# Accelerating safely with continuous agentic security

## How to move faster and test thoroughly – without breaking things

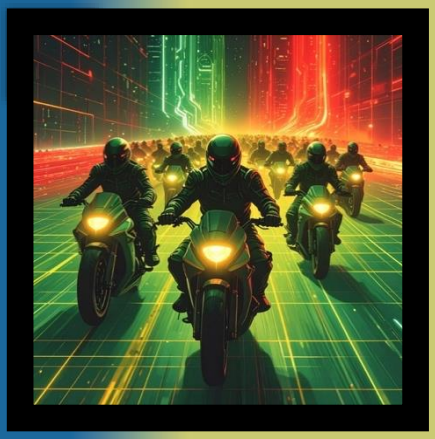
*An Intellyx Brief White Paper for Ridge Security*
*by Jason 'JE' English, Director & Principal Analyst*

**August 2026**

## *Turning agents into a security advantage*

Whether we are ready for it or not, agentic AI is being adopted by organizations at an unprecedented rate. McKinsey reported that 88% of enterprises have already used agents in at least one business function, and Gartner predicts 40 percent of companies will run software with embedded agents in 2026, up from just 5 percent in 2025.

*Optimistic executives would love to be able to trust agents as flawless co-workers that enhance everyone's productivity. Agents can work faster, but they can also make mistakes faster, and expose novel vulnerabilities, especially when deployed at scale.*



The resulting speed of agentic implementation is far outstripping the security organization's ability to secure the resulting expanded attack surface of customer-facing application environments. Even security testing itself can become a risk and take down production systems when pentesting agents are used.

In this brief paper from Intellyx, CISOs, SecOps leads, and development teams will learn strategies for safely employing agentic security measures, including offensive agent testing, to counteract the chaos and cut through the noise with insight.

## Why agentic security creates new challenges

Whether they are deployed on a user's laptop, on a core system, or running in a cloud service, agents are already becoming ubiquitous across our application environments. They may be used for just about any purpose, from developer coding, to customer service, analytics, financial calculations, observability, and security automation.

Risk-averse IT executives will seek to employ stricter governance standards to prevent unvetted agent use, but this must be balanced with the business demands of achieving faster agentic AI adoption and quick ROI wins. Since nobody wants to be a roadblock to

progress, security teams become the last line of defense for closing down agentic vulnerabilities.

## Agentic security challenges

- **Cybersecurity teams are understaffed** across the board, with 4.8 million jobs unfilled globally. Most companies don't have the luxury of security personnel who are ready to validate agentic scenarios at the point of implementation, so there is no proof that applications will survive a flood of outside agentic attacks.
- **Teams are underequipped** to conduct security exercises. Conventional security tools don't recognize novel agentic threats such as indirect prompt injection, excessive agency, and cognitive hijacking, and access to safe production-like environments to test in is severely limited.
- **Overwhelming alert data** makes spotting agentic security issues nearly impossible. Agents can exhibit non-deterministic behaviors, especially when interacting with other agents and systems in production. Procedural tests looking for predefined attack chains return false positives that obfuscate higher-risk agentic attack vectors.
- **Generalized AI security investigations are unsustainable at scale.** Typing a carefully worded seed prompt into an unspecialized GenAI such as Claude or ChatGPT, or asking a coding agent to "check your own work" won't reliably uncover agentic risks that can make a real impact. Besides that, the token costs of pentesting with generalized AI tools will break the budget if used at scale.

To provide context, we don't just need a "human in the loop" to make endlessly repetitive decisions as agentic loops run. Humans need to respond and provide business context – ideally at the exact point changes and agentic activities identify and escalate a significant threat.

# Meeting the demands of agentic security

New threats are emerging every day thanks to the widespread adoption of GenAI and agents, The OWASP Top 10 report demonstrates how agents can be compromised, misled, or hijacked for nefarious purposes.

Most conventional TEM (threat exposure management) and XDR (extended detection and response) security tools are pretty good at running daily or weekly scans to discover

instances of known exploits that would appear as CVEs in MITRE ATT&CK. To discover the unknown vulnerabilities agents can introduce, we need security agents that are continuously running and laser-focused on finding and resolving AI-specific threats.

Rather than relying on an all-in-one platform approach to security that rests upon a massive data lake of earlier findings, we need to orchestrate a new class of security agents that are highly specialized to understand agentic behavior, so that agent-borne threat vectors can be triaged, evaluated, and resolved in real time according to risk.

Since agentic development tools are evolving fast, separating concerns should encourage the right agent, model, and data combination for each job in terms of specialization for target type, speed, accuracy, effectiveness and cost. For instance, one agent to orchestrate offensive test runs, more agents to execute tests for certain scenarios, another to take in results, and others that evaluate reachability and predict blast radius. All of these agents should provide a feedback loop into the organization's agentic SOC and ITSM platforms for tracking and auditability.

While there's no substitute for continuously monitoring and validating live systems, pointing a flock of agents at production is a recipe for disaster. 95 percent of agentic security testing should be done in a safe, bounded environment to prevent data leakage, latency, and unexpected impact on critical applications.

# How Ridge Security delivers continuous agentic security

Ridge Security offers a CTEM (Continuous Threat Exposure Management) security platform that installs within your enterprise network and orchestrates offensive test automation runs, providing analytics and observability for validating findings, and prioritizing and remediating vulnerabilities.

Conventional pentesting practices are often run in weekly or daily intervals, and involve both manual human test runs and automated scripts or 'bots' that conduct sequences of attack chain behaviors against pre-production or production systems, often looking for previously encountered exploits and known CVEs, then delivering findings for human teams to triage and figure out how to remediate.

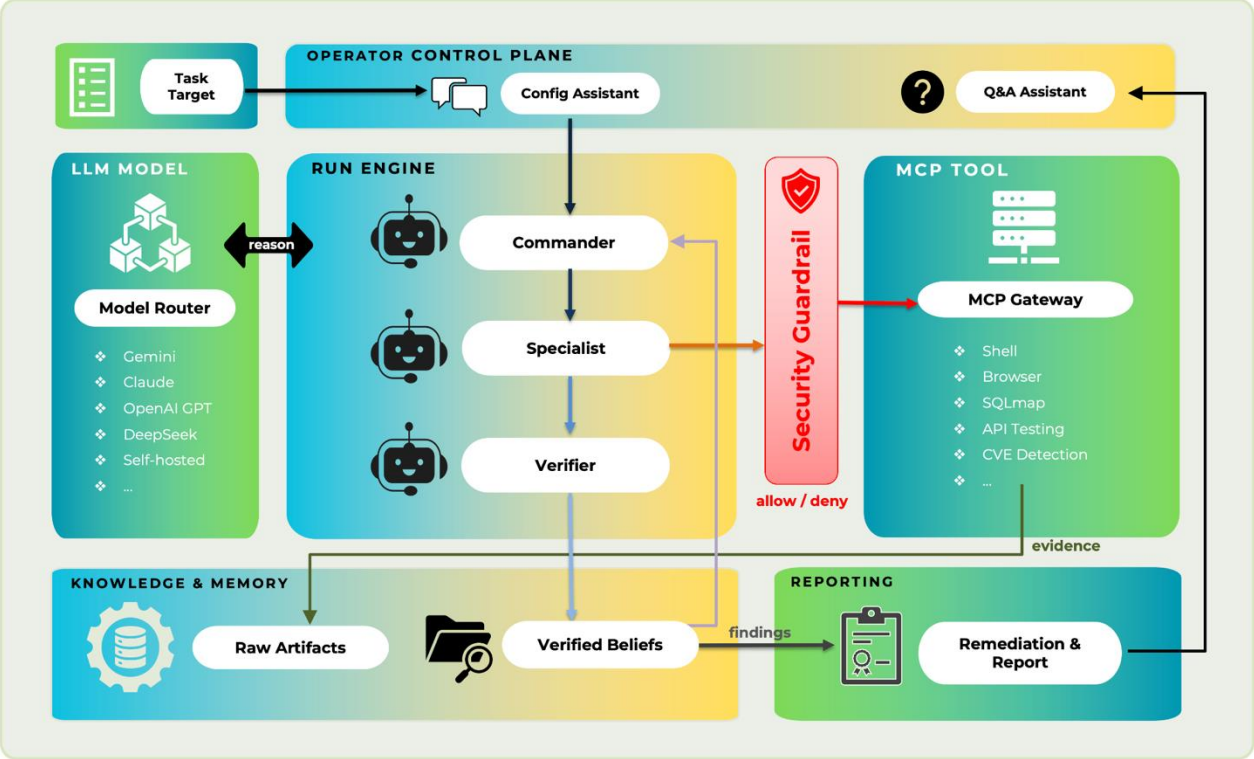



*Figure 1. RidgeGen offensive agents are equipped with tools, expert knowledge and context for finding and exploiting specific application, system and network-level vulnerabilities.*

To meet the unique needs of continuous agentic security testing, their platform is paired with their new RidgeGen autonomous AI product, which seeds and customizes a pool of expert offensive attack agents that can take action against the system under test. Each agent is armed with system knowledge, business context, user goals, and attacker intent, while the system maintains secure guardrails over their system access and behavior.

## Safely governing the agentic testing lifecycle

To protect critical systems, Ridge replicates the production environment in a ***fixed safety envelope***. This agentic testing space is sort of like a pre-production environment you might push out to test a deployment in a CI/CD pipeline, but with contractual boundaries that prevent destructive agent actions and data from getting promoted or escaping the envelope before findings can be investigated and resolved or dismissed.


*Figure 2. RidgeGen fixed safety envelope with different levels of governance and human oversight.*

There is a safe or 'green zone' for safely bounded testing with read-only obfuscated data, ideal for validating the 'as is' state of a production system as it is accessed by human or agentic users.

The envelope also includes a supervised or 'amber zone' that provides partial exposure to test more system resources with human approval, and an aggressive 'red zone' requiring direct human oversight of the most malicious types of agentic test runs that could have a large blast radius in the real world.

## The Intellyx Take

Whether driven by top-line improvement goals for better customer experience and agility, or bottom-line efficiency and productivity metrics, agentic AI is a promising value lever companies are pulling today.

Agentic systems can behave in unexpected ways, especially when exposed to swarms of agent attacks. For agentic initiatives to survive past the pilot project stage, we need a way to continuously test this rapidly emerging class of software before it falls on its face in production.

Rather than expecting short-handed security teams to find and fix agentic vulnerabilities in ever-changing production environments using conventional pentesting procedures, why not turn agents to our advantage without the risk?

Continuous agentic security testing doesn't just put a "human in the loop." Expert offensive agents can give security teams more context, less noise, and better visibility of the most significant risks, so everyone can breathe a little easier with each push to production.

## About the Author

**Jason "JE" English** is Director & Principal Analyst at Intellyx. Drawing on expertise in designing, marketing and selling enterprise software and services, he is focused on covering how agile collaboration between customers, partners and employees accelerates innovation.

A writer and community builder with more than 25 years of experience in software dev/test, cloud computing, security, blockchain and supply chain companies, JE led marketing efforts for the development, testing and virtualization software company ITKO from its bootstrap startup days, through a successful acquisition by CA in 2011. He co-authored the book *Service Virtualization: Reality is Overrated* to capture the then-novel practice of test environment simulation for Agile development. Follow him on LinkedIn.

## About Ridge Security

**Ridge Security** delivers autonomous cybersecurity validation solutions that help organizations proactively manage risk and improve resilience. It develops agentic AI-based adversarial risk platforms that support continuous threat exposure management programs. Recognized by Gartner in the Market Guide for Adversarial Exposure Validation, Ridge has earned industry honors including the 2026 Frost & Sullivan Global New Product Innovation and Top Emerging Cyber Security Company. The company serves customers worldwide across finance, government, telecom, and enterprise sectors. For more information, go to https://ridgesecurity.ai/.