CEQUENCE

Whitepaper

A CISO's Guide to Agentic AI Security

Executive Summary

Enterprise application development is undergoing its most significant change in a generation as artificial intelligence (AI) driven agents begin to augment and even replace legacy applications. While this change will deliver more actionable data faster to decision makers and create new ways to automate labor-intensive tasks, this trend comes with significant security risks. In much the same way that aerial drones are fundamentally changing the way in which armed conflict is waged, AI-driven (agentic) agents will change the way cybersecurity "wars" are fought.

While agentic AI will revolutionize the way enterprise systems are developed and deployed, they will also be misused by cybercriminals to retool their existing attack methodologies and create new, more dangerous, and adaptable classes of attacks. And, as enterprises create new agents that depend even more on APIs than the applications they replace, CISOs will need to ensure the APIs in use by agentic AI based systems are secure from malicious attacks. The Cequence API Protection (UAP) platform empowers cybersecurity teams to identify and mitigate these new attacks as they appear.

AI-Based Agents Are Coming 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 will enable new and more efficient cyberattacks, having 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 will impact cybersecurity in general and API security specifically, we need to define some terms. And before we can define agentic AI, we need 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 110 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 has 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

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

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.

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

Example 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 will not be materially impacted by the deployment of agentic AI solutions as API use becomes the dominant form of data transactions. In fact, AI-driven agents will significantly increase 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 will replace.

Given the opportunities presented by agentic AI, it's little surprise that it is being embraced by forward-thinking

Both cybercriminal gangs and cybersecurity practitioners and the vendors that serve them are beginning to deploy agentic AI based agents and this trend will accelerate and expand in the coming years.

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 economy by 2030 only 20% of enterprises currently have effective API security measures in place[3] to address current API threats, let alone the agentic AI threats they know are coming.

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 will be 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 are beginning to deploy AI-based agents and this trend will accelerate and expand in the coming years. We're already seeing AI-driven phishing and ransomware attacks and as the agents driving these attacks learn more about their target victims, they will only 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 are expected to use agentic AI for attacks in two different ways. First, they will 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. There is a second class of never-before-seen attacks that will begin to emerge in the near future that will only be possible by using agentic AI.

Agentic AI won't change the goals of this first class of attacks or the types of vulnerabilities they'll exploit. However, from the threat actor's perspective they will 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 will allow cybercriminals to fully automate the process of gathering information about a target's attack surface to identify vulnerabilities they can exploit. It will empower 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 won't change very much, agentic AI will allow attackers to do a much better job of circumventing existing defense tactics such as lock out policies and allow 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 will 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 will 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 adapting their approach based on API responses.
- **Data & resource theft:** This is one area where agentic AI based attacks will really shine as agents will be able to very quickly identify sensitive data at rest and in motion and then steal it or alter it in place.

2 Gartner, Inc. - Top Strategic Technology Trends for 2025: Agentic AI – October 2024
3 Sources unknown
4 Jim Malone – The Untouchables – 1987

- **Business logic abuse:** These attacks appear as valid interactions because the attacker is exploiting intended app or API functionality, which also enables them to bypass traditional security solutions without detection. These attacks can be automated and massively scaled through bots potentially leading to data loss, theft, or fraud.
- **Obfuscation of malicious behavior:** This is another area where agents will be 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.
- **AI enhancements of existing malware families:** Agentic AI will allow 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."[5]

The new emerging class of attacks that will leverage agentic AI will include, but not be 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) will only 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.

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 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.

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. Research done by Cequence Security reveals that 70% of online transactions are API based[6]. This number will only go up as agentic AI based business systems are deployed and API use becomes the dominant form of data interaction.

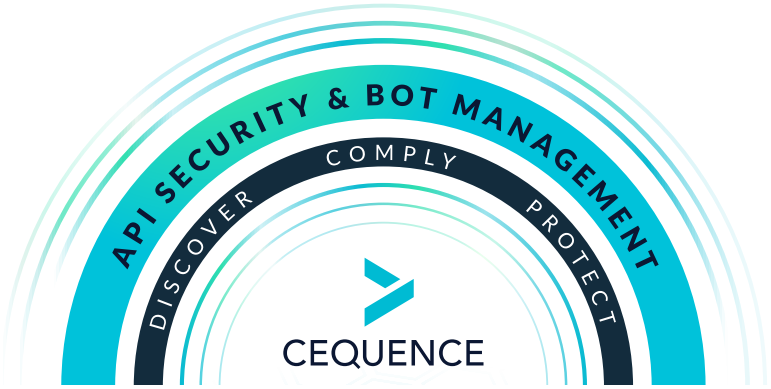
5 Gartner, Predicts 2025: AI's Impact on the Future of Enterprise Technology, 18 Mar 2025, Arun Chandrasekaran et al., https://www.gartner.com/document-reader/document/6273683

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.

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.

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.



Cequence Unified API Protection Platform

The Cequence Unified API Protection (UAP) platform uses a three-part framework to protect legacy APIs to discover and protect the APIs currently 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, AI, and 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

6 https://www.cequence.ai/news/cequence-security-releases-report-revealing-top-3-attack-trends-in-api-security/

While legacy APIs can all display these issues, the advent of agentic API driven agents will cause them to appear more frequently with ever-increasing 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's UAP platform 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.

The Cequence UAP platform 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 UAP 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:
The UAP platform 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 then 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 UAP 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.

The Cequence UAP platform 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. The Cequence UAP platform 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, the Cequence UAP platform 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 coming agentic AI agent attacks by focusing on three four 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 UAP 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 UAP 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 exploit whatever vulnerability is involved.

Dynamic threat detection and response: Cequence UAP 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.

Agentic AI Security Is a Job for Us All

As noted above, the good news is that the threats CISOs face from agentic AI cyberattacks are not immediate, but we will be seeing them soon. Unlike many threats, CISOs won't be surprised when the first ones start appearing in the wild, or at least they shouldn't be. The bad news is that legacy cyber-defense strategies will not 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 will exploit with their own AI-driven agents can be quickly enumerated, prioritized, and remediated before they become active attacks. Cequence's API security and bot management 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 know we will all be facing.

5201 Great America Pkwy, Suite 240, Santa Clara, CA 95054 | 1-650-437-6338 | info@cequence.ai | www.cequence.ai