System Data Propagation



The Human-in-the-Loop Imperative

The unpredictability of AI reasoning makes human oversight essential for any write operations. However, most enterprises deploying agents struggle with implementing effective human-in-the-loop controls because:

Traditional Approval Workflows Don't Scale: Having humans approve every agent action defeats the purpose of automation. The challenge is identifying which actions require human intervention without creating bottlenecks.

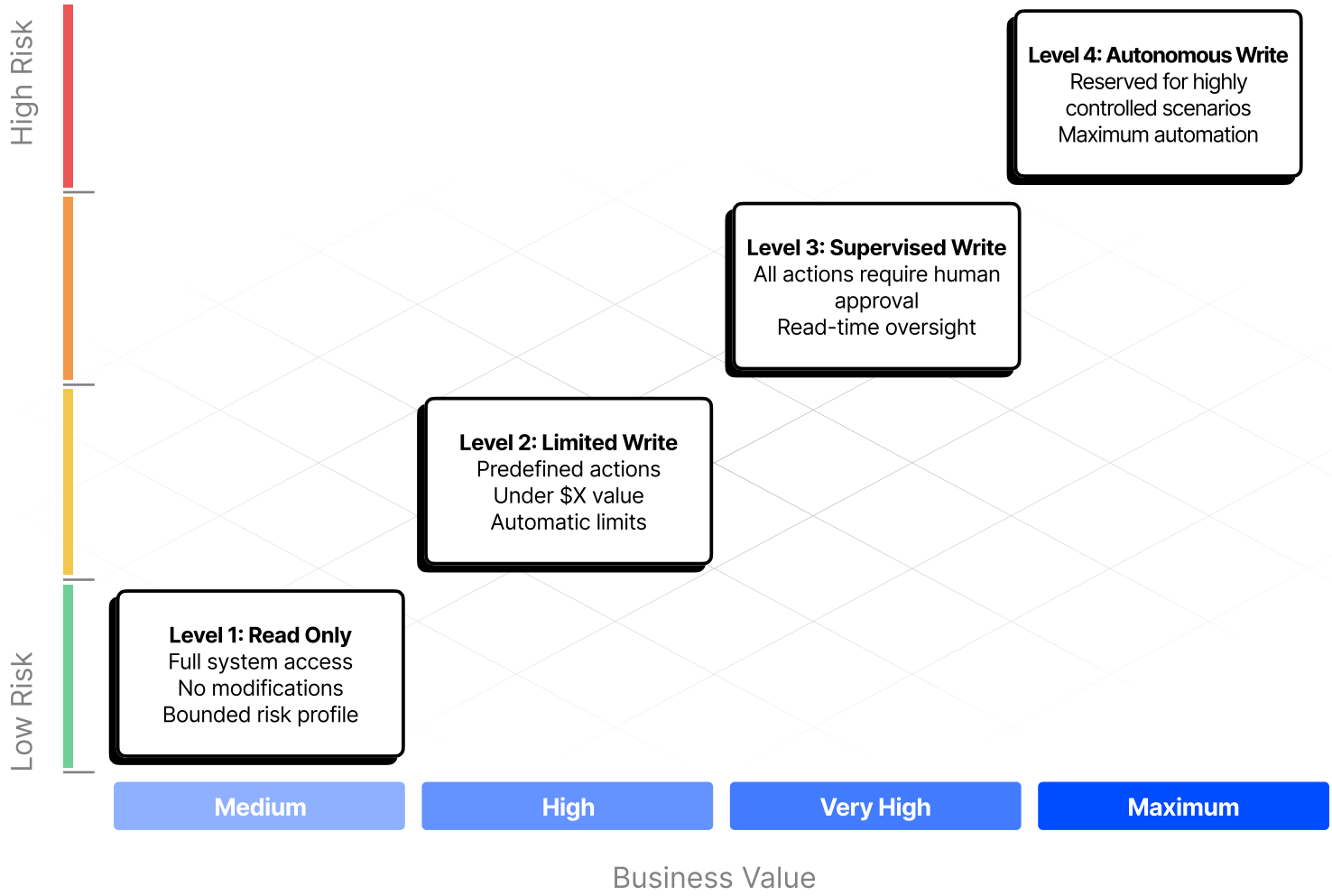
Context Loss in Approval Processes: By the time a human reviews an agent's proposed action, they may lack the full conversational context that led to the decision, making meaningful oversight difficult.

Speed vs. Safety Trade-offs: Business pressure for fast automated responses conflicts with the time needed for proper human review of agent reasoning.

Effective Governance Strategies for Write-Enabled Agents

Enterprises deploying agents need the appropriate governance frameworks put in place:

1. Tiered Permission Architecture



The Paradox: Highest business value = Highest security risk
Enterprise success requires balancing automation benefits with governance controls

Agent Permission Levels



Level 1 (Read-Only): Full system access, no modifications



Level 2 (Limited Write): Predefined actions under $X value



Level 3 (Supervised Write): All actions require human approval



Level 4 (Autonomous Write): Reserved for highly controlled scenarios

2. Business Impact-Based Controls

Rather than blanket write restrictions, implement controls based on potential business impact:

- **Medium Impact:** Automated responses, data logging, notification sending
- **High Impact:** Account modifications under preset limits, routine data updates
- **Very High Impact:** Financial transactions, account status changes, cross-system modifications
- **Maximum Impact:** Security-related changes, bulk operations, irreversible actions

3. Real-Time Reasoning Validation

Deploy systems that can analyze agent reasoning in real-time and flag potentially problematic logic:

- Unusual tool combination patterns
- Actions outside normal business workflows
- Requests involving high-value customers or sensitive data
- Operations that affect multiple systems simultaneously

The Write Permission Paradox

Enterprises face a fundamental paradox: the agents that provide the most business value are often those with write permissions, but these are also the agents that pose the greatest risk. Organizations that stick to read-only agents may achieve safety but miss the transformational potential of agentic AI. Those that deploy write-enabled agents without proper governance face potentially catastrophic consequences.

Resolution requires acknowledging that write-enabled agents are not just "more powerful" versions of read-only agents - they represent an entirely different class of technology with fundamentally different risk profiles and governance requirements.

The enterprises that succeed in deploying agentic AI safely will be those that design their systems around this critical distinction, implementing sophisticated controls for write operations while preserving the autonomous capabilities that make agents valuable in the first place.

Enterprise Agentic AI Deployment Patterns

Understanding how agentic AI is actually being deployed across enterprises helps clarify where NHI security risks emerge and why different approaches require different governance strategies.

Based on observations across enterprise environments, agentic AI deployments fall into three distinct tiers, each representing different levels of organizational control and security governance. These tiers reflect the actual evolution path most enterprises follow as they adopt agentic AI, starting with ungoverned experimentation and progressing toward managed platforms. Each tier presents fundamentally different NHI security challenges that require tailored governance approaches.

The three-tier model ranges from completely ungoverned shadow deployments (highest risk) to enterprise-managed cloud platforms (lower risk), but importantly, even the most "governed" tier presents novel NHI challenges that traditional security controls cannot adequately address. Understanding this spectrum is crucial for enterprise security leaders because different tiers require different detection, governance, and risk mitigation strategies.

Tier 1: Shadow Agentic AI Deployments

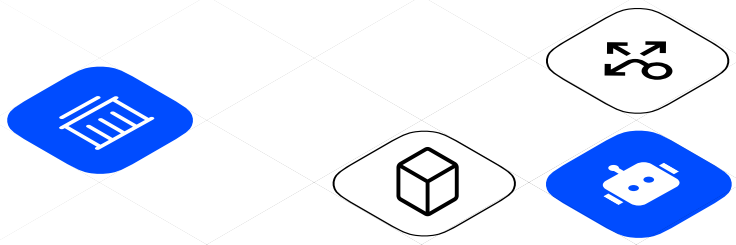
What "Shadow" Means: Shadow agentic AI refers to AI agent deployments that occur outside official IT governance and approval processes. These systems are typically implemented by individual developers or teams without security review, using personal credentials or over-privileged access that wasn't intended for AI automation.

Shadow agentic AI represents the highest-risk deployment pattern because it combines the dynamic NHI usage of true agents with complete absence of governance controls.

Characteristics:

- Deployed by individual developers or teams without security review
- Uses personal or over-privileged credentials for system access
- No audit trail or monitoring of credential usage
- Often bypasses traditional security controls entirely

The Model Context Protocol (MCP) Problem: MCP enables developers to expose enterprise systems to AI reasoning without any identity governance. A developer can create an MCP server that gives Claude Desktop access to production databases, Slack channels, and payment systems using their personal credentials - all running locally with no visibility to security teams.



Typical NHI Risk Profile:

- Production database admin credentials repurposed for agent access
- Personal Slack tokens with organization-wide permissions
- API keys hardcoded in configuration files
- Service account keys downloaded to developer machines
- Cross-system access with no boundary controls

Why This is Uniquely Dangerous: Unlike traditional shadow IT where usage patterns are predictable, shadow agents can autonomously discover and exploit credential combinations in ways that weren't anticipated, creating unlimited attack surface expansion.

Real-World Impact: A single shadow agent with database admin access and Slack permissions could autonomously query sensitive customer data and post findings to public channels - a perfectly logical sequence for AI reasoning but catastrophic for data protection.

Tier 2: Platform-Integrated Agents

What "Platform-Integrated" Means: These are agentic AI systems built on existing enterprise platforms (like Retool, Anthropic, or OpenAI) that provide some level of credential management and governance, but weren't originally designed for the dynamic NHI usage patterns of true agentic AI.

Retool Agents: The Gold Standard: Retool represents the most mature approach to enterprise agentic AI, with built-in credential management and governance controls. However, even well-designed platforms face NHI challenges when agents use credentials dynamically.

NHI Management Approach:

- Platform-managed credential storage with role-based access
- Built-in audit logging for agent actions
- Permission boundaries enforced at the integration level
- Automated credential rotation and lifecycle management

Remaining NHI Risks:

- Agents may still be granted broader permissions than necessary for flexibility
- Cross-integration credential correlation not fully governed
- Platform-level compromise could expose all agent credentials
- Limited real-time anomaly detection for unusual credential combinations

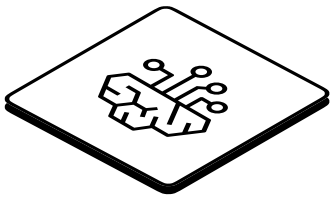
Platform Agent Example: Customer service agent that can access customer database, create support tickets, and send Slack notifications. While each integration is properly credentialed, the agent's ability to correlate data across systems and make autonomous decisions about which tools to use creates new risk scenarios not covered by traditional access controls.

Tier 3: Cloud-Native Agent Platforms

What "Cloud-Native" Means: These are managed agentic AI services offered directly by major cloud providers (AWS Bedrock Agents, Google Vertex AI Agents) that integrate with the provider's existing Identity and Access Management systems, offering the most comprehensive NHI governance available today.

AWS Bedrock Agents:

- Full integration with AWS IAM for credential management
- CloudTrail logging for complete audit trails
- VPC isolation and KMS encryption for data protection
- Guardrails for content filtering and safety controls



Google Vertex AI Agents:

- Integration with Google Cloud IAM and Workload Identity
- BigQuery and Cloud Storage integration for knowledge bases
- Service account management with automatic token rotation
- Cloud Audit Logs for comprehensive monitoring



Why "Governed" is Relative: While these platforms provide the best available NHI governance, they still face fundamental challenges with agentic AI's dynamic credential usage. An agent with legitimate access to customer data and billing APIs might autonomously decide to correlate sensitive information in ways that weren't anticipated during permission provisioning.

Enterprise Reality: Even with cloud-native platforms, organizations can still implement poor NHI security by over-provisioning permissions, sharing credentials across agents, or failing to monitor usage patterns effectively. Additionally, while the cloud provider manages the platform's identity infrastructure, all third-party service NHIs (Stripe, Slack, Salesforce, etc.) that agents use remain completely unmanaged and difficult to track or govern - creating significant blind spots even in "governed" deployments.



clutch

Enterprise Capabilities Required for Agent NHI Governance

Traditional IAM systems are fundamentally inadequate for managing the dynamic, context-aware credential usage patterns of agentic AI. Enterprises need new capabilities specifically designed for agent NHI governance.

Core Infrastructure Requirements

Dynamic Identity Lifecycle Management: Rather than static credential provisioning, enterprises need systems that can provision credentials just-in-time based on agent context, automatically rotate them without disrupting agent operations, and revoke access immediately when sessions end or anomalies are detected.

Cross-System Identity Federation: Agents often need to access multiple systems within a single workflow. This requires universal identity translation capabilities that can map agent identities across different authentication systems while maintaining unified audit trails and governance policies.

Real-Time Risk Assessment: Traditional periodic access reviews are insufficient for agents that may use credentials hundreds of times per day in novel combinations. Enterprises need behavioral analytics that can detect unusual patterns and automatically adjust permissions or trigger human review.

Advanced Governance Requirements

Agent Identity Provenance: Every agent deployment must maintain complete traceability of which human user or process initiated the agent, what permissions were granted and why, and how those permissions have evolved over time. This becomes critical for compliance and incident response.

Multi-Dimensional Compliance Automation: Agents operating across different data types and regulatory frameworks require automated compliance validation. For example, an agent processing both healthcare and financial data must automatically enforce HIPAA and PCI DSS requirements simultaneously.

Context-Aware Authorization: Traditional role-based access control is insufficient for agents that need different permissions based on conversational context. For example, a customer service agent might need billing access only when explicitly asked about account charges, not for general inquiries.

What Enterprises Are Missing

Most organizations lack the fundamental infrastructure required for agent NHI governance:

- No inventory of agent deployments or their credential usage
- No monitoring for agent-specific access patterns
- No policies for agent credential lifecycle management
- No incident response procedures for compromised agent identities
- No compliance frameworks for agent data processing

This infrastructure gap explains why many enterprises are unprepared for the security implications of widespread agentic AI adoption.

clutch

Core Principles for Agent NHI Security

Rather than complex technical implementations, enterprises should focus on establishing fundamental principles that can guide their approach to agent NHI governance:

Principle 1: Identity-First Agent Design

Every agent action must be traceable to a specific Non-Human Identity with clearly defined permissions. Agents should never operate with shared or overly broad credentials.

Principle 2: Agent Identification, Inventorying, and Accountability

Every AI agent must be uniquely identifiable as an active entity within the enterprise environment. While agents are not identities themselves, they must be trackable, auditable, and accountable for their credential usage, with clear lineage connecting agent actions to responsible human owners and business justifications.

Principle 3: Minimal Credential Scope

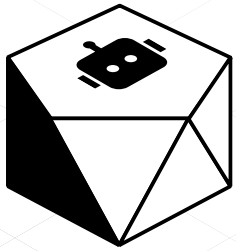
Agents should receive only the minimum credentials necessary for their intended function (Principle of Least Privilege), with automatic escalation processes for unusual requests rather than over-provisioning for flexibility.

Principle 4: Continuous Validation

Agent credential usage should be continuously monitored and validated against expected patterns, with automatic intervention for anomalous behavior.

Principle 5: Human Accountability

Every agent deployment must have a responsible human owner who is accountable for the agent's actions and credential usage.



clutch

The Path Forward

The enterprise adoption of agentic AI represents a fundamental shift that requires new approaches to Non-Human Identity security and management. Organizations that recognize this challenge early and implement proper governance will gain sustainable competitive advantages through safe, scaled deployment of agentic AI.

Those that treat agent security as an afterthought - or worse, mistake traditional automation for agentic AI - will face increasingly severe consequences as the technology proliferates throughout their operations.

The choice is binary: implement comprehensive NHI governance for agentic AI proactively, or face the inevitable crisis when ungoverned agent proliferation overwhelms existing security controls.

The window for proactive implementation is narrow. The exponential growth curves of agent adoption ensure that reactive approaches will be overwhelmed by the scale and complexity of ungoverned NHI proliferation.

The question is not whether agentic AI will transform enterprise operations - it already is.

The question is whether organizations will develop the capabilities required to govern the Non-Human Identities that make such transformation possible.



About Clutch Security

Clutch Security specializes in securing Non-Human Identities across enterprise environments. As agentic AI fundamentally changes how NHIs operate within organizations, Clutch provides the visibility, governance, and risk management capabilities required for safe AI adoption at scale.

clutch