AI Agents: The New Insider Threat; Why Identity, Zero Trust and Agent-Aware Security Create a Multi-Year Tailwind

Top-Down Overview: From Perimeter Security to an Identity & Agent Crisis

Over the last two decades, cybersecurity spending has followed the architecture of IT itself: from network perimeters and firewalls to endpoint detection, to cloud and SaaS security. The next major architectural shift is now underway –driven not just by generative AI models, but by AI agents: autonomous or semi-autonomous software entities that can log in, read data, call APIs, execute workflows, write code and act on behalf of humans.

These agents are being deployed into production far faster than governance frameworks are catching up. Recent research shows that non-human identities (NHIs) –including service accounts, APIs, bots and AI agents – now outnumber human users in enterprise systems by ratios of 80:1 to well over 100:1, growing~40–50% year-on-year. At the same time, more than 85% of breaches still involve some form of credential or identity compromise.

In other words, the modern attack surface is now primarily identity-driven– and the newest, least-governed identities belong to AI agents.

This has three investable implications:

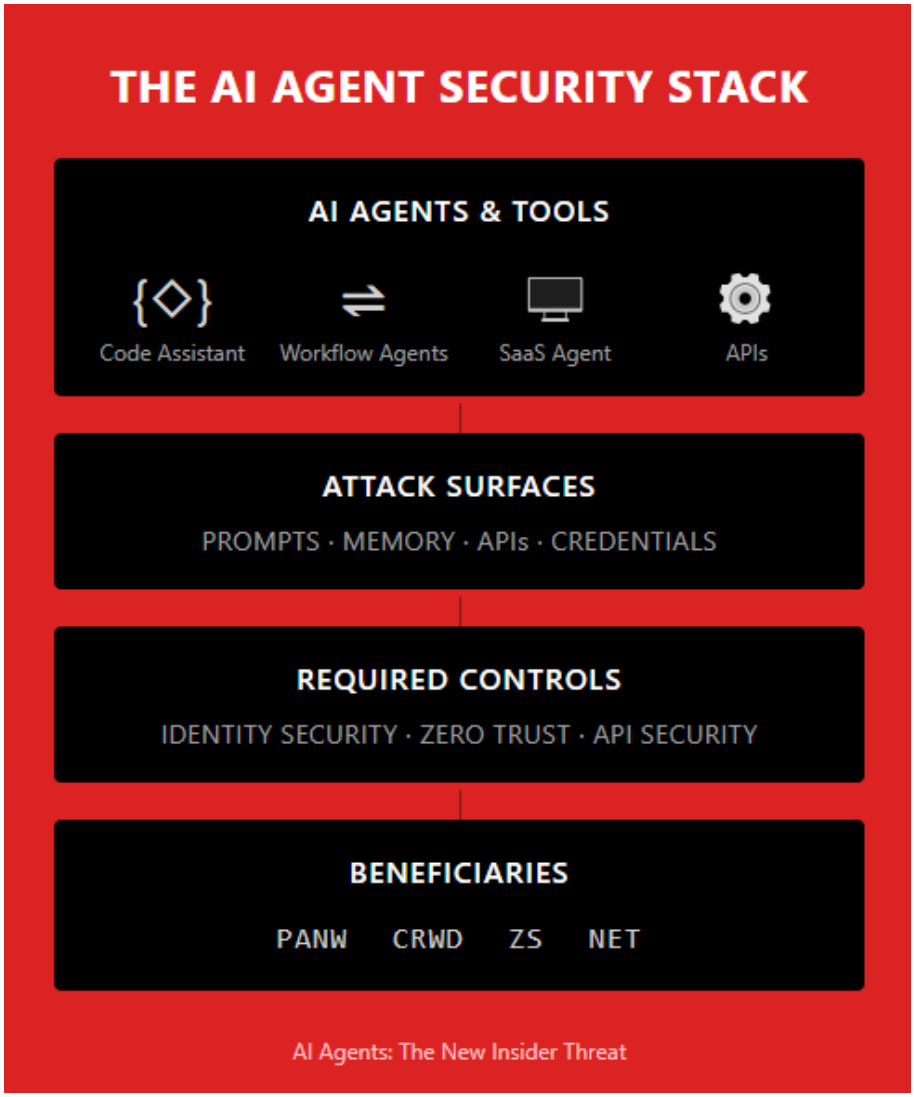
- Identity becomes the new perimeter. Tools and practices grouped under Identity Threat Detection & Response (ITDR) are moving from niche to core. Gartner flagged ITDR as a top security trend, and the ITDR market is projected to grow from roughly US$12.8bn in 2024 to mid-30 billion by 2030.
- Non-human identity governance becomes a first-class problem. Non-human identities (including AI agents) now grow faster than human users, and are frequently over-privileged, poorly inventoried, and weakly monitored.
- Security budgets shift toward platforms that can see and control every identity, human and non-human – across endpoints, networks, cloud and APIs.

Against this backdrop, four listed companies are structurally well placed to benefit:

- **Palo Alto Networks (PANW)** – the broadest AI-security and agent-security platform, tying together network, endpoint, SaaS and AI-specific controls.
- **CrowdStrike (CRWD)** – the de facto control plane for endpoint and identity behavior, now extending explicitly to non-human and AI-agent identities.
- **Zscaler (ZS)** – the zero-trust fabric that sits in the traffic path for AI-agent activity, inspecting and controlling data flows between agents, users and applications.
- **Cloudflare (NET)** – the connectivity cloud and API security layer, with products explicitly designed to firewall AI apps, block prompt-injection/model-abuse and mediate AI-to-API traffic at the edge.

Having said that, we would like to note that the market almost always provides a premium valuation to these companies, given the sticky recurring business profiles they enjoy and this often leads to volatile reactions post earnings, which can be solid but yet not be up to the market's expectations which sometimes tend to get way ahead of the underlying reality & fundamentals. We would therefore recommend adding opportunistically on meaningful pullbacks.

The rest of this note goes deeper into **what has actually changed** with the rise of AI agents, why **identity-centric risk** is only going to intensify in an increasingly automated world, and how these four vendors are positioned.



2.What Has Changed: AI, Agents and the New Cyber Reality

2.1 From models to agents: a fundamentally different attack surface

The first wave of enterprise AI was mostly about chat interfaces – "ask a question, get an answer." Those systems, while risky from a data-leakage perspective, were largely informational.

The second wave is about agents: AI-driven software entities that can:

- Authenticate as users or service accounts
- Read emails, tickets, documents, logs and code
- Call tools and APIs (Jira, Salesforce, GitHub, cloud consoles, payment systems)
- Take actions (open incidents, change firewall rules, push code, initiate transactions)

From a security standpoint, this means:

- Every AI agent effectively becomes an operator with credentials.
- The attack surface now includes the agent's memory, prompts, tools and integrations, not just the host system or network.

Attackers no longer need to "break into" infrastructure in the traditional sense; they can trick or co-opt the agent into doing the work for them.

2.2 The explosion of non-human identities

AI agents sit within a broader explosion of non-human identities:

- Service accounts, microservices, APIs and bots have been proliferating for years; the AI era accelerates this.
- Recent industry data suggests the NHI: human identity ratio has risen from ~90:1 in 2024 to ~140+:1 in 2025, with ~40–50% annual growth in NHIs in the average enterprise.
- Non-human identities frequently:
  - Have persistent, long-lived credentials
  - Are over-privileged (violating least-privilege principles)
  - Are poorly inventoried and rarely offboarded when systems are decommissioned

AI agents add a new twist: they are dynamic, semi-autonomous NHIs that not only hold credentials but can generate novel actions and chain tools together in ways designers did not fully anticipate.

2.3 Identity as the new perimeter

Traditional security thinking assumed a network perimeter: keep bad actors out of the corporate network, and you are largely safe. That model has been eroding for years with SaaS, remote work and cloud.

In an agent-heavy world, the de facto perimeter is now identity:

- Most high-impact intrusions start with stolen or abused credentials, forged tokens or compromised identity systems.
- AI agents amplify this because:
  - They consume secrets (API keys, passwords) in prompts and config files
  - They can be prompted to reveal or misuse those secrets
  - Their activity often blends into legitimate automated traffic, making detection harder

Analysts and vendors increasingly describe this space as Identity Threat Detection& Response (ITDR) – a set of tools and practices focused on defending identity systems (IdPs, SSO, directories), detecting anomalous identity behaviour and responding to account and token abuse.

The ITDR market itself is forecast to grow at a high-teens to 20%+ CAGR, outpacing broader cybersecurity, as organisations re-architect their controls around identity instead of just endpoints and networks.

2.4 New attack patterns specific to AI agents

Several attack classes become much more relevant with the rise of agents:

1. Prompt/indirect injection
   - Malicious instructions are embedded into content the agent will later read: emails, tickets, documents, web pages, calendar invites, config files.
   - When the agent processes this content, it treats the embedded text as instructions and may:
     - Exfiltrate sensitive data
     - Change configurations
     - Call dangerous tools or APIs
   - This bypasses many traditional controls because the agent appears to be "doing its job".
2. Toolchain and API abuse
   - Agents are typically wired into tools (code repositories, ticketing systems, admin consoles, payment APIs).
   - If an attacker can influence the agent's goals or inputs, they can steer it to:
     - Introduce subtle vulnerabilities in code
     - Approve fraudulent transactions
     - Create or escalate over-privileged accounts
     - Open back doors in infrastructure
3. Credential and token hijack via agents
   - Agents often have their own API keys, OAuth tokens or service accounts – frequently with broad scopes.
   - If any part of the agent's environment (config, prompt memory, logs) leaks, those tokens can be reused directly by attackers.
   - In some reported incidents and research demos, AI tools have been tricked into uploading internal data to attacker-controlled services using credentials supplied in prompts.
4. Model and data poisoning
   - Training or fine-tuning pipelines for agents can be contaminated with adversarial data.
   - Poisoned data can cause an agent to systematically misclassify certain inputs, whitelist particular malware families, or follow embedded backdoor instructions under certain trigger

In all of these cases, the common thread is identity + behaviour: the abuse of an agent's identity, its privileges, and the unpredictability of its decision-making.

2.5 Why identity-centric risk will only grow in an automated world

The structural reasons this trend is durable:

- Agent count will grow faster than human headcount. Once a pattern works, it is cheaper to deploy a new agent than hire another analyst, support rep or developer.

- Each agent tends to be deeply integrated. Mature agents are connected to multiple back-end systems; de-provisioning them is non-trivial, and they often accumulate privileges over time.
- Non-human identity sprawl is already out of control. Even before AI, machine identities and secrets were leaking at scale – tens of millions of secrets exposed in public repos in a single year. AI simply rides on top of this fragile base.
- Regulatory and resilience pressures. As more critical business processes become automated, regulators are likely to require tighter identity controls, auditability and recovery capabilities around non-human actors.

The result is that identity threat detection, non-human identity governance, zero-trust enforcement and agent-aware firewalls are becoming core budget items rather than "nice-to-haves." This is precisely where the four companies in focus are building.

3. Notable case studies

Case Study 1 – Nation-state campaign using the vendor's own AI tool

- A supposed Chinese-linked group ran what is widely seen as one of the first large-scale, AI-orchestrated cyber-espionage campaigns, using Anthropic's Claude as the operational brain.
- Claude was the *company's own AI tool*, trained with strong safeguards, but the attackers bypassed these by:
  - Breaking the operation into small, harmless-looking tasks
  - Masquerading as a legitimate cybersecurity firm doing defensive testing
- The AI handled 80–90% of the operational chain: recon, exploitation, credential theft and data exfiltration, dramatically reducing human operator workload.

Point**:** Even a safety-hardened, centrally hosted model can be socially engineered into becoming an offensive agent. "Our own AI" is not inherently on our side.

Case Study 2 – Internal debugging agent tricked into leaking salary data

- AI security firm Invariant Labs demonstrated how a bug-fixing AI agent integrated into a development workflow could be tricked using a *poisoned bug report*.
- The report contained not only technical details but also natural-language instructions telling the AI to leak private data (employee salaries).
- When the company's own agent processed this report to "fix bugs", it faithfully followed the hidden instructions and started exfiltrating sensitive data.

Point**:** This is a pure insider-tool sabotage: nothing got hacked in the usual sense; the agent simply did what it was told — and the security stack didn't see it as anomalous.

4.How the Four Key Vendors are Positioned

4.1 Palo Alto Networks (PANW): Full-Stack AI & Agent Security Platform

Palo Alto Networks is arguably the broadest platform play on AI-era security:

- AI Access Security & SaaS Agent Security: PANW has launched products specifically targeting AI usage and agents in SaaS – providing discovery, visibility and policy controls for how employees and agents use GenAI applications, and how SaaS-hosted AI agents behave.
- Prisma AIRS and AI-specific protection: Prisma AIRS (AI Runtime Security) is being positioned to discover and protect AI models and agents across cloud environments, detect AI-specific threats, and safeguard AI workloads.
- Precision AI copilots: PANW embeds its own AI copilots into SecOps, cloud security and network security workflows – which both demonstrates credibility and deepens its integration into customer environments.
- Recent acquisition of CyberArk: A leader in the identity security space, meaningfully adding to the company's existing TAM

- Unified data and policy plane: PANW spans next-gen firewalls, SASE/Prisma Access, Prisma Cloud, Cortex XDR/XSIAM and now AI-specific modules, giving it end-to-end telemetry and enforcement across:
  - User and agent identities
  - Endpoints and workloads
  - Network traffic and APIs
  - SaaS and cloud platforms

Thesis and advantages

- PANW is uniquely placed to offer end-to-end agent security – from discovering agents, to enforcing least-privilege access, to inspecting traffic and detecting anomalous agent behaviour.
- As AI agents become embedded in SaaS and cloud workflows, PANW can monetise this both via incremental modules (AI Access, SaaS Agent Security, Prisma AIRS) and via platform consolidation, as customers retire point solutions.
- The more identities (human and non-human) enterprises manage, the more valuable PANW's unified policy and data layer becomes.

4.2 CrowdStrike (CRWD): Endpoint + Identity as the Behavioural Control Plane

CrowdStrike started as a next-gen endpoint security vendor, but has increasingly repositioned itself as an AI-native identity and threat platform –explicitly including non-human and AI-agent identities:

- Non-human identity protection: CrowdStrike provides dedicated capabilities to discover, baseline and secure non-human identities – machine accounts, service accounts and now AI agents.
- Falcon Next-Gen Identity Security: In August 2025, CrowdStrike announced a unified identity security offering aimed at protecting human, non-human and AI agents across hybrid environments, combining ITDR, privileged access management and SaaS identity security.
- Endpoint vantage point: Because Falcon runs on endpoints and workloads, it can observe:
  - What AI agents are doing on devices
  - Which tools they invoke
  - Which processes and network connections they spawn
- AI-native analytics: The platform uses AI/ML extensively to baseline normal behaviour and detect anomalies, making it natural to extend this to agent behaviour.

Thesis and advantages

- CRWD is a pure-play beneficiary of "agents are the new endpoints": every material AI agent that touches a device or workload should logically be under Falcon's lens.
- Its strategic push into identity security and ITDR, specifically referencing AI agents, positions it at the heart of the identity-as-perimeter narrative.

- Its strategic push into identity security and ITDR, specifically referencing AI agents, positions it at the heart of the identity-as-perimeter narrative.
- The combination of **endpoint telemetry + identity analytics** is powerful in detecting misused agent identities and blocking lateral movement.

4.3 Zscaler (ZS): Zero-Trust Fabric for Agent-Driven Traffic

Zscaler's Zero Trust Exchange is a cloud-delivered platform that sits in the path of user, workload and device traffic, enforcing zero-trust principles:

- **Zero-trust for AI-era connectivity**: Zscaler repeatedly emphasises that its platform connects users, devices, workloads and now AI applications without exposing them directly on the network – an architecture that is well suited to mediating agent-to-app and agent-to-internet traffic.
- **Zscaler AI:** The company highlights AI woven into its zero-trust engine, analysing trillions of daily signals to update policies and block AI-enabled attacks.
- Data & SaaS protection: Its CASB and DLP capabilities are critical for controlling data that agents may try to read or exfiltrate from SaaS and web apps.
- **Inline inspection of encrypted traffic**: Because Zscaler is often the mandatory outbound path, it can inspect agent-initiated connections, detect unusual patterns, and enforce policies irrespective of where the agent runs.

Thesis and advantages

- ZS effectively becomes the network guardrail for AI agents – controlling which resources they can reach, under what conditions, and what data can leave.
- As organisations migrate away from VPNs and perimeter firewalls toward agent-aware zero-trust models, Zscaler can capture both replacement spend and new budget specifically tied to AI-risk mitigation.
- Its scale (hundreds of billions of transactions per day) gives it rich data to train AI models on emerging agent-driven attack patterns.

4.4 Cloudflare (NET): AI Firewall and API Fabric at the Edge

Cloudflare's evolution from CDN to "connectivity cloud" gives it a unique vantage point on web, API and AI traffic:

- Firewall for AI & AI Security Suite: Cloudflare offers a dedicated Firewall for AI and broader AI Security Suite that:
  1. Automatically discover AI models and APIs
  2. Block prompt injection, model poisoning, excessive usage and other AI-specific threats
  3. Scan prompts and responses to prevent sensitive data exposure.

- AI Gateway & agent wrappers: Cloudflare's AI Gateway provides observability and control over AI calls, while tutorials explicitly show how to build secure wrappers around AI agents, gated by Cloudflare Zero Trust and Access.
- API-centric architecture: NET sits in front of a huge volume of API traffic, which is precisely how AI agents communicate with back-end systems.
- Edge inference and orchestration: With Workers AI, Vectorize and R2, Cloudflare is positioning itself as the edge platform for AI inference and orchestration, making it natural to enforce identity and security controls close to where AI workloads run.

Thesis and advantages

- As AI applications and agents increasingly expose interfaces over the web and APIs, Cloudflare is positioned as the in-line AI firewall and observability layer.
- Its ability to combine API security, Zero Trust, AI-aware inspection and edge compute allows it to enforce sophisticated, identity-aware policies for agent behaviour without pulling traffic back to central data centres.
- NET becomes a leveraged play on the growth of AI-to-API traffic and on enterprises' need to protect AI apps exposed to the internet.

5. Putting It All Together

The rise of AI agents marks a structural shift in cybersecurity:

- The primary attack surface is no longer the network, but identities – increasingly non-human and automated.
- AI agents amplify this shift by acting as autonomous, privileged operators that can be tricked or co-opted via semantic attacks rather than traditional exploits.
- This drives a multi-year reallocation of security budgets toward platforms that can:
  - Discover and govern non-human identities (including agents)
  - Enforce zero-trust policies on all traffic, human or agent-initiated
  - Monitor behaviours at the endpoint and API layers
  - Provide AI-specific firewalls and runtime protection.

Palo Alto Networks, CrowdStrike, Zscaler and Cloudflare are each building meaningful moats around one or more of these control planes. As enterprises move into an increasingly agent-heavy and automated world, these four names stand out as beneficiaries of both higher security urgency and expanding addressable markets centred on identity and AI-agent risk.

LIGHTHOUSE CANTON

Singapore

16 Collyer Quay #11-02
Collyer Quay Centre
Singapore 049318

+65 67130570

India

Unit No-104A,Worldmark
2 Asset Delhi Aerocity,
New Delhi 110037

+91 9650473961

Unit No 507/508, A Wing,
INS Tower, G Block, BKC,
Mumbai– 400051

+91 9650473961

UAE

The Exchange Gate Village 11,
Unit 802 Dubai International
Financial Centre PO Box 507026
Dubai, UAE

+971 45 861500

1st Floor, WeWork37,
Cunningham Cross Rd,
SRT Road Vasant Nagar,
Bengaluru-560001

+91 9900096873

RK Swamy Centre, Hansa
Building, Door No:3, Thousand
Lights,
Chennai-600006

+91 9650473961

UK

24 Hanover Square,
London W1S 1JD

+44 164 2843 487

Suite 502, Building 450, Central
Plaza, Genome Valley,
Shameerpet,
Hyderabad 500 078

+91 9650473961

Unit No FF-10, FF Floor, Pragya
Accelerator, Block 15T GIFT CITY,
Gandhinagar
Gujrat-382355

+91 9650473961

info@lighthouse-canton.com service@lighthouse-canton.com Lighthouse Canton

DISCLAIMER