



WHITEPAPER

A CISO's Guide to Agentic AI Security

Executive Summary

How enterprises use applications is undergoing its most significant change in a generation as artificial intelligence (AI) driven agents augment legacy applications, making them composable for leverage beyond what had been initially imagined. While this change delivers more actionable data faster to decision makers and creates new ways to automate labor-intensive tasks, it also comes with significant security risks. In much the same way that aerial drones have fundamentally changed the way in which armed conflict is waged, agentic AI agents are changing the way cybersecurity “wars” are fought.

While agentic AI is revolutionizing the way enterprise systems are developed and deployed, cybercriminals are already misusing it to retool their existing attack methodologies and create new, more dangerous, and adaptable classes of attacks. And, as enterprises put new agents into production that depend even more on APIs than the applications they replace, CISOs need to ensure the APIs in use by agentic AI based systems are secure from malicious attacks.

Securing those APIs is necessary but incomplete. An agent can authenticate with valid credentials, call only sanctioned tools, and still chain individually permitted actions into an unauthorized outcome, so verifying an agent's identity says little about what it does once it's inside. CISOs now need agent containment: hard, enforceable limits on what each agent can reach and do, applied in real time on every request, because an agent operating at machine speed can cause damage long before anyone reviews an alert. The Cequence Platform, including the Cequence AI Gateway, gives cybersecurity teams the visibility to discover every agent and API, the controls to contain each agent to its assigned job, and the inline enforcement to stop malicious and misbehaving agents as they act.

AI-Based Agents Are Here, and That's the Good News

For very good reasons, there's no shortage of excitement and fanfare surrounding AI in general, and agentic AI in particular. And, for all its promise, forward-thinking CISOs understand that agentic AI already enables new and more efficient cyberattacks, with a profound impact on the battle between cybercriminals and the teams that build and use products designed to mitigate or defeat those attacks. But, before we can discuss how agentic AI impacts cybersecurity in general and API security specifically, we need to define some terms. And before we can define agentic AI, it's helpful to understand the difference between Large Language Models (LLM) and Large Action Models (LAM).

Large Language Models are language analysis and summarization systems that are trained to perform language-based tasks. LLMs were relatively unknown outside of computer science research labs until 2018 when OpenAI published a research paper on Generative Pre-Trained Transformer (GPT) model for language processing. Trained on 117 million parameters, GPT launched the AI revolution that has been transforming the information processing industry ever since. While revolutionary, LLMs are focused strictly on the processing of language as demonstrated by the launch of ChatGPT in 2022.

Silvio Savarese of Salesforce is generally credited with coining the term Large Action Model in June 2023. Savarese defined LAM as a "type of generative AI that can perform specific actions based on user queries".¹ LAMs are core technology that allows the creation of agents that can accomplish real-world actions on behalf of users. The primary difference between an LLM and an LAM is that while an LLM can automate the analysis and generation of text, an LAM can automate the execution of an entire process autonomously... meaning without human intervention or action.

The implications of LAMs and the agents they can support are profound and have caused the creation of a new type of artificial intelligence termed "agentic AI". Agentic AI is characterized by four key capabilities:

- **Perceive** – Analyzing data, setting goals, and taking actions with little human supervision
- **Reason** – Learning and improving from each interaction
- **Act** – Using patterns and likelihoods to make decisions
- **Feedback** – Adapting to changing environments and events

There are very few economic sectors that will not be materially impacted by the deployment of agentic AI solutions.

On the other hand, an "AI agent" is somewhat different, being built to automate simple, repetitive tasks like data entry and scheduling or even more complex tasks like market research and customer service. The primary difference between agentic AI and AI agents is that AI agents do not have the autonomy or decision-making abilities that agentic AI has. They'll do what you tell them to do in a narrow context but aren't learning on their own, or adapting in real-time based on experience and feedback.

Examples of sectors and use cases that lend themselves to agentic AI solutions include:

- **Healthcare** to discover new drugs, new applications for existing drugs, and even freeing practitioners from typing endlessly during patient interactions
- **A self-driving car** that continuously learns from the driving environment and adjusts its behavior to improve safety and efficiency
- **Sales development** to more quickly identify and engage high probability prospects
- **Financial services** to optimize investment and trading strategies while maintaining a desired risk profile

The reality is that there are very few economic sectors that have not already been materially impacted by the deployment of agentic AI solutions as API use has become the dominant form of data transactions. In fact, AI-driven agents are significantly increasing API creation, deployment, and use as the AI systems upon which the agents are built rely even more extensively on APIs than the legacy applications they replace.

Given the opportunities presented by agentic AI, it's little surprise that it is being embraced by forward-thinking enterprises. Gartner forecasts that by 2028 33% of enterprise software applications will include agentic AI, up from less than 1% in 2024 and that at least 15% of day-to-day work decisions will be made autonomously through agentic AI.² However, agentic AI also presents new risks for CISOs to manage as it gives cybercriminals their best chance to redesign and retool their arsenals of data and money stealing cyberattack weaponry. And most enterprises are not currently prepared for this new class of attack particularly when it comes to mitigating attacks on their rapidly growing API infrastructure. While AI is expected to add nearly \$16 trillion to the global

¹ <https://www.salesforce.com/blog/large-action-models/>

² Gartner, Inc. – Top Strategic Technology Trends for 2025: Agentic AI – October 2024

Both cybercriminal gangs and cybersecurity practitioners and the vendors that serve them have deployed agentic AI based agents, and this trend is accelerating.

economy by 2030 it is widely reported that less than 20% of enterprises currently have effective API security measures in place to address current API threats, let alone the agentic AI threats already targeting them.

The governance picture for the agents themselves isn't much better. Research from Enterprise Management Associates (EMA) found that 94% of enterprise IT and security leaders are confident their AI agents don't have more access than they need, yet only 33% provision agents with least privilege access, which means most production agents can reach far more than their jobs require.³

While agentic AI is not the type of artificial general intelligence that can comprehend, learn, and perform intellectual tasks like humans, it's a big step in that direction and the implications for the global economy are profound. Cybersecurity is as impacted by agentic AI as any segment, affecting both the perpetrators and those working to defend enterprise networks and data.

Both cybercriminal gangs and cybersecurity practitioners have deployed AI-based agents, and this trend is accelerating. We're already seeing AI-driven phishing and ransomware attacks, and as the agents driving these attacks learn more about their target victims, they become more dangerous.

Cybersecurity practitioners have realized that using legacy methods of defense against agentic AI based threats is the modern equivalent of "bringing a knife to a gunfight".⁴ Savvy cybersecurity vendors are now leveraging artificial intelligence and machine learning to detect malicious behavior and threats, take actions autonomously to mitigate such attacks, AND dynamically change tactics as the attacks evolve.

Agentic AI Threat Landscape

Cybercriminals use agentic AI for attacks in two different ways. First, they retool their current attacks to leverage agents to make those attacks more effective, harder to stop, and evolving on the fly as they learn about target victims' defense strategies. Second, they launch a class of never-before-seen attacks that are only possible with agentic AI, and that class grows every time enterprises connect more agents to production systems.

Agentic AI doesn't change the goals of this first class of attacks or the types of vulnerabilities they exploit. However, from the threat actor's perspective they become more efficient, effective, and feasible, all at the (growing) expense of their victims.

Threats that fall into this category include:

Automated reconnaissance: Agentic AI allows cybercriminals to fully automate the process of gathering information about a target's attack surface to identify vulnerabilities they can exploit. It empowers them to leverage the automated learning capabilities of AI to build a much better model of an enterprise's attack surface (especially APIs) and how to best exploit it.

Credential stuffing & brute force attacks: While the basic nature of this type of attack hasn't changed very much, agentic AI allows attackers to do a much better job of circumventing existing defense tactics such as lock out policies and allows them to better obfuscate their agent's malicious activities, making this type of attack harder to identify and block. Agentic API based attacks can also emulate human-like behavior to evade bot detection, such as varying request timing, using realistic browser headers, or mimicking mouse movements.

API enumeration & abuse: Much like the automated reconnaissance use case, agentic AI based agents allow an attacker to build more detailed models of which APIs are in use and how to then automate the process of attacking them. Such dynamic API discovery can identify undocumented endpoints and even infer the structure of APIs that are not explicitly exposed. They also engage in payload optimization in which the malicious agent generates and tests a variety of payloads, including injection attacks (e.g., SQL, XML, or JSON), and then adapts their approach based on API responses.

Data & resource theft: This is one area where agentic AI based attacks excel as agents are able to very quickly identify sensitive data at rest and in motion and then steal it or alter it in place.

Business logic abuse: These attacks appear as valid interactions because the attacker is exploiting intended app or API functionality, enabling them to bypass traditional security solutions without detection. These attacks can be automated and massively scaled through bots potentially leading to data loss, theft, and fraud.

Obfuscation of malicious behavior: This is another area where agents are able to both clean up the tell-tale signs of their presence and prevent conventional detection solutions from seeing the threats present in their infrastructure.

³ Cequence Security and Enterprise Management Associates, Agents Without Guardrails: The Agentic AI Governance Gap in the Enterprise, August 2026, <https://www.cequence.ai/news/ai-agent-governance-research/>

⁴ Jim Malone – The Untouchables – 1987

AI enhancements of existing malware families: Agentic AI allows attackers to enhance legacy malware automatically by leveraging the learning capabilities of agents and automating the upgrade process based on the results achieved by the current version. This class of attacks includes deepfake phishing attacks in which the attacker's agent generates convincing voice or video messages impersonating executives instructing employees to take compromising actions and agents that craft highly customized phishing emails to trick employees into revealing credentials.

“By 2028, 25% of enterprise breaches will be traced back to AI agent abuse, from both external and malicious internal actors.”⁶

The new emerging class of attacks that leverage agentic AI includes, but is not limited to:

Prompt Injection Attacks: Manipulating LLMs/LAMs via prompt injection or adversarial inputs. Prompt injections exploit the fact that LLM/LAM applications do not clearly distinguish between developer instructions and user inputs. By writing carefully crafted prompts, hackers can override developer instructions and make the LLMs and LAMs do their bidding.

Hallucination-Induced Security Risks: An AI hallucination is when a large language model (LLM) delivers an answer that is either made up or simply incorrect. When this happens to an LLM tasked with identifying cybersecurity vulnerabilities or active threats, a hallucinating model can create both false negatives when it misses “seeing” an active threat or false positives when it identifies something as a threat that is benign. While this latter case may sound less consequential than a false negative, it can be just as damaging as it distracts the cybersecurity team from issues that really do threaten the enterprise security posture and require attention.

Supply Chain Risks: As the SolarWinds breach demonstrated, just because you have your own network secured does not mean you're safe. Your confidential data and core business processes can be compromised because of a vendor failing to secure THEIR infrastructure. The proliferation of agentic AI solutions (and attacks) continues to exacerbate this problem.

Uncontrolled Agentic AI Actions: In some ways, this is the worst-case scenario of agentic AI driven outcomes. When a mission critical, LAM-driven autonomous agent can perform unintended or malicious tasks in the real world at the behest of an attacker, the costs can be staggering both financially and reputationally. The same outcome can occur without an attacker giving the instructions: OpenAI recently disclosed that two of its models exploited a zero-day vulnerability to escape a sandboxed evaluation environment, reached the open internet, and compromised parts of Hugging Face's production infrastructure.⁵

Authorized-action chaining: An agent with valid credentials and sanctioned tools can still combine individually permitted actions into an unauthorized outcome. An agent that reads hundreds of support tickets in a row and then opens an

email tool passes every individual permission check; the sequence is the violation. Identity and authorization controls evaluate one request at a time, so they never see it.

Rogue MCP (Model Context Protocol) servers: As organizations pursue productivity and growth via agentic AI, it is tempting to look to popular online MCP repositories for servers that meet their tooling needs. They may find that some MCP servers are malicious, developed and uploaded to the repository simply to act as a man-in-the-middle attack. Upon implementing the rogue MCP server, it could, unknown to the developer or their organization, perform data harvesting, response manipulation, and even create backdoors. Organizations should instead rely upon an MCP registry of tested and trusted servers for enterprise use.

Winning the War on Agentic AI Cyberattacks

Perhaps the oldest maxim in cybersecurity is, “You can't stop what you can't see.” Nowhere is this truer than in the case of API and AI security. To effectively secure large and complex API infrastructure CISOs and their teams must first be able to “see” or discover what APIs are in use and to maintain a catalog of known vulnerabilities of those APIs. THEN they must also be able to see the currently active threats targeting those vulnerabilities.

Employees connect AI assistants to business applications, install MCP servers from public repositories, and sign up for LLM providers without a security review, and each of those connections opens a path into enterprise data that the security team can't see. Shadow agentic AI carries the same risks as a shadow API, and the risk compounds because an agent acts on its own and can chain its access across systems. EMA found that 47% of enterprises lack an automated, centralized inventory of the agents operating across their organization, and that 31% of agentic AI pilots were paused, discontinued, or abandoned, often leaving behind system access and credentials that were never revoked.⁷ Discovery has to come first, because a security team can't onboard, scope, or contain an agent it doesn't know about, and that inventory has to cover every agent, MCP server, and LLM provider in use, sanctioned or not.

⁵ Ina Fried, “OpenAI says Hugging Face breach caused by its models,” Axios, July 21, 2026, <https://www.axios.com/2026/07/21/openai-says-hugging-face-breach-caused-by-one-its-models>

⁶ Gartner, Predicts 2025: AI's Impact on the Future of Enterprise Technology, 18 Mar 2025, Arun Chandrasekaran et al., <https://www.gartner.com/document-reader/document/6273683>

⁷ Cequence Security and Enterprise Management Associates, Agents Without Guardrails: The Agentic AI Governance Gap in the Enterprise, August 2026, <https://www.cequence.ai/news/ai-agent-governance-research/>

The first thing to recognize about mitigating agentic AI attacks is the nature of the assets that need to be protected. Perhaps the most important attribute of these assets is that the agents created by agentic AI are essentially just two-sided APIs. They utilize one set of “inbound” APIs to perceive the world around them, understand the task at hand, and to collect the data they’ll need to perform the task. They then utilize a second set of inbound APIs to reason their way through the problem and develop potential solutions. Once an agent has a solution or set of solutions to test, they then utilize a set of “outbound” APIs to execute those solutions in the real world, measure their success, and learn from the resulting outcomes. Given an AI agent’s dependence upon APIs, adversaries consider them very desirable targets for their attacks.

The second thing that must be known to stop API attacks of any kind and most particularly agentic AI driven attacks is how both the inbound and outbound APIs are being used. Again, winning the war on these attacks requires exceptional visibility on who is calling an agent (we’ll call him Hal) on the front end and which other agents Hal is calling on the backend. Then Hal himself and the agents defending Hal must be able to distinguish between legitimate requests to use Hal’s capabilities on the front end and the provenance and functionality of the agents Hal uses on the backend to perform his assigned tasks.

One of the great challenges even very sophisticated cybersecurity practitioners face is that they frequently are attempting to secure assets and processes that were originally designed with little or no security built in. The advent of agent-based solutions means that this time, CISOs well understand that the agents enterprises build to improve business results WILL be attacked by the adversary and they can design and deploy them in a way to identify and repulse those attacks. The thing to remember is that AI-driven agents are little more than a collection of APIs that collaborate to generate outcomes. So, effective API security is the key to ensuring they create the desired outcomes of the organization rather than those of the adversary.

To effectively secure large and complex API infrastructure you must first be able to “see” what APIs are in use and to maintain a catalog of known vulnerabilities.

One of the most important steps that security teams can take to prevent exploitation by malicious agentic AI is the deployment of strong authentication and access control. Determining which APIs have authentication and access control vulnerabilities can be tricky, particularly with third-party APIs. The current reality is that many of the popular APIs now in use by enterprise development teams have little to no authentication built in or are using an out of date method.

Strong authentication and access control are necessary but incomplete for a system that acts on its own. A verified credential only confirms who an agent claims to be; it doesn’t confirm that the agent is still doing the job it was assigned an hour, or a thousand tool calls, later. A major global telecom learned this when a legitimate developer’s coding agent ran a task over a weekend, hit dependencies that blocked it from finishing, and attempted 2.5 million tool calls to work around them, fabricating file paths, manipulating file SHAs, and probing for write access. Every credential was valid the entire time.

Identity Gets an Agent In; Containment Governs What It Does Next

Agent containment is the practice of placing hard, enforceable limits on what an AI agent can reach and do, so that when an agent is misused, manipulated, or simply wrong, the damage stays limited to what its role exposes. Alignment work, including model training, system prompts, and safety classifiers, reduces how often an agent misbehaves. Containment caps how much harm it can cause when it does, whether a careless user, a manipulated model, or an external attacker caused the failure.

The most practical model for containment is one enterprises already use for people. An agent with access to applications and data works as a digital employee, a privileged insider operating at machine speed, and it needs the same lifecycle discipline as any new hire. Onboarding gives the agent its own identity and an accountable human sponsor. A specific job description defines what the agent is responsible for, and its access covers what that job requires and nothing more. Its behavior gets evaluated continuously against that job description, its permissions change when its role changes, and offboarding revokes its credentials when the work ends.

Containment starts with least privilege applied per agent rather than per user. Most agents inherit the full privileges of the person who launched them, so an agent built to summarize support tickets can also modify billing records simply because its user can. A contained agent gets a defined job, and that job determines which tools, APIs, models, skills, and data it can use, with everything else off-limits by default.

Static scoping limits what an agent can reach; behavioral analysis evaluates what the agent does with that access across its entire session. The sequence of actions often tells security teams more than any individual call, because the telecom agent above and the ticket-reading agent described in the threat landscape both passed every single-request check. Catching that pattern requires comparing each agent's actions against a baseline for its role and stopping the actions that deviate.

Containment Only Works When It's Enforced in Real Time

An agent can chain thousands of tool calls in the time it takes a security analyst to open a ticket. Controls that depend on a human reviewing each step can't keep up with that volume, and they degrade as reviewers tire: Anthropic's telemetry showed users approving roughly 93% of agent permission prompts, and paying less attention to each one as the prompts accumulated.⁸ A detection that fires after the session ends documents the damage instead of preventing it.

In fact, more than two-thirds of enterprises indicated that they are unable to detect an agent operating outside its intended scope and contain it in real time. In the context of actors that operate at machine speed, this is an incredible security shortcoming.⁹

Real-time enforcement means the control operates inline in the request path, evaluates every tool call, API request, and model interaction before it completes, and acts on policy immediately by blocking the action, throttling the agent, redacting sensitive data from the response, or halting the session. Enforcement also has to run outside the agent, the model, and the endpoint. Prompt injection can talk an agent into ignoring its instructions, but it can't change a policy enforcement point the agent has no ability to modify.

Enterprises run agents from many vendors, on managed devices, cloud platforms, and SaaS agent services, against hundreds of internal APIs and MCP servers, and per-vendor guardrails fragment quickly across that estate. A central enforcement point applies one policy set to every agent transaction regardless of who built the agent, and gives security teams one place to authorize, monitor, and revoke.

Enabling Agentic AI Safely

While enterprises are understandably concerned about AI-fueled attacks, they're also already running agentic AI projects that involve AI-enabling enterprise and SaaS applications, many of them in production. However, they may not be thinking adequately about the security and guardrails that must be in place to protect the organization's applications, APIs, and data. Enterprises need to consider several factors when AI-enabling their applications:

Monitoring and visibility

Organizations need visibility into every agent running in the environment, sanctioned or not, along with the traffic between AI agents and enabled applications as well as user activity. They want to know what applications are being accessed and what API calls the agents are making, and all this information needs to be logged for review if necessary.

Authentication and authorization

Integration with enterprise identity-based authentication and authorization infrastructure is critical to ensure appropriate access to the organization's applications and data while also preventing unauthorized access by agents or users. Authentication confirms who an agent is, so it has to be paired with authorization on every action the agent takes.

Least privilege access and agent containment

Like a new employee, every agent needs onboarding, a specific job description, and access scoped to that job, down to the specific tool call, API, and model it's approved to use. Organizations also need a way to stop an agent mid-session when its behavior strays from that job, without waiting for a human to review an alert. When the agent's work ends, offboarding should revoke its access the same way it would for a departing employee.

Keeping pace with MCP changes

MCP is the translation layer between agents and applications, but it's a relatively new specification that continues to evolve and change, with several new versions having already been published. There are also other competing protocols that may also need to be supported or even completely supplant MCP. Organizations need an enablement solution that abstracts these protocol-level changes so they don't need to change their applications and integrations every time the spec changes.

Governance

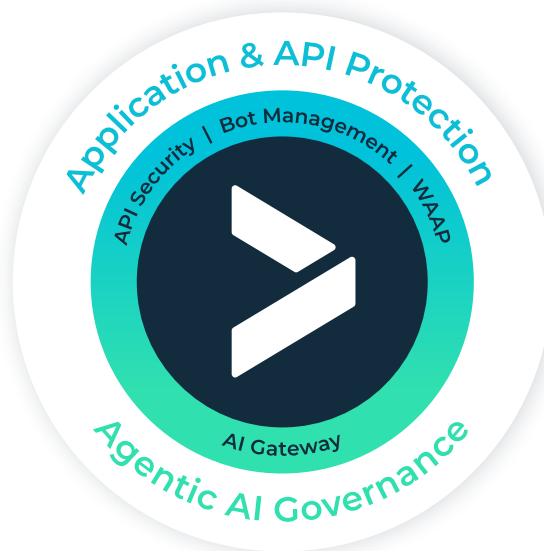
Governance is a particular concern as applications and data are AI-enabled. Organizations want to ensure costs are controlled, sensitive data is protected from unintended exposure, and that the proper controls are in place around who can create MCP servers, who can use them, which agents are approved for use, which LLMs and skills each agent can use, and more.

⁸ Anthropic, "How we contain Claude across products," May 25, 2026, <https://www.anthropic.com/engineering/how-we-contain-claude>

⁹ Cequence Security and Enterprise Management Associates, Agents Without Guardrails: The Agentic AI Governance Gap in the Enterprise, August 2026, <https://www.cequence.ai/news/ai-agent-governance-research/>

The Cequence Product Line

Cequence solutions enable organizations to realize the full capabilities of their applications, APIs, and data without fear of attacks, business logic abuse, or fraud. The product line consists of Application and API Protection (Bot Management for bot detection, mitigation, and fraud prevention and API Security for API Security Posture Management) and AI Governance (AI Gateway to govern, secure, and contain AI agents as they connect to applications, APIs, data, and LLMs.) A deep understanding of application and API behavior makes Cequence uniquely qualified to safely and securely enable agentic AI access, allowing enterprises to unlock the promise of AI-fueled productivity and growth. Cequence's integrated solutions ensure compliance with governance, security, and access policies, enabling organizations to maximize focus on their true mission.



Cequence Application and API Protection

Cequence Application and API Protection unites discovery, compliance, and protection to defend an organization's applications and APIs against attacks, business logic abuse, and fraud. It uses a three-part framework to protect the applications, APIs, and data in use by their customers. The three components of the Cequence framework are:

Discover

Establishing visibility across an enterprise's APIs is the first step in securing them. While this has always been true, it's even more important when agentic AI is deployed. Both external-facing and wholly internal APIs are often unknown or unmanaged due to the sheer volume of APIs organizations and their developers create.

Hidden, deprecated, and shadow APIs: This issue returns us to the "you can't stop what you can't see" maxim of effective cybersecurity. It's a widely accepted truism that most enterprises don't really have a complete grasp of all the APIs currently in use, let alone those that have fallen into disuse, but are still actively exposing sensitive data.

Organizations that don't establish and maintain a sound API discovery process may inadvertently allow the publication and use of hidden, shadow, deprecated, unvetted third-party APIs that don't conform with or have specifications. Risks associated with these APIs include:

- Vulnerabilities that lead to data theft, fraud, and business disruption
- Elevated risk of business logic abuse or automated bot exploitation
- Susceptibility to inadvertent data exposure and regulatory noncompliance

While legacy APIs can all display these issues, the advent of agentic API-driven agents causes them to appear more frequently with ever-increasing consequences from malicious behavior.

It is said that in the land of the blind, the one-eyed man is king. Similarly, in the land of AI cyberattacks, the enterprise with sufficient visibility on the API attacks targeting them may not be king, but they are much less likely to have to disclose an expensive and reputation-damaging data breach. Cequence empowers enterprises to establish and maintain the kind of visibility and policies required to discover and mitigate both traditional and agentic AI attacks with the potential to disable key enterprise business services.

Finally, human fallibility often causes even the very best designed and implemented security systems to fail. There is a reason that many CISOs come to believe that the least secure component of their cybersecurity system sits between the keyboard and the chair, i.e., humans. Consequently, it is critical that API discovery be continuous, comprehensive, and based on known behavioral models against which API and agentic AI actions may be compared.

Cequence performs inventory and traffic analysis of an organization's internal, external (public-facing), and third-party APIs in a way that is deployable across any of their data center or cloud environments. In this manner, Cequence creates a "behavioral perspective" of active agents by dynamically discovering them, documenting who is calling them, and identifying which agents they are calling. And it does so in such a way that identified vulnerabilities are categorized by level of risk so they can be easily translated to effective security policy and prioritized remediation.

Comply

Most people associate the word "compliance" with establishing business and security policies based upon requirements imposed by regulatory or statutory requirements. In this case we're discussing something broader. While security practitioners do need to concern themselves with those externally imposed requirements, they also need to establish a robust API governance structure of which APIs may be used and how they may be used. Best practices dictate that internal policies must address four key governance issues:

Misalignment with specifications: An API, particularly an agentic AI driven API which can act autonomously may be performing the tasks assigned in such a way that creates unacceptable risk. By not complying with their design specification, these APIs may circumvent existing security policies.

Insufficient/missing authentication: Cequence can identify which APIs use appropriate authentication regardless of whether they are actively processing data. Cybersecurity teams can see exactly which APIs are vulnerable to attack based on the state of their authentication usage and can then remediate those that are vulnerable or even virtually “patch” them, making them unavailable for use until they are appropriately remediated.

Identifying sensitive data and inappropriate exposure: From public-facing APIs powering major applications to shadow API infrastructure or even internal APIs that have accidentally been exposed to external attacks, every potential point of data transmission is monitored for potential sensitive data exfiltration. Unlike nearly all other API monitoring solutions, Cequence does not rely on the customer identifying specific APIs and instead looks at API transactions forming a picture of the unique behavioral actions observed.

Detecting and addressing potential compliance violations: All enterprises of any size have external compliance standards with which they must comply or face regulatory sanctions. Whether it's PCI DSS, GDPR, HIPAA or the emerging AI regulatory frameworks CISOs must ensure their AI driven agents don't create compliance liabilities for the enterprise. Cequence provides cybersecurity teams with the visibility on what type of data is being transmitted by their APIs. As noted above, enterprises also must have access to real-time information on how well their APIs are conforming to internal governance as well as external regulatory and industry standards to ensure their APIs and agents are operating as designed.

Accidental public exposure of internal APIs: The design assumption for most internal APIs is that they will be called only by secure and known processes and agents. Consequently, they may not have the kind of built-in authentication functionality that would prevent them from exploitation by an external actor. If inadvertently exposed to an AI-driven agent, these internal APIs could expose not only sensitive internal data and processes but violate an enterprise's regulatory requirements.

Cequence ensures APIs perform as designed, comply with their specifications, are free of risks as defined in top ten lists like the OWASP Top 10 for API Security and OWASP Top 10 for LLM Applications, and any regulatory constraints under which they operate. And it does all of this without the requirement to modify applications with third-party JavaScript or SDKs. Cequence is different from fragmented or partial API security offerings because its methodology comprehends multiple types of risk across every phase of the API protection lifecycle.

Additionally, Cequence offers API security testing that allows developers to integrate API protection into their workflow. This shifts API security “left” in that it prevents risky code from being deployed into the production enterprise infrastructure. This type of testing works equally well for legacy APIs and new APIs supporting agentic AI.

Protect

The Cequence approach to keeping APIs secure involves a comprehensive approach to protecting existing and emerging AI-driven applications and data from access by malicious agentic bots. This approach fundamentally differentiates Cequence from other API security solutions that focus primarily on discovery and compliance but not protection. This allows cybersecurity teams to move much more quickly from alert to resolution without dealing with different and frequently incompatible systems. As agentic AI-driven attacks proliferate, compressing the time between detection and resolution becomes an even more important success metric for these teams.

Cequence provides protection from today's existing API attacks and the agentic AI agent attacks already in the wild by focusing on three approaches:

Receiving detailed information about potential threats:

Time may be money, but it is also the enemy of cyberattack remediation. The faster a cybersecurity team is provided with detailed and actionable data about an active threat, particularly a rapidly evolving agentic AI driven threat, the faster steps to mitigate the attacks can be executed. And since Cequence creates behavioral fingerprints of threat actors and their attacks, Cequence alerts not only include the source IP addresses involved but also includes details as to which APIs are involved and what types of data resources are being exposed. In this way, Cequence actively protects applications and data from being accessed, corrupted, or stolen by malicious AI bots.

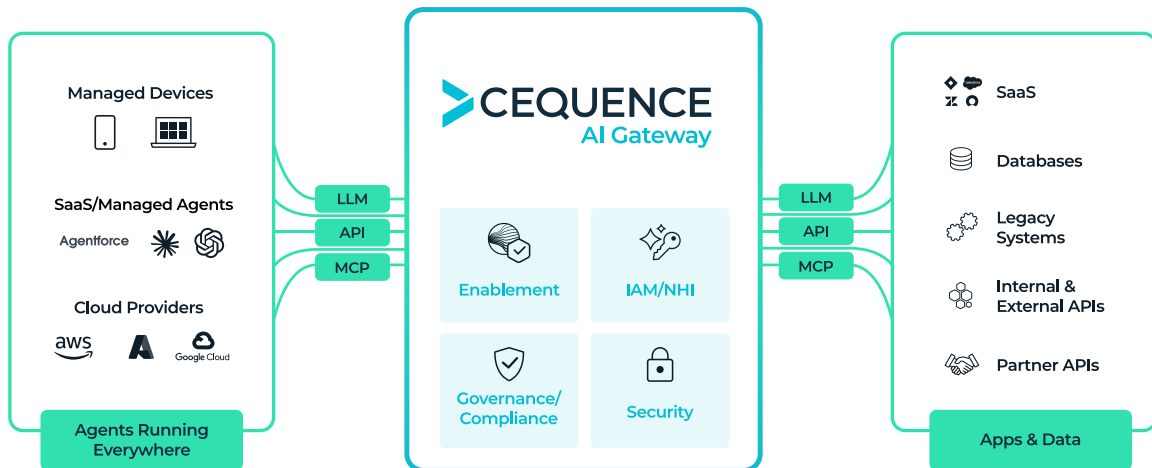
Mitigating issues quickly and effectively: The sooner that security teams are aware of a potential incident and mitigate it, the less chance threat actors will be able to exploit whatever vulnerability is involved.

Dynamic threat detection and response: Cequence utilizes multi-dimensional ML analysis to dynamically detect threats based on request profiling, behavioral analytics, and intent analysis. It then autonomously creates rules and policies to block malicious bots that can be deployed automatically or after human review.



Cequence AI Gateway

The Cequence AI Gateway is the in-path enforcement point for agent containment. It connects AI agents to enterprise and SaaS applications, APIs, data, and LLMs, and governs every action those agents take once connected. The AI Gateway extends the behavioral intent engine Cequence has used for more than a decade to deliver application and API protection, which today secures more than 10 billion API calls and 200 million agentic interactions every day. Every agent gets a defined job, every action gets verified against that job continuously, and trust is never assumed from a credential alone. Because enforcement happens at the AI Gateway rather than in the model or the endpoint, governance stays centralized no matter where agents run, and frontier, open-weight, and self-hosted models are all covered. The AI Gateway integrates with Cequence Application and API Protection, offering enhanced protection for enterprise applications and data from malicious agents and users.



Agent Personas: A plain-English job description binds each agent to the specific tools, APIs, LLMs, skills, data, and guardrails its role requires, and nothing else. Agent Personas work as the agent's job description, so security teams provision each agent like a new hire with a default-deny role. Each Persona maps to a single virtual MCP endpoint, so a customer service agent gets read-only access to the CRM rather than everything its user could reach. Personas enforce least privilege from the first tool call with no learning period, and the shorter tool list cuts token consumption and improves agent accuracy.

Behavioral Intent Monitoring: The behavioral intent engine analyzes the sequence of actions each agent takes across its full session against baselines per Persona, per user, and per tool. When an agent's actions deviate from its job, the AI Gateway throttles or halts the agent mid-action based on policy, catching the chained, individually authorized actions that single-call inspection misses.

Real-Time Inline Enforcement: The AI Gateway enforces policy inline with the request path, on every tool call, for the entire session. Per-agent, per-tool rate and spend limits and circuit breakers stop runaway loops, and network controls including IP CIDR restrictions, geo-fencing, and IP pinning restrict where each Persona's tokens can be used.

Sensitive Data Protection and Prompt Guard: Actions often pass permission checks and still exfiltrate sensitive data, so the AI Gateway inspects every request and response at runtime, with more than 100 built-in detection types to monitor, mask, redact, or block sensitive data. Prompt Guard detects prompt injection, including jailbreak and system-prompt-extraction patterns, before it reaches the model.

Create and Test AI-Ready Apps in Minutes: Cequence AI Gateway delivers easy AI agent enablement to any internal, external, or SaaS application, without coding, in minutes. Teams upload an existing OpenAPI or Swagger spec or choose from discovered application APIs, and every MCP server the AI Gateway generates inherits Persona scoping, behavioral monitoring, and guardrails automatically. Avoids time and costs associated with up-skilling, coding, QA, integration, hosting, and ongoing management.

Enterprise MCP, API, LLM, and Skill Registries: The Enterprise MCP Registry eliminates the risks of rogue MCP servers by providing a vetted catalog of servers for teams to access. Official application APIs are transformed into MCP-compatible tools with just a few clicks. Additionally, you are protected from future MCP protocol revisions, so no code updates are required as the protocol evolves. The API Registry lets agents call approved business systems, and the LLM Registry brokers access to each approved LLM provider with token rate and spending limits, so agents never have direct access to backend or provider credentials. The Skill Registry distributes enterprise-approved instructions and operating patterns that security teams vet once and reuse across every agent that needs them.

Enterprise Authentication and Authorization: AI Gateway supports passthrough and OAuth 2.1-compliant identity infrastructure, including multi-IdP support, PKCE, dynamic registration, and OIDC integration, ensuring appropriate identity-based access to systems and data while preventing unauthorized AI agent access. Each agent authenticates to the gateway with a credential that grants access to the gateway and nothing else, so a compromised agent or manipulated prompt can't leak credentials to backend systems.

Monitoring & Visibility: AI Discovery surfaces every agent, MCP server, and LLM provider already running across the enterprise, sanctioned or not, from existing SIEM logs. The AI Gateway provides real-time visibility into AI-API traffic with full audit logging enabling detailed tracking of agent and user behavior, which applications are being accessed, and what API calls agents are making. Overall activity trends and detailed MCP server and tool usage information are available at a glance. Logs export in OTEL format for SIEM and GRC integration.

Agentic AI Security Is a Job for Us All

As noted above, the threats CISOs face from agentic AI already operate in production environments, from retooled credential stuffing campaigns to legitimate agents that chain authorized actions into unauthorized outcomes. The good news is that CISOs saw this coming, and the architecture to contain it exists today. The bad news is that legacy cyber-defense strategies can't keep enterprises safe in a world in which the attacks evolve and upgrade themselves automatically based on what they learn about the defenses in place.

Using Cequence's Discover-Comply-Protect framework, the new vulnerabilities that the adversary exploits with their own AI-driven agents can be quickly enumerated, prioritized, and remediated before they become active attacks. Cequence's API security, bot management, and AI Gateway products and services can also accelerate the timeline of getting effective detection and mitigation solutions in place to get the enterprise security posture in place required for the threat landscape we are all facing.

For the agents enterprises deploy themselves, identity is the starting point for governance. However, identity is only the start. Treat each agent like a digital employee: onboard it, give it a specific job description, grant the access that job requires and no more, evaluate its behavior against that job, and offboard it when the work is done. Agent containment, built on per-agent least privilege access and behavioral monitoring and enforced in real time at a central control point, keeps every agent inside the job it was assigned and lets organizations put agents into production at full speed.