Emerging Tech Impact Radar: AI Cybersecurity Ecosystem

17 October 2025 - ID G00835447 - 76 min read

By: Mark Wah, David Senf, Alfredo Ramirez IV, Tarun Rohilla

Initiatives:Emerging Technologies and Trends Impact on Products and Services; Increase Product Traction

AI presents significant risks and opportunities, spanning the protection of AI systems and the application of AI to enhance security. Product leaders must decide how to harness emerging technologies to secure the AI life cycle and leverage applied AI to accelerate and differentiate security outcomes.

Overview

Key Findings

- C-level product executives need to act quickly to take advantage of multiple points of security market disruption as AI ushers in rapid security consolidation trends.
- Securing the AI engineering pipeline requires a holistic, governance-guided approach that integrates AI supply chain security for component integrity, AI code security assistants for AI-generated code, information governance to manage data risks, and AI security testing to validate AI-enabled applications, AI models and guardrails before deployment.
- Securing AI in runtime requires robust, real-time defense mechanisms to protect a diverse range of AI implementations, from built and blended to embedded and external AI services, against a broad range of adversarial attacks that include prompt injection, misuse, and data leaks.
- Applied AI represents opportunities to differentiate AI outcomes in security where synthetic data, security-tuned domain-specific language models (DSLMs) and intelligent simulation can support high-value and effective outcomes.
- Agentic AI represents diverse implementations with building blocks such as prebuilt AI agents, AI agent builders and the evolution toward guardian agents as a longer-term approach to secure AI agents at scale.

Recommendations

- C-level product executives must prioritize and decide on areas of AI investments with urgency based on the mass and impact of AI capabilities that align with their target and adjacent markets by prioritizing high-impact profiles to explore first, such as AI governance and AI usage control.
- Address preproduction AI risks by developing an effective feedback loop between policy-setting capabilities, like AI governance and information governance, and validation capabilities, like AI security testing.
- Reduce runtime exposure across diverse AI implementations by designing an extensible architecture that integrates continuous runtime protection with capabilities to handle multimodal inputs and secure complex agentic AI workflows.
- Build a roadmap that includes applied AI capabilities to expand the effectiveness and speed of agentic-based security outcomes across security domains by leveraging security-tuned DSLM and synthetic data to differentiate.
- Accelerate development of agentic AI features with a high level of abstraction and architecture flexibility to pivot quickly when protocols, architectures and use cases shift dramatically from this very early stage of market adoption.

Strategic Planning Assumptions

By 2027, AI governance will be integrated into 75% of platforms, making them a main area of competition.

By 2028, 80% of data for AI will be synthetic to simulate reality and future scenarios and to derisk AI, up from 20% in 2024.

Through 2029, over 50% of successful cybersecurity attacks against AI agents will exploit access control issues, using direct or indirect prompt injection as an attack vector.

By 2030, more than 60% of enterprises will secure the AI life cycle and various AI implementations within an AI security platform, up from less than 10% in 2025.

Analysis

This Impact Radar analyzes the emergence, market momentum, and influence of new technologies and trends related to the AI cybersecurity ecosystem. It serves to guide product leaders on their strategy development and product roadmap planning regarding relevant AI cybersecurity ecosystem technologies and services, considering their velocity to early majority adoption (defined as more than 16% market adoption) and their expected impact on securing AI and applying AI within security. (See Note 1 for an explanation of the methodology we adopted for this research.)

Gartner identified four central themes across 18 emerging technologies and trends for the first Impact Radar for the AI cybersecurity ecosystem:

- Securing AI along the AI engineering pipeline
- Securing AI at runtime
- Applied AI in security
- Agentic AI

Secure the AI Life Cycle Effectively With Broad AI Engineering Pipeline Security Integrations

Securing the AI engineering pipeline is a holistic approach guided by AI governance, which has expanded from simple inventorying to actively overseeing critical implementation elements. This broad oversight extends to AI supply chain security and AI security posture management (SPM), which are important for discovering, evaluating and tracking a wide range of artifacts from models to applications and AI configurations. A central component is information governance, which manages risks across the data life cycle and can extend to runtime monitoring. AI code security assistants can be leveraged for efficient inspection and analysis of both AI-generated and human-written code. Ultimately, this entire process is validated through AI security testing, which provides an offensive security perspective to uncover vulnerabilities in both models and applications. Additionally, sovereign AI represents a broader, national-level effort where countries control their own AI development to advance unique strategic objectives.

Effective Runtime Security Must Address All AI Implementations, From Built AI to Embedded AI and External Services

Securing AI at runtime addresses a broad range of implementations spanning AI controlled by an organization, including built AI and blended AI approaches (e.g., DSLM and hosting AI models), embedded AI within an application (e.g., Microsoft Copilot, prebuilt agents) or consumed as an external service (e.g., ChatGPT, Google Gemini). AI runtime defense provides a foundation for enforcing policy and detecting anomalies like prompt injection for internally controlled AI. This is complemented by AI usage control (AI-UC) for enforcing organizational security policies to mitigate misuse and data leaks of embedded AI and external services. These defenses are made more effective by multimodal AI protection across data, text, images, video, and audio to uncover threats invisible to traditional tools.

Given the need to align individual capabilities for consistent AI policy enforcement and guardrails, a few profiles — such as AI SPM, AI usage control, AI runtime defense, and AI security testing — are being supported collectively as an AI security platform.

Applied AI Is Crucial to Differentiating Security Outcomes

Applied AI within cybersecurity is modernizing security by combining several technologies. Synthetic data is artificially generated to achieve scale and provide extensive coverage for edge cases and future scenarios. It can enable the safe and thorough testing of diverse attack scenarios within intelligent simulation implementations for security. To further assist humans, AI assistants use generative AI to automate routine tasks, summarize threat data, and suggest fixes. These assistants can be powered by specialized DSLMs trained on cybersecurity data to overcome hallucinations and provide more efficient and cost-effective operations. Together, these tools form a cohesive system that significantly improves the speed, accuracy, and efficiency of cybersecurity defenses.

Evolution Within Agentic AI Unlocks the Ability to Tackle Unaddressed Exposures and Incidents

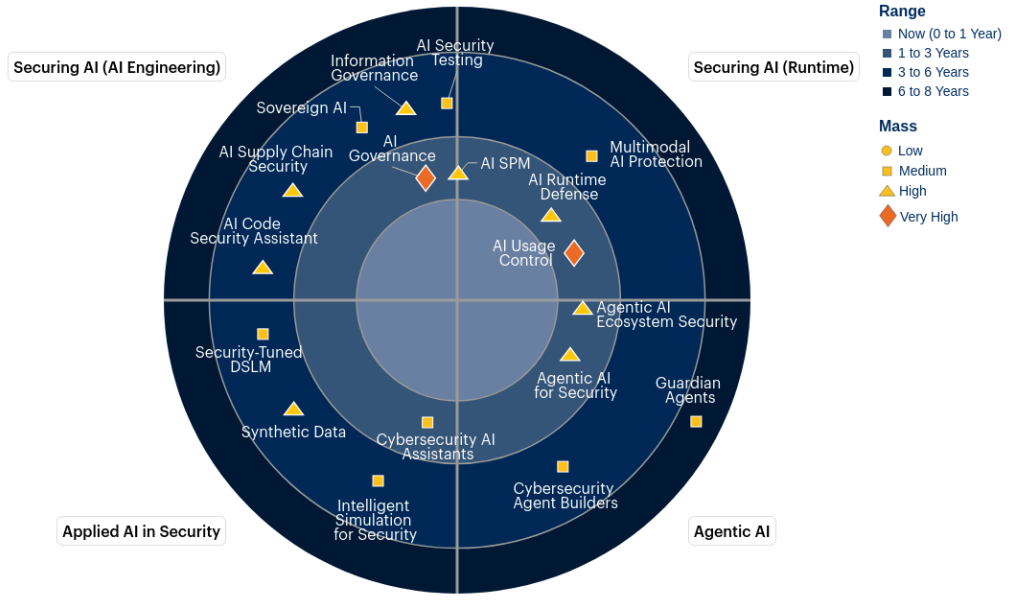
The integration of AI assistants into cybersecurity has seen mixed outcomes, yet they are a key stepping stone toward improved automation. Agentic AI is being applied in various security domains, including SecOps, AppSec, cloud security, and identity management, with the security operations center (SOC) remaining a primary focus due to challenges with staffing and round-the-clock coverage. Innovators are developing platforms that enable users, from nontechnical individuals to expert developers, to build agents for different security applications. As the number of AI agents grows, a new concept of guardian agents is emerging. These guardians, along with orchestrators, are designed to monitor and validate the behavior, outcomes, and overall security of the AI agents they oversee, ensuring they remain secure and effective.

Figure 1 depicts the 18 emerging technologies and trends across the four different themes in the respective quadrants within this impact radar.

The Impact Radar

Figure 1: Emerging Tech Impact Radar: AI Cybersecurity Ecosystem

Emerging Tech Impact Radar: AI Cybersecurity Ecosystem



Source: Gartner

As of October 2025

Gartner

The objective of this research is to guide product and service leaders on how emerging technologies and trends (ETTs) are evolving and impacting areas of interest in the AI cybersecurity ecosystem. Providers can leverage this knowledge to determine which technologies or trends are most important to their business's success and when it makes sense to advance their products and services by investing in them.

Product leaders should use the Radar Profile range to plan investment timing in the related emerging technology or trend. "Range" represents Gartner's estimate of the time it will take each supporting technology to reach early majority adoption (more than 15% target market adoption), not when product leaders should act on an investment. Considering the time to plan for, develop, and launch, a starter guide to product leader investment timing, based on product strategy, is as follows:

- First movers should be acting now on items in the six-to-eight-years ring (or beyond).
- Fast followers should be acting now on ETTs in the three-to-six-years ring.

- The majority of followers should be acting on ETTs in the now and one-to-three-years rings.
- Laggard followers can wait until the ETT has passed through to early, or even late, majority.

Refer to the “About the Impact Radar” section for more information.

Emerging Technologies or Trend Profiles

The Priority Matrix lists emerging technologies and trends identified in the Impact Radar for AI cybersecurity ecosystem according to their range (see the About the Impact Radar section for our methodology). Click on a technology name in the table to jump to a profile of the Emerging technology or trend.

Table 1: Priority Matrix for AI Cybersecurity Ecosystem

(Enlarged table in Appendix)

Mass	*Range*			
↓	*Now (0 to 1 Year)* ↓	*1 to 3 Years* ↓	*3 to 6 Years* ↓	*6 to 8 Years* ↓
Very High		AI Governance AI Usage Control		
High		Agentic AI Ecosystem Security Agentic AI for Security AI Runtime Defense AI Security Posture Management	AI Code Security Assistant AI Supply Chain Security Information Governance Synthetic Data	
Medium		Cybersecurity AI Assistants	AI Security Testing Cybersecurity Agent Builders Intelligent Simulation for Security Multimodal AI Protection Security-Tuned DSLM Sovereign AI	Guardian Agents
Low				

Source: Gartner

1 to 3 Years

Agentic AI Ecosystem Security

Analysis By: Alfredo Ramirez IV, Mark Wah

Definition:

Agentic AI ecosystem security is an integrated framework designed to protect and govern advanced AI systems that operate with a high degree of autonomy within multiagent systems (MAS). This includes the AI agent platform, the AI agent patterns, behaviors and characteristics, identity and access management (IAM) implications and coverage of emerging protocols such as model context protocol (MCP) for tool use and Agent2Agent (A2A) protocol for interagent communications. The ecosystem extends to integrations and AI agent tool use that includes misconfigurations. Each of these represents attack vectors that could be taken together to approximate the blast radius of a compromised agent.

Sample Vendors

Aembit; Astrix; Linx Security; Google; Microsoft; Noma Security; Palo Alto Networks; Prompt Security (SentinelOne); Zenity

Range: 1 to 3 years

The range is one to three years to reach early majority adoption due to the rapid adoption of AI agent implementations on diverse AI agent platforms.

While 53% of organizations have already deployed custom-built AI agent automation as of 2025, indicating significant adoption of AI agents themselves, the comprehensive "Agentic AI Ecosystem Security" framework, which uses Guardian Agents (See Guardians of the Future: How CIOs Can Leverage Guardian Agents for Trustworthy and Secure AI) and integrates protective measures across the entire ecosystem, is still emerging. Most implementations provide partial coverage from the communications, network security and MCP server vulnerabilities perspective. Few focus on the identity, agent memory, footprint, and behaviors to secure multiple conversations and multistep processes. Given the broad spectrum of AI agent platforms, tight integrations between Guardian Agents and these platforms, such as Microsoft and Google ecosystems, are required.

A rapid surge of interest in AI agents has fueled growth for agentic AI ecosystem security, reflected by an increase in Gartner AI-agent-related inquiries. The increasing complexity of agentic AI security risks has spurred venture capital interest in real-time platforms for securing AI agents, identities, and applications. The rapid rise of MCP servers that front-end many emerging and established services is another important indicator. The market for broader AI security solutions is anticipated to scale as enterprises seek comprehensive protection for their AI investments.

Mass: High

The mass is high because agentic AI ecosystem security shifts the focus from how organizations manage and protect their digital assets and operations to how they protect them.

Agentic AI ecosystem security impacts how organizations manage and protect digital assets and operations across numerous industries and business functions as agents are adopted. It will be critical for highly regulated industries such as banking, capital markets, insurance, public sector and healthcare, where access to sensitive data and outcomes by AI agents is heavily scrutinized.

Agentic AI ecosystem security is revolutionary for adopters and providers, requiring new capabilities such as AI agent discovery, observability and control that will augment and replace existing capabilities. Traditional security controls are inadequate for agentic AI systems due to their dynamic behaviors, external dependencies, and real-time decision-making capabilities, which will require new safeguards.

Recommended Actions

- Gain early credibility as an agentic AI ecosystem security vendor by helping customers gain comprehensive visibility into AI agents through AI agent platform integrations, traffic inspection and API and MCP call monitoring.
- Enhance dynamic access controls and implement real-time runtime defenses for agent interactions by deploying strong, dynamic authentication methods such as policy-based access control and utilizing AI runtime defense solutions to detect prompt injections and behavioral anomalies.

Gartner Recommended Reading

How to Secure Custom-Built AI Agents

IAM for LLM-Based AI Agents

Agentic AI for Security

Analysis By: Alfredo Ramirez IV

Definition:

Agentic AI for security is the use of semiautonomous or autonomous multiagent systems that augment the cybersecurity workforce. Unlike cybersecurity agent builders, which provide tooling for creating and integrating cybersecurity agents, this trend is for self-contained solutions with little or no exposure to the underlying agent creation or design for the end user. Agentic AI for security is focused on doing the digital labor in cybersecurity that is not being done today due to a lack of available talent or lack of resources to hire talent. As such, it mainly works with existing cybersecurity solutions rather than attempting to replace them.

Sample Vendors

Command Zero; Crogl; Dropzone AI; Exaforce; Intezer; Imperum; Shift Security; Simbian; Tuskira; Twine Security

Range

The range for agentic AI for security is one to three years from reaching early majority adoption mainly due to the need to establish trust in the reliability of these systems. However, increasing investment and emerging vendors indicate that the market is bullish on the success of this model in the near to midterm.

Signals from early adopters indicate that it can provide coverage and efficiency gains for security operations center (SOC) teams that have been dealing with a constant state of too few experienced staff members and too many security alerts over the last three to five years. Many alerts below a certain risk level, which is not always correct, go unworked due to lack of resources. Therefore, any system that can help organizations investigate all alerts to a reliable outcome is seen as a welcome addition.

Growth in the last 12 months has been exponential. Incumbent cybersecurity vendors are investing in go-to-market activities, claiming they are now agent-powered. Many new vendors are coming to market with significant funding and claiming to be agent-powered by design. Cybersecurity vendors in general are not new to AI usage. Most have been using various forms of machine learning technologies to aid in things like classification, triage and recommendations for years. GenAI adds the ability to do last-mile human work — including asking questions from other humans, writing up tickets and researching alerts — that has been missing from these systems' ability to complete investigations.

Mass

The mass of agentic AI for security is high because almost all organizations with a digital footprint have cybersecurity products generating alerts that require investigation, which could enhance security by increasing coverage.

Given the speed at which GenAI tooling has made it easier to conduct cyberattacks, most organizations and industries will consider adopting for the efficiency and automation enabled by agentic AI for security. This assumes that these solutions are made reliable, transparent and trustworthy and that they are available at price points amenable to multiple organizational sizes and budgets.

Although many cybersecurity vendors have utilized machine learning and other forms of AI in their products for years, agentic AI for security represents fundamental new capabilities that will replace existing capabilities for incumbent vendors. Examples of this include log normalization for agent context — in some cases bypassing security incident and event management (SIEM) — and playbook creation at runtime based on context rather than brittle static playbooks. Also, significantly, new companies are being created solely to architect and bring to market cybersecurity solutions built with agentic architecture from the ground up.

Recommended Actions

- Gain early market share by emphasizing how agentic AI for security solutions expand coverage of any SOC team to include investigating alert levels that previously had to be ignored due to resource limitations.
- Expand the addressable market by developing a go-to-market plan that is focused on how your agent-powered solution can serve as a SOC team for organizations that otherwise do not have the resources to hire humans for that purpose.
- Increase relevance to buyers by expanding agentic capabilities across the full SOC life cycle from detection and alert triage to incident investigation and response.

Gartner Recommended Reading

Emerging Tech: 'Time to Trust' Is the New Vital Agentic AI Metric

Emerging Tech: Customer Trust Is a Critical Barrier to Agentic AI Adoption

AI Governance

Analysis By: Tarun Rohilla, Tushar Jain

Definition:

AI governance is the process of creating policies, assigning decision rights and ensuring organizational accountability for risks and decisions for the application and use of AI techniques. AI governance is part of adaptive data and analytics governance, addressing the predictive and generative capabilities of AI.

Sample Vendors

Cranium; Credo AI; Fairly AI; FICO; Holistic AI; IBM; ModelOp; Monitaur; Saidot; SolasAI

Range

The range for AI governance is one to three years because of the evolving AI technologies and increasing regulatory pressure, indicating a relatively near-term horizon for significant market penetration.

We estimate AI governance to be very close to early majority adoption. The key drivers are generative AI (GenAI) adoption, which is aggressively pushing enterprises to establish and evolve governance frameworks. This is accelerating its current relevance and ongoing adoption. Global data, AI regulations and compliance requirements, such as the EU AI Act and various U.S. executive orders, are making AI governance goals more concrete and add urgency to defining responsible AI procedures. By 2027, AI governance is projected to be a requirement of all sovereign AI laws and regulations worldwide.

The growth in velocity for AI governance is fast and is driven by the following:

- The rapid evolution of AI, particularly GenAI, compels organizations to quickly establish and adapt their governance strategies. AI initiatives are increasingly becoming top-down directives from CEOs.
- More than two-thirds of organizations are opting for centralized AI governance structures and consolidating AI capabilities, which signifies rapid internal acceleration in adoption.

- Globally, inconsistent but urgent regulations are driving a concrete push for defining responsible AI governance procedures.
- AI governance increasingly includes security, compliance and governance controls in model development and deployment (e.g., SoC2 for AI, ISO 42001).

Mass

The mass for AI governance is very high, primarily due to its influence across all organizational functions, industries, and geographies. This is also driven by the mission-critical and strategic nature of AI driving risk-based governance, transparency and human oversight within the enterprise and the need for comprehensive oversight.

The breadth of impact of AI governance is very high as it spans across diverse industries, geographies, and multiple business units within enterprises. It is essential for managing applications, models, and AI agents, as well as critical areas like risk management, privacy, regulatory compliance, trust, transparency, data quality, technology deployment, and workforce roles. AI is becoming “mission-critical” and exists “throughout the enterprise,” making governance non-negotiable. Regulations are emerging globally, including in the EU, U.S. and China, which regionalizes governance goals and adds to the urgency for defining good practices.

AI governance is a revolutionary shift for both adopters and providers, fundamentally altering existing governance paradigms. Unlike deterministic technologies, AI’s unique probabilistic and generative nature necessitates new policies, participants and oversight methods, requiring continuous feedback loops for unexpected outcomes. This profound change is evidenced by high search interest for “generative AI governance,” “AI governance frameworks,” and “AI governance platforms,” highlighting the demand for entirely new structures and technological support. While leveraging existing governance mechanisms is a best practice, AI governance requires specific additions like trust, transparency and diversity pillars be applied across people, data and techniques. The widespread adoption of centralized AI governance structures further signifies this significant organizational transformation.

Recommended Actions

- Embed AI governance capabilities into your platform elements such as data, analytics rather than offering siloed add-ons. Ensure that your solution supports AI-specific requirements like trust, transparency, which can be tailored to align with the client's existing frameworks.
- Develop and prioritize AI trust, risk and security management (TRiSM) features within your product suite to help customers enforce AI governance policies across a variety of use cases. Provide built-in tools for continuous monitoring, automated testing, validation and compliance reporting throughout the AI life cycle.
- Offer frameworks within your platform that allow clients to define use cases criticality, helping them to focus on governance of high-impact AI deployments allowing for flexibility across scenarios.
- Scale GenAI governance controls for hallucinations, risk, prompt injection defenses, intellectual property (IP) and security compliances before models are deployed or exposed to users.

Gartner Recommended Reading

Executive AI Governance Playbook

Hype Cycle for Artificial Intelligence, 2025

AI Runtime Defense

Analysis By: David Senf

Definition:

AI runtime defense implements intent-based policy enforcement and anomaly detection for AI applications and models. It offers content inspection for application abuse and attacks (e.g., prompt injection) and aims to detect intent or content anomalies (toxicity, hallucinations). Some AI runtime defense tools are application-specific (e.g., chatbots). AI runtime defense is a common feature of AI governance platforms but can also be a stand-alone product when focused on cybersecurity. Solutions typically involve continuous, real-time monitoring to enforce guardrails, provide risk scoring, and initiate basic responses such as redacting personally identifiable information (PII), triggering alerts and blocking malicious activity. Although threat actor breaches of large language models (LLMs) have been low so far, there is significant potential for sensitive data oversharing and customer-facing abuse of generative AI (GenAI)-based chatbots and applications.

Sample Vendors

Cato Networks (Aim Security); Check Point (Lakera); F5 (CalypsoAI); HiddenLayer; Lasso; Noma Security; Palo Alto Networks; Pillar Security; SentinelOne (Prompt Security); TrojAI

Range: 1 to 3 Years

The range for AI runtime defense is one to three years from early majority adoption because GenAI technology deployments and use cases are maturing, the attack surface and resource access are rapidly expanding, and threats and accidental exposure continue to emerge.

As organizations move beyond exploring and testing GenAI use cases to deploying GenAI in production, increased demand for AI runtime defense will follow. Moreover, as customer-facing AI agents are deployed in greater numbers, the need to prevent jailbreaking and other runtime concerns, such as biased output, will drive greater demand for in-line security solutions.

AI runtime defense is growing quickly because the need to reduce data exposure and other risks will become acute with increasing AI deployments, particularly for externally facing AI-enabled capabilities such as customer service agents. GenAI-enabled models, applications and use cases are expanding rapidly, and new agentic AI workloads are being actively explored. The myriad AI startups and established vendors are expected to go through a period of market education and trust-building that further generates demand and drives adoption.

Mass: High

The mass for AI runtime defense is high because of the breadth of AI use cases that require inspection in production, the number of business functions involved and the growing security challenges associated with agentic AI.

Generative and other forms of AI are being deployed across diverse industries in which the models and applications have access to, may be trained on, and can produce sensitive and confidential data. Ensuring the security of production workloads is paramount to preventing data breaches, intellectual property theft and compliance violations. The regulatory landscape is growing more stringent, with laws mandating strict data protection practices. As AI systems become more autonomous and integral to decision-making processes, ensuring their integrity, quality of output and security is crucial to maintaining trust and reliability in AI-driven operations.

Expected innovations in chain of thought, model reasoning, agentic AI and new data science approaches will generate new security concerns requiring automated oversight at runtime. Prompt injections and other attacks against LLMs and their applications will evolve as the foundation models are upgraded to new versions. Development of specialized detection algorithms may be required to shield against more advanced natural language hacking, model manipulations and other adversarial GenAI attacks.

Recommended Actions

- Establish realistic expectations of the threat landscape by clearly articulating the immediate and practical value AI runtime defense solutions provide in preventing real risks from accidental data oversharing by employees and customer abuse, rather than the yet-unrealized threat of malicious attackers.
- Future-proof adoption of AI runtime defense by designing solutions with a forward-looking, extensible architecture capable of defending diverse AI models (beyond just LLMs), handling multimodal inputs and outputs, and securing complex agentic AI workflows as the breadth of use cases expands.
- Focus on integrating AI runtime defense capabilities into existing, more mature security platforms (e.g., API security) in order to accelerate adoption and address the reliance on alternative controls.

Gartner Recommended Reading

Market Guide for AI Trust, Risk and Security Management

Emerging Tech Impact Radar: Cloud Security

AI Security Posture Management

Analysis By: Mark Wah

Definition:

AI security posture management (AI SPM) technology solutions help enterprises scan their infrastructure to discover deployed AI models, assistants and agents, and their associated data pipelines. AI SPM evaluates how enterprise AI infrastructure creates risks when exposed to data storage and processing workloads. This enhanced monitoring functionality is primarily achieved through API integrations with public cloud infrastructure providers' cloud AI developer services (CAIDS), AI engineering tools, and AI agent platforms such as Amazon Bedrock, Azure AI Foundry, Google Vertex AI, Databricks, Microsoft Copilot Studio and Salesforce Agentforce. It is the primary mechanism used to extract, monitor and normalize settings and activities over time for security risks such as misconfigurations and the lack of proper guardrails.

Sample Vendors

Cato Networks (Aim Security); BigID; Microsoft; Noma Security; Orca Security; Palo Alto Networks; Securiti; Strac; Wiz; Zenity

Range: 1 to 3 years

The range of AI SPM is one to three years due to accelerated AI adoption across multiple AI deployment options in the buy versus build spectrum.

Despite accelerated AI adoption and the escalating configuration risks driving the need for solutions, broader integration of AI SPM with diverse AI engineering tools remains inconsistent among providers. Initial adoption was focused on CAIDS, which was addressed by cloud-native application protection platform (CNAPP) technology providers. This expanded to AI engineering tools, data-centric use cases — particularly from data security posture management (DSPM) technology providers — and AI agent platform coverage. Data is the fuel for AI, and use cases are found in AI training, fine-tuning, data retrieval techniques such as retrieval-augmented generation (RAG), and AI agent data access. This fragmented landscape and the evolving nature of numerous AI services integrations and API support continue to contribute to the timeline for widespread adoption.

The growth of AI SPM is influenced by active venture capital (VC) investments in startups focusing on AI trust, risk and security management (AI TRiSM), and adjacent markets in cloud security and data security. The rapid evolution of AI and related technologies is escalating configuration risks, which in turn drives the need for a more comprehensive AI SPM approach. Agentic AI advancements and the broad spectrum of AI agent platforms have led to additional risks that some startups are addressing with solutions that are being labeled as AI SPM. The coverage will be fragmented along the AI deployment spectrum from embedded AI within software or SaaS applications to build or blend AI options. AI SPM coverage depends on integrations to gain visibility into AI assets and transactions, including their prompts and responses. The APIs and integrations to support security use cases are mixed, based on the maturity of the AI implementations.

Mass: High

The mass of AI SPM is high, given broad industry adoption from different product lines such as AI TRiSM, CNAPP and DSPM.

The impact of AI SPM is broad and driven by accelerating AI adoption across diverse industries, which is introducing new AI-native and amplified infrastructure risks. It secures AI training, fine-tuning, RAG and agent data access, preventing misconfigurations and vulnerabilities. Coverage is fragmented, with CNAPP addressing CAIDS, AI TRiSM used more broadly within AI engineering pipelines and runtime, and DSPM addressing AI SPM from a data access governance perspective and for RAG or AI agents. Adoption will expand across the banking, finance, insurance, government and manufacturing industries, as confirmed by Gartner search analytics.

AI services, AI engineering tools and AI agent platforms are now expected to offer secure configuration coverage. However, the default settings of such offerings often prioritize ease of adoption over security. CNAPP's AI SPM coverage for CAIDS has the longest tenure, while the broader coverage for various AI engineering tools and AI agent platforms remains fragmented. This fragmentation indicates that while AI SPM will significantly enhance security practices as it evolves and matures, adoption rates will be mixed across AI providers and solutions.

Recommended Actions

- Provide coverage for popular CAIDS, AI engineering tools and AI agent platforms by conducting product discovery research and surveys among ideal customer profiles to help prioritize AI SPM support and integration.
- Differentiate coverage by providing early support for popular AI engineering tools and AI agent platforms such as Databricks, DataRobot, LangChain and hyperscalers' no-code agent builder platforms.
- Address gaps and fragmented AI SPM coverage by partnering with providers in adjacent markets such as DSPM or CNAPP.

Gartner Recommended Reading

Hype Cycle for AI and Cybersecurity, 2025

Market Guide for AI Trust, Risk and Security Management

Market Guide for Cloud-Native Application Protection Platforms

AI Usage Control

Analysis By: Mark Wah, David Senf

Definition:

AI usage control (AI-UC) is a combination of technology approaches to discover the use of third-party AI and embedded AI and enforce the security policies of an organization. AI-UC discovers and categorizes third-party AI consumed as SaaS apps or AI embedded within applications. It defines and enforces usage policies, inspects content for sensitive data and responsible AI usage, assesses risk and alerts on anomalous activities, and manages AI configurations to reduce risk.

Sample Vendors

Aurascape; Cato Networks (Aim Security); Harmonic; Lasso Security; LayerX; SentinelOne (Prompt Security); Tenable (Apex); WitnessAI

Range: 1 to 3 years

The range of AI-UC is one to three years, as it addresses a prevalent need; third-party AI and embedded AI are found in more than 80% of existing software and services.

AI-UC is an emerging technology; currently, less than 10% of companies have it. However, adoption is predicted to grow rapidly as AI governance teams increasingly require comprehensive visibility and controls over AI usage, including third-party AI and AI embedded within SaaS or software applications. Embedded AI accounts for over 40% of enterprise AI use, and that figure will continue to grow as existing software providers embed AI capabilities. End-user organizations are using network-based controls to address this problem, but lack the needed visibility by leveraging network data alone. AI-UC is primarily delivered as a cloud-based service and may include multiple inspection points such as browser extensions, web proxy, network, DNS traffic analysis, email data and local agents to provide more granular visibility and control.

The growth of AI-UC is driven by the rapid adoption of AI embedded in third-party and SaaS-based services and the resulting local use of AI models, which often outpaces an organization's ability to monitor, govern, and control AI usage. Organizations implementing AI governance frameworks need AI-UC to fill blind spots in AI trust, risk and security management (TRiSM) tools lacking comprehensive discovery for embedded AI. Concerns about sensitive data leaking to AI applications from the datasets used for their underlying training models are on the rise. Some AI-UC vendors can help secure the configuration of these emerging services. Many AI-UC vendors are early-stage startups that align with AI TRiSM, agentic AI security and SaaS security, with healthy venture capital (VC) funding in this area.

Mass: Very High

The mass of AI-UC is very high given the growth in VC investments in this area and increasing use among early AI adopters in multiple industries.

The requirement for AI-UC crosses all industry verticals and organization sizes. Early adoption was observed in highly regulated industries such as finance, government, and healthcare. Many are using existing security controls to address this problem, such as security service edge (SSE) and enterprise browsers that may appear "good enough" to address some of the risks. AI-UC vendors are quick to address some of the gaps in AI discovery via multiple inspection points with network analysis and browser extensions. AI-UC is largely a siloed capability that has not been integrated into broader security ecosystems, though integrations with security operations center (SOC) tools or security platforms such as SSE are often requested. Supporting the SOC and established security teams will require mapping security data to established frameworks such as MITRE ATLAS and OWASP Top 10 for large language models (LLMs).

AI-UC is adjacent to a few established markets, such as SSE and SaaS security posture management. AI-UC startups have uncovered gaps in AI discovery and related security policies, many of which have been supported by VC investments in the last few years (see Emerging Tech: Top-Funded Early-Stage Startups in GenAI TRiSM). Further innovation will be needed to address agentic AI and emerging AI browser use cases. Some of these are emerging within the current layers of AI-UC deployment options. These include third-party risk management (TPRM) systems, AI discovery tools, endpoint agents, browser extensions, and network traffic analysis, all of which are different methods for detecting AI usage within an organization. While each method has unique strengths, they each also have limitations in their ability to identify and control AI, especially in cases of shadow AI.

Recommended Actions

- Enable your AI discovery services to assess current visibility gaps for embedded AI usage, including specific AI models and relevant metrics. Contrast the results with those from existing security products by leveraging multiple inspection points and service integrations to uncover third-party AI applications.
- Gain competitive advantage and expand sales channels by forging technology partnerships and integrations with security vendors to ensure seamless integration into the security ecosystem. This will help ensure you support established functions such as the SOC, SSE, and AI governance platforms.
- Improve outcomes of the end user's broader security portfolio by aligning additional security data to MITRE ATLAS and OWASP Top 10 for LLM to support security integrations, existing security processes and workflow.

Gartner Recommended Reading

Innovation Insight for AI Usage Control

Exposing and Managing Embedded AI: Tools for Transparency and Vendor Oversight

AI Security Platforms Are Core to Cybersecurity Revenue Growth Strategy

Cybersecurity AI Assistants

Analysis By: Alfredo Ramirez IV

Definition:

Cybersecurity AI assistants use generative AI to help security teams with everyday tasks. Their purpose is to augment human operators by automating routine tasks, synthesizing threat intelligence, and generating remediation suggestions serving as a gateway toward semiautonomous and autonomous cybersecurity tools. They pull in knowledge from existing security tools, generate content or code, and support investigations or responses. Most are built into existing products as helpful features, but some work as stand-alone interfaces and can even trigger actions through connected software agents.

Sample Vendors

Airrived; CrowdStrike (Charlotte AI); Darktrace; Exabeam (Nova); IBM (QRadar and Guardium); Microsoft; Palo Alto Networks (Cortex XSIAM); ReliaQuest; Vectra AI (Cognito); Wiz (AI-SPM)

Range

The range of cybersecurity AI assistants is one to three years, driven by the integration of generative AI (GenAI) capabilities that help address urgent industry demands for efficiency (generate content or code) amid growing cybersecurity threats and talent shortages (assist security teams in their daily tasks).

We estimate that cybersecurity AI assistants are in an emerging phase, with a moderate benefit rating and 5% to 20% penetration, with nearly 88% of security leaders piloting or planning use. The distance to crossing the chasm into early-majority adoption (more than 16%) reflects market factors such as CIO confidence in GenAI reliability, evolving regulatory clarity, smoother integration and stronger demonstration of measurable value.

Cybersecurity AI assistants are getting traction across verticals. They are mostly embedded as product-bound features rather than operating across multiple platforms for a full 360-degree view, though broader integration is on the horizon. Their adoption of use cases is fueled by the availability of generative AI, which enables teams to generate actionable guidance, synthesize threat intelligence, automate early incident response and deliver remediation suggestions.

By addressing cloud misconfigurations, supporting secure coding, custom/natural language query, and reducing repetitive workload, these assistants help close skills gaps and increase agility through automation.

Mass

The mass of cybersecurity AI assistants is medium because their strongest impact is currently limited to specific, high-value use cases like alert triage, incident response and threat intelligence summarization for industries with advanced security maturity. Broader transformation is moderated by varied industry readiness, need for human oversight and integration dependencies.

We determine the volume to be high due to the broad interest and impact of cybersecurity AI assistants across SOC, IAM, SIEM/MDR and SSE. Industries showing interest in "cybersecurity AI assistants" include government, banking, finance and insurance, services and education, followed by technology and telecom, and healthcare. This suggests a broad interest and potential for adoption across various sectors.

These assistants enhance operator accuracy; cut training time amid high turnover; and support secure coding, misconfiguration fixes, script generation, and risk/compliance analysis — driving multifold value for any sector managing software, cloud or complex IT security.

Cybersecurity AI assistants have the potential for a high impact on the business of IT by transforming how security operations centers, incident responders and compliance teams handle day-to-day workloads. In target industries with high security maturity, such as finance, telecom and defense, cybersecurity AI assistants can accelerate investigation, automate repetitive tasks and improve decision quality, freeing scarce talent for more strategic

work. Over time, they could reshape SOC staffing models, shorten response cycles and enable more proactive threat management.Op

However, realizing this potential depends on overcoming current barriers that affect the one- to three-year adoption range, including limited integration beyond single-vendor ecosystems, CIO concerns over GenAI reliability and explainability, evolving regulatory and compliance frameworks, and the need for robust safeguards around privacy and trust. Until these are addressed, adoption will skew toward tactical use cases, task automation, playbook execution and contextual insights rather than wholesale replacement of human expertise.

Recommended Actions

- Prioritize outcome-driven cybersecurity AI assistant initiatives by launching small, measurable experiments that build user trust through transparency and control. Develop teamwide AI literacy efforts to set realistic expectations and implement robust evaluation frameworks to assess agent reliability, explainability and alignment with security workflows.
- Adopt AI-powered cybersecurity workflows with measurable caution, integrating semiautomated assistants where outcomes are transparent and verifiable. Preserve and empower human expertise by embedding manual verification loops in key decision points while leveraging cybersecurity AI assistants.
- Build and develop industry- or workflow-specific, vertically aligned cybersecurity AI assistants for highly regulated industries and critical infrastructure where churn is high and tactical shifts are needed. Leveraging DSLMs enables context-aware threat detection, policy interpretation and compliance alignment, making AI assistants more actionable, trusted and relevant at the tactical level.

Gartner Recommended Reading

How to Evaluate Cybersecurity AI Assistants

Innovation Guide for AI Code Assistants

3 to 6 Years

AI Code Security Assistant

Analysis By: Mark Wah

Definition:

Artificial intelligence (AI) code security assistants (ACSAs) are technologies that help developers identify and remediate security vulnerabilities in code. They offer autoremediation suggestions and direct code assistance or chatbots, using forms of AI, such as generative AI (GenAI) and AI agents. ASCAs seamlessly integrate with static analysis application security testing (AST) tools to deliver just-in-time, data-driven guidance by embedding early in the software development life cycle (SDLC). They optimize multiple variables to address security vulnerabilities and code-quality issues simultaneously. This end-to-end integration is vital for evaluating and securing both AI-generated and human-written code, ensuring risks are identified and mitigated before they propagate.

Sample Vendors

Checkmarx; GitHub; GitLab; Harnesss (Qwiet AI); Mend.io; Semgrep; Snyk; Symbiotic Security; Synopsys; Veracode

Range

The range is three to six years because ACSA offerings are provided by established application security vendors and there is an influx of startups offering similar capabilities with AI agents.

Investments in generative AI (GenAI) by existing application security (AppSec) vendors are actively laying the groundwork for robust AI code security assistants. Many of these vendors are incorporating ACSAs as features within their existing solutions. This approach supports customers in transitioning to comprehensive platforms like application security posture management (ASPM), which consolidates various AppSec capabilities. The increasing speed of code generation by AI code editors, such as Cursor and Windsurf, introduces additional risks that can greatly benefit from ACSA capabilities. Vendors are strategically guiding this transition toward platform adoption by positioning ACSAs as one of the key features within their offerings. This approach is expected to result in consistent, moderate growth in market adoption over time. The ability to differentiate through fine-tuned large language models (LLMs) and domain-specific language models (DSLMs) can expand the focus of ACSAs beyond security vulnerabilities to also include code quality, open-source risks and consistency.

The adoption trajectory of ACSA is expected to closely align with that of general AI code assistants, which currently have a market penetration of 20%. AI code assistants are also projected to reach the Plateau of Productivity on the Hype Cycle within the next two to five years. As developers grow increasingly familiar with — and receptive to — these tools, the adoption of AI code assistants is expected to steadily accelerate the uptake of ACSAs at a similar, yet steady pace. A significant factor in the growth of ACSA is the investment by existing AppSec vendors using their extensive datasets on vulnerability remediation to train ACSAs to produce secure code. Emerging technology providers such as Qwiet AI and Semgrep are incorporating AI agents to deliver automated and proactive security capabilities.

Mass

The mass is high because it is being adopted across several industries to complement developer productivity and AI code assistant usage.

Various industries such as banking, finance, insurance, government and manufacturing are using ACSA, closely mirroring the adoption patterns of general AI code assistants. However, persistent concerns remain regarding privacy and confidentiality, similar to those associated with broader GenAI adoption. Organizations remain vigilant about the risks of data breaches and the potential misuse of sensitive information, including proprietary code. As a result, they mandate stringent safeguards to ensure ACSAs are implemented in a secure and compliant manner. As ACSA implementations mature, they are evolving beyond simple AI assistants into AI agent implementations, offering more autonomy and deeper integrations into the developer workflow.

ACSA has the potential to improve the security posture by significantly reducing the burden of AppSec testing. ACSAs help with:

- **Streamlining processes:** By generating secure code, ACSAs reduce the manual effort required to track and fix software vulnerabilities.
- **Addressing just-in-time training gaps:** They provide developers with real-time guidance on secure coding practices, helping to ensure that security is embedded into the code from the earliest stages of the development cycle.
- **Enhancing security remediation:** They provide secure recommendations when vulnerabilities in the code are detected.

By shifting security left in the modern, secure software development life cycle, ACSAs emphasize the principle that prevention is better than cure. However, the competitive advantage offered by ACSAs may diminish over time as general AI code assistants evolve to generate secure code, potentially narrowing the unique benefits ACSAs currently provide.

Recommended Actions

- Enhance integration with existing development tools so development teams can easily adopt ACSAs and integrate them into their workflow by creating seamless connections with popular integrated development environments, including AI code editors.
- Address privacy and confidentiality concerns by implementing secure deployment options that safeguard developers from leaking sensitive code to ACSAs for AI training.
- Differentiate ACSA security outcomes by fine-tuning LLM or developing DSLM with proprietary data that incorporates established secure coding practices, open-source risk metrics and code quality standards.

Gartner Recommended Reading

Hype Cycle for Application Security, 2025

Magic Quadrant for AI Code Assistants

AI Security Testing

Analysis By: David Senf, Mark Wah

Definition:

AI security testing uses offensive security methods to identify both the unique LLM and standard application security vulnerabilities of AI apps, models, library frameworks and notebooks. Prompt injection, multiturn attacks, model inversion and other test attacks are launched against AI models and applications. A primary goal is to manipulate model behavior to identify the potential for sensitive information disclosure and to determine under which conditions models will generate potentially harmful, unethical and inaccurate output. Testing also evaluates a model's resilience against denial-of-service attacks, model theft and supply chain vulnerabilities. AI security testing tools can be leveraged in automated scenarios and to support red teams generating probing prompts.

Sample Vendors

Cato Networks (Aim Security); Check Point (Lakera); F5 (CalypsoAI); HackerOne; HiddenLayer; Mindgard; NetSPI; Noma Security; Palo Alto Networks; PointGuard; TrojAI

Range: 3 to 6 Years

The range for AI security testing is three to six years for early majority adoption because, although the growth rate is high, it is starting from a low base of market adoption, and the tools are still evolving to higher degrees of effectiveness.

Gartner research shows that the number of generative AI deployments within an organization has an outsized impact on its view toward the prioritization of security and, by extension, testing solutions. Only those in the highest maturity category, with broadest deployments, place security at the top of their list of challenges. Thus, AI security testing will be constrained to organizations moving into production with a greater number of AI workloads rather than those that are experimenting and in early stages of deployment.

The growth rate for AI security testing is fast because of the significant potential reputational, regulatory and other damage that can occur from AI models, applications and agents. Organizations will require assurances that use of generative AI, particularly for externally facing automated chatbots and other AI-enabled services, will respond predictably and not create security and privacy exposures. Particularly for customer-facing services, organizations will need the comfort that their service output will remain within expected guardrails and not provide information that it shouldn't — either passively or if the customer is actively trying to extract discounts or sensitive data or cause other issues.

Mass: Medium

The mass for AI security testing is medium because demand for proactive application security and red teaming of AI solutions extends across most industry verticals and requires significant innovation to effectively test probabilistic models while providing prioritization and remediation guidance.

Like other forms of application testing and offensive security solutions, the appeal of these capabilities is broad across industries. The limiting factor is typically the size of the organization and the complexity of its application layer attack surfaces. AI security testing will have greater traction among larger, heavily regulated industries and those with more external, consumer-facing AI-enabled services.

Performing tests on AI applications and models requires further development and a better understanding of how to act on the results. Tool innovations will need to navigate the complex challenges as AI models evolve, new types of models emerge, and agentic AI spans multiple models and trust zones, noting that many organizations may test only small subsets of the total attack surface.

Large language models (LLMs) exhibit inherent variability in their outputs, making consistent testing of exploits difficult. This variability stems from the complex process of converting words to numerical embeddings and applying rule frameworks to generate and refine responses, meaning the same input can generate different results and higher false positives and false negatives. Consequently, internationally compromising these systems — especially multimodal AI that integrates audio, video, images and documents — requires understanding their dynamic interactions and designing consistent tests.

Recommended Actions

- Increase tool relevance by collaborating with other tool vendors to develop standard benchmarking and predictable expectations for appropriate prioritization and remediation actions based on testing.
- Reduce sales cycles by verticalizing offerings with contextualized tests, including tailored modules for specific industry verticals, to allow for deeper integration with industry-specific regulatory requirements, terminology and processes.
- Prioritize programmatic, automated adversarial testing capabilities for emerging multiagent environments, including developing tools that can generate diverse probing prompts, simulate a wide range of attack scenarios and efficiently assess harmful outputs across various AI models.

Gartner Recommended Reading

Market Guide for AI Trust, Risk and Security Management

Hype Cycle for AI in Finance, 2025

AI Supply Chain Security

Analysis By: Mark Wah

Definition:

AI supply chain security refers to the proactive and continuous protection of all components, processes and interactions involved in the lifecycle of AI systems, including design, development, deployment, operation and decommissioning. This security discipline encompasses the safeguarding of AI models as intellectual property, securing APIs for AI services, and ensuring the integrity of data used in training, fine-tuning, grounding and retrieval, and securing the infrastructure upon which they run. It also addresses the distinct security challenges of multiagent and multiparty systems in AI and AI-specific vulnerabilities, such as indirect adversarial attacks, in addition to established software supply chain security practices.

Sample Vendors

Apiiro; Cato Networks (Aim Security); Cisco; Google; Microsoft; Noma Security; Palo Alto Networks; Snyk

Range: 3 to 6 years

The range is three to six years because while AI adoption is rapidly accelerating, the specialized domain of AI supply chain security is emerging, indicating it is still maturing toward widespread, integrated enterprise adoption.

Organizations are leveraging specialized AI security solutions to address the unique vulnerabilities present in AI models, data and APIs. A significant challenge arises with the advent of agentic AI, which introduces a new frontier of security concerns due to its enhanced autonomy and interconnectedness, leading to an amplified attack surface and complex, unpredictable emergent behaviors. These concerns are particularly pronounced in multiagent and multiparty scenarios, which are part of the AI supply chain risk. The development of emerging standards, such as the Model Context Protocol (MCP) and Agent-to-Agent (A2A) Protocol, further complicates the security landscape by introducing new attack vectors that exploit how AI models process context and instructions. As AI development and adoption of its ecosystem including MCP servers becomes more widespread, organizations are increasingly recognizing the critical need for these security solutions to mitigate substantial financial, reputational and regulatory risks.

The fast-paced growth of AI adoption is driven by substantial investments in the AI sector, particularly in specialized AI security solutions. In 2024, global venture capital investment in AI companies surpassed $100 billion, with generative AI funding nearly doubling (see this Mintz article called The State of the Funding Market for AI Companies: A 2024 - 2025 Outlook). This rapid proliferation of AI applications is fueling demand for AI supply chain security, making it a strategic business and investment imperative. The velocity of this growth is reflected in the robust supply-side expansion, with dedicated venture capital investments actively backing AI security initiatives. These efforts focus on securing AI agents and ensuring comprehensive lifecycle security for data and AI, signaling strong growth in vendor and service organizations.

Mass: High

The mass is high because it transforms how organizations protect their digital assets and operational integrity in AI-driven environments, affecting numerous business units and industries.

This technology will have an impact across multiple business units and industries, reshaping operational paradigms in sectors such as banking, finance, insurance, government and technology. By enhancing real-time responsiveness and cybersecurity, it affects all components of the AI-enabled supply chain ecosystem, from AI models and APIs to data sources and agentic AI use cases. The integration of agentic AI, characterized by multiagent systems and multiparty coordination, will make it necessary to extend security measures across the entire AI life cycle. Risks such as instruction manipulation, including prompt injection and data poisoning, pose significant threats by potentially causing widespread and unpredictable harm across interconnected AI agents and systems. This can disrupt multiple business processes globally. The management of an AI Bill of Materials (AI-BOM) is critical across any department or industry adopting AI, providing essential transparency for all AI components and dependencies, thus ensuring a comprehensive understanding and management of AI risks worldwide.

AI supply chain security represents a shift for adopters and providers, transitioning from static, reactive measures to agile, autonomous and context-aware systems. While traditional cybersecurity models are evolving to address AI's unique risks, they remain incomplete without specialized AI-aware security solutions. The emergence of agentic AI introduces additional security challenges due to its increased autonomy and interconnectedness, leading to unpredictable emergent behaviors. Agentic AI transforms the security paradigm; vulnerabilities are no longer isolated, and a compromise can propagate exponentially, leading to widespread and unpredictable harm. The evolution from indirect adversarial AI attacks to tool manipulation by AI agents, which exploits how models process context and instructions, shifts the focus to controlling AI behavior rather than just securing data access. This requires a revolutionary approach to security models, with intent-based security and AI-aware security controls. The AI-BOM is a response to the need for transparency and integrity in AI artifacts, altering how providers build and users adopt AI by requiring deeper visibility into provenance and dependencies.

Recommended Actions

- Adopt and integrate AI security frameworks, such as the NIST AI Risk Management Framework (AI RMF) and the OWASP AI Security and Privacy Guide, into existing enterprise risk management and AI engineering processes to ensure a consistent and comprehensive security coverage including third-party and nth-party elements.
- Invest in specialized AI security solutions that offer capabilities like indirect adversarial attack detection and mitigation, AI security gateways, information governance and tools for monitoring agentic AI behavior, including emerging protocols such as MCP and A2A, to address gaps in existing security approaches.
- Prioritize robust data governance and privacy by implementing stringent controls for all data utilized in AI systems, including data minimization, comprehensive encryption, strict access controls and the adoption of privacy-preserving techniques, to mitigate legal and ethical risks throughout the AI lifecycle.

Gartner Recommended Reading

Market Guide for AI Trust, Risk and Security Management

How to Secure Custom-Built AI Agents

Cybersecurity Agent Builders

Analysis By: Alfredo Ramirez IV, Tushar Jain

Definition:

Cybersecurity agent builder platforms enable the creation and management of AI agents tailored for cybersecurity applications. These platforms let users design and modify agents using natural language, connect them to resources and other agents with low-code or no-code tools, and come with built-in cybersecurity tooling integrations.

Sample Vendors

AIrrived, Amazon Web Services (AWS), BlinkOps, Google, Lyzr, Microsoft, Relevance AI

Range

The range for cybersecurity agent builder platforms is three to six years based on the existence of a mix of purpose-built startups, general agent builder platforms already adopted for cybersecurity, and incumbent cybersecurity vendors extending into the space.

In the short term, signals indicate that this trend will either take hold as its own unique segment or may be absorbed as a set of cybersecurity functionalities serviced by the winning general agent builder platforms. The general agent builder platforms can be adapted to cybersecurity use cases but today that requires in-house cybersecurity expertise. The cybersecurity-focused platforms remove some of this need though not all.

Sample startup vendors for this trend have raised just under $90 million in 2025 alone. See The AI Workforce Revolution: $24M Series B to Accelerate Our Mission, BlinkOps Closes $50 Million Series B Funding Round Led by Eyal Ofer's O.G. Venture Partners and Lyzr's Funding and Investors.

Gartner has seen high inquiry interest in utilizing agentic AI for cybersecurity, indicating market support for new entrants. Meanwhile, incumbent vendors are planning, or have added, agentic AI build-and-deploy capabilities to their existing cybersecurity offerings.

Mass

The mass for cybersecurity agent builder platforms is medium because while the need for cybersecurity agents is growing, there is less investment in dedicated agent builder platforms than in agent-powered cybersecurity solutions.

The cybersecurity agent builder platform is one way to deploy agents for cybersecurity work, but time will tell if it is the market's preferred model. Currently, the difference between this trend and cybersecurity agent-powered solutions mainly impacts the strategies of cybersecurity providers and agentic AI vendors rather than the organizations adopting these technologies.

Cybersecurity agent builder platforms demand significant technical advancement over existing cybersecurity products. Incumbents must transform existing product capabilities to enable building and deploying agents within their platforms, and startups in the space are building platforms from the ground-up to take advantage of these new agentic design patterns.

Recommended Actions

- Product leaders for agent builder platforms should broaden their market and business reach by offering agent templates for cybersecurity tasks, such as policy enforcement, threat alert investigation and response, and security posture reviews.
- Product leaders for cybersecurity agent builder platforms should maintain their edge over general agent builders by showcasing specialized cybersecurity expertise through domain-specific models, building more and stronger integrations with leading cybersecurity tools, and incorporating reliable, deterministic features, e.g., ML-based triage and classification, to improve speed and consistency in resolving issues.

Gartner Recommended Reading

Innovation Insight: No-Code Agent Builders

Emerging Tech: Top-Funded Startups in Agentic AI

Emerging Tech: 'Time to Trust' Is the New Vital Agentic AI Metric

Information Governance

Analysis By: David Senf, Tarun Rohilla, Mark Wah

Definition:

Information governance is the management, protection and optimization of information throughout its life cycle to improve decision making while meeting regulatory and business requirements. Mature information governance controls are critical for securing sensitive data as AI deployments rapidly expand for both internal and external customer-facing use cases. Provisioning and monitoring access to only relevant and permissioned data for AI models, applications and agents is essential. Defining and enforcing information governance policies requires multiple technologies and spans various business functions, including compliance, legal, security and marketing.

Sample Vendors

AvePoint; BigID; Concentric AI; Cyera; Daxa; Fortanix; Microsoft; Rencore; RightData; Rubrik

Range

Information governance is three to six years from early majority adoption because executing on it involves multiple teams, technologies and processes. Also, the many and varied tools to deliver it remain fragmented, and organizational practices outside of heavily regulated industries tend to be less mature.

Although GenAI adoption has created a new data attack surface and exposed a lack of proper data controls, information governance practices are still ramping up. Moreover, the information governance technology layer is fragmented, with vendors that tend to focus on one or two of the required solutions. DSP, DSPM, DLP and DAG vendors focus on data security, while other vendors focus on permission management, and compliance vendors focus on retention and data life cycle. AI governance and runtime enforcement startups are tackling aspects of information governance as well.

The growth rate for information governance of AI workloads will be in significant double digits across the solutions that comprise this market. Organizations will initially follow the typical approach of deploying mitigating controls to secure data, such as web gateways and access controls in front of data exposure problems, instead of solving the lack of information governance. However, because of the impending scale of the AI and agentic AI challenge, practices and technologies to support improved information governance will need to grow significantly. But history has shown that information governance significantly lags new technology adoption.

Mass

The mass for information governance is high because the number of industries involved is expansive and the forthcoming innovations will be numerous.

All industries and all organization sizes have heightened requirements to improve information governance controls and practices. Momentum in this space will first come from larger organizations, those with stricter regulatory requirements and those that sell to end consumers. Adoption will expand downmarket over time as commodity and turnkey integrated solutions are offered. Organizations will accelerate the rollout of information governance given their acute internal GenAI data oversharing problems, especially those with externally facing consumer chatbots. Multiple buyers across compliance, data and analytics, and security will drive demand for these solutions.

For many organizations, weak and fragmented information governance is emerging as the major obstacle to wider GenAI rollouts. This presents an opportunity for vendors specializing in this space to consolidate functionality. New vendors and new approaches to integrating information governance silos are emerging due to the unique challenges of securing the information life cycle in a GenAI context. These solutions will only partially come by way of model hosting and model providers that do not typically have access to all enterprise data and cannot provide comprehensive information governance. This provides a more predictable market opportunity for vendors in this space, as their solutions won't be obviated by the foundation model providers.

Recommended Actions

- Integrate disparate information governance capabilities to offer cross-team solutions, and start with data discovery and classification to help organizations avoid the issue of unclear ownership of compliance and data security.
- Expand the market for information governance technologies by designing offerings that simplify deployment for downmarket adoption by providing automated discovery, classification and policy generation based on common compliance frameworks and industry requirements.
- Accelerate information governance activity around generative AI by focusing product messaging and initial capabilities on data oversharing, as it is a top generative AI adoption challenge.

Gartner Recommended Reading

Market Guide for AI Trust, Risk and Security Management

Intelligent Simulation for Security

Analysis By: Mark Wah

Definition:

Intelligent simulation solutions are designed to offer accurate modeling and what-if scenarios for both physical and digital systems at an unprecedented scale, and at lower cost, by leveraging digital technologies, such as AI, digital twins and spatial computing. This emerging technology is actively expanding beyond traditional applications, such as manufacturing, and is now revolutionizing critical sectors like security. It aims to transform security data and processes by accelerating proactive measures, shifting from reactive detection and response to preemptive cybersecurity.

Sample Vendors

Google; Microsoft; Reach; RedSeal; Skyhawk Security; Stream Security; Trend Micro; Tuskira; XM Cyber

Range: 3 to 6 Years

The range for intelligent simulation for security is three to six years because enabling technologies and advancements in digital twins augmented with AI capabilities, such as managed graph database services and AI model advancements, will lower the barriers to adoption.

Digital twin approaches have existed within cloud security for several years but have only seen adoption in the last few years. Key enabling technologies and advancements in AI, such as simulation twins and synthetic data, are significantly advancing capabilities and helping to reduce adoption friction. Early adopters of intelligent simulation are primarily found within cloud security, with emerging capabilities observed within exposure management. The widespread adoption of simulation data itself requires multiple enabling technologies, such as AI models, digital twins and data fabric, to reach early majority. Product leaders must also invest in driving business process change to trust and incorporate simulation data into decisions.

More security vendors are actively incorporating intelligent simulation capabilities, with recent announcements extending beyond just cloud security to include hybrid and operational technology (OT) environments. This indicates a broadening interest across different security domains. Furthermore, startups that are integrating these differentiated capabilities have received venture capital funding within the last two years, and established vendors are also adding these capabilities to differentiate their offerings in the market. The speed of adoption for using simulation to address security use cases is driven by its ability to mitigate risks, reduce testing costs, enable continuous risk assessment and testing cycles, and address the increasing complexity of enterprise infrastructure.

Mass: Medium

The mass for intelligent simulation for security is medium. Adoption is still in early stages within broad industries, with implementation mostly observed among younger technology companies and limited adoption by large enterprises.

Early adoption of intelligent simulation for security has been observed among younger technology companies. Initial implementations have largely focused on cloud security. However, security vendors are already extending the scope to hybrid environments, which is expected to help address some obstacles to broader adoption. This technology is revolutionizing critical sectors such as security, thereby enhancing exposure management and accelerating proactive measures. Intelligent simulation offers numerous use cases in cybersecurity, including exposure validation, threat detection, security control assessment, risk analysis, guided remediation, incident response planning, and security training and preparedness. It enables organizations to model and analyze complex environments, simulate potential attacks, and predict changes to their security posture via “what-now,” “what-if” and “what-next” scenarios.

Intelligent simulation for security provides a more comprehensive and continuous approach to exposure validation, which is poised to disrupt established products and services that offer periodic assessment and validation capabilities. This approach represents a revolutionary shift from traditional methods, which often rely on agents in live environments and periodic testing. Instead, intelligent simulation, often built on a robust digital twin, enables continuous and comprehensive validation in a controlled, safer environment, allowing for continuous assessment with low operational risk. This capability positions intelligent simulation to transform security operations by shifting from reactive detection and response to preemptive cybersecurity, creating new opportunities. Early adoption of intelligent simulation for security has been observed in highly regulated industries, such as healthcare, banking and insurance, particularly within cloud environments where mature cloud security practices are in place.

Recommended Actions

- Develop and differentiate simulation-based security capabilities that provide dynamic and continuous analysis within safe, controlled digital twin environments, embedding intelligent simulation concepts, such as digital twin, graph databases and AI, into security product portfolios.
- Address limitations of real-world data by integrating synthetic data generation and management tools into intelligent simulation offerings, enabling thorough and safe testing of diverse and extreme attack scenarios and continuous improvement of models.
- Articulate the value and differentiation of intelligent simulation as either a new product offering or integrated into existing security portfolio by educating the market on how its continuous and safer validation outperforms periodic testing in reliability and scope.

Gartner Recommended Reading

Emerging Tech: Intelligent Simulation Accelerates Proactive Exposure Management

Emerging Tech: Tech Innovators for Emerging Sectors in Intelligent Simulation

Emerging Tech Disruptors: Top 5 Early Disruptive Trends in Cybersecurity for 2025

Multimodal AI Protection

Analysis By: Mark Wah

Definition:

Multimodal AI protection is an integrated security approach that ensures AI systems are protected across all data types and modalities that they process. By moving beyond the constraints of single-modal analysis, this paradigm enables holistic, context-aware threat detection and protection through the processing and integration of various data types, including text, images, audio and video. It targets the emerging capabilities of AI systems that handle diverse data modalities, safeguarding against malicious attacks, privacy breaches and indirect risks. This capability enables the uncovering of sophisticated threats that are often invisible to traditional, unimodal security tools.

Sample Vendors

ActiveFence; Clarity; Check Point (Lakera); Coralogix; DeepKeep; Enkrypt AI; Mindgard; Reality Defender

Range: 3 to 6 years

The range of multimodal AI protection is three to six years due to the adoption of multimodal AI models by broad AI-enabled applications that will expose new security risks.

Multimodal AI is in its early adopter and deployment phase, with adoption focused on sectors where risks are significant and ROI is evident. The current focus of multimodal AI protection is on practical applications that yield measurable results, such as fraud and deepfake detection. By 2027, it is expected that 40% of generative AI (GenAI) solutions will be multimodal, making this capability a standard feature in enterprise software (see Emerging Tech: Top Emerging Use Cases in Generative AI). As multimodal AI becomes a core component of enterprise technology stacks, protecting against weaponized threats, such as malicious content in images and videos and the proliferation of deepfakes, will increase the importance of multimodal AI protection.

The growth of multimodal AI protection is driven by both demand and supply. Demand is propelled by the increasing sophistication of cyberthreats, such as deepfake-driven social engineering and phishing campaigns that integrate text, images, audio, video and other sensors to circumvent single-modal defenses. Early use cases are found in online fraud, identity and steganography-based attacks. On the supply side, the market is experiencing robust growth, with both startups and established companies advancing multimodal capabilities. The substantial investment of over $1.5 billion in AI trust, risk and security management (TRiSM) and disinformation security startups (see Emerging Tech: Techscape for Early-Stage Startups in GenAI TRiSM and Emerging Tech: Techscape for Startups in Disinformation Security) highlights the rapid technological advancement and the growing number of vendors supporting the expansion of multimodal AI protection solutions.

Mass: Medium

The mass of multimodal AI protection is medium as multimodal AI adoption itself is still growing and vendors' capabilities within are still limited to specific use cases.

Multimodal AI protection has a broad and expanding impact across several industries. Its core applications are critical in combating disinformation, online fraud and deepfakes, affecting sectors from finance and government to media and entertainment. Leading adopters are predominantly in high-risk and highly regulated industries, such as financial services, which are already moving GenAI use cases, including multimodal capabilities, into production to enhance fraud detection and compliance. Multimodal AI's capabilities also extend to other domains, like autonomous driving and healthcare, that will require security coverage.

Multimodal AI protection is highly revolutionary for both adopters and providers, representing an expanded role in cybersecurity that presents significant challenges and opportunities. The inherent nature of multimodal AI blends program code and the data it processes. Thus, maliciously crafted inputs — such as a subtly altered audio file, a carefully worded text prompt, or even a hidden element within an image — are not merely "bad data" but can effectively function as a form of code injection or instructions for AI systems to perform unintended and harmful actions. Established security tools are often blind to these new vulnerabilities, as the weakness is an emergent and often unpredictable property of the model's architecture and its training data, rather than a specific line of faulty code. Another key related area is disinformation security to combat deepfakes and impersonation that leverage multimodal AI techniques.

Recommended Actions

- Prioritize initial multimodal AI protection efforts on high-risk, regulated industries by focusing on pragmatic use cases like enhanced fraud detection and deepfake detection, starting with thorough integration testing of two to three critical modalities.
- Adopt a balanced approach for multimodal protection by conducting thorough multimodal risk assessments to identify blended attack vectors and implementing continuous monitoring of data integration points to customize security coverage.
- Invest in automated and dynamic security policies that explicitly cover text, image, audio, video and other sensors.

Gartner Recommended Reading

Emerging Tech: Top Emerging Use Cases in Generative AI

Emerging Tech: Top-Funded Early-Stage Startups in GenAI TRiSM

Emerging Tech: Top-Funded Startups in Disinformation Security

Security-Tuned DSLM

Analysis By: David Senf, Tarun Rohilla

Definition:

Security-tuned domain-specific language models (DSLMs) are AI models that have been trained or fine-tuned on cybersecurity-specific data, language, threats and workflows. These models are purpose-built to understand the semantics, structures and threat patterns within cybersecurity domains. Security-tuned DSLMs are reshaping how security teams analyze, respond to, and automate the protection of systems and data.

Sample Vendors

Cisco; Conifers; Darktrace; Drata; Dropzone AI; Google; Microsoft; Palo Alto Networks; Picus Numi AI; SentinelOne

Range

The range for security-tuned DSLMs is three to six years, driven by rapid model proliferation, growing integration into cybersecurity workflows and increased availability from both traditional AI providers and emerging pure-play generative AI (GenAI) startup vendors.

Security-tuned DSLMs are 20% to 60% on the way toward early adoption and thus at the mid stages from crossing the chasm, with early adoption strongest in SecOps, AppSec and GRC. Security-tuned DSLMs offer precise, high-fidelity automation aligned with operational needs, with expansion of AI GenAI use cases. Adoption is still early but accelerating, as mature organizations seek domain-specific AI to enhance detection, reduce alert fatigue, and support evolving guidance and regulations like NIST AI Risk Management Framework and the EU AI Act. Vendors such as Palo Alto Networks, SentinelOne and Drata are embedding DSLMs trained on telemetry, threat intel and analyst workflows to automate triage, secure coding and compliance

The speed at which GenAI capabilities are being added to domain-oriented solutions is rapid because the technical effort is relatively low for the first wave of GenAI augmentation. Early implementations are rapidly progressing from simple pattern recognition to more advanced, domain-specific reasoning and planning. The emergence of security-tuned DSLMs reflects this velocity; these models are being rapidly trained and deployed to address specific operational needs, often leveraging existing large language model (LLM) infrastructure and conventional AI techniques. Their ability to operate efficiently in resource-constrained or on-premises environments ensures improved latency, enhanced data privacy and minimized exposure risk. This momentum reflects a broader trend in which organizations are adopting a combination of conventional AI techniques and GenAI approaches to address complex, domain-specific security challenges.

Mass

The mass for security-tuned DSLMs is medium because of their ability to leverage vast amounts of data within a specific security domain to generate results that are impactful, relevant and accurate. While near-term gains may be incremental, these models are evolving rapidly as the technology matures and security use cases become more specialized.

Security-tuned DSLMs offer wide-reaching impact across sectors by aligning tightly with SOC workflows, automating log parsing, and linking evidence across SIEM, EDR and IAM systems. Compact models support real-time use on-premises, while larger models enable deeper threat analysis and hypothesis testing. Regulated industries such as healthcare, finance and manufacturing benefit from vertical-tuned DSLMs trained on HIPAA, PCI-DSS and OT telemetry. This improves precision, reduces false positives and accelerates threat response. Their adaptability across business units and geographies makes them a scalable solution for diverse security environments.

Domain-specific models are initially used to incrementally automate workflows in AI/ML applications. But as AI adoption grows, attackers are targeting the ML life cycle areas traditional tools miss. Security-tuned DSLMs fill this gap, understanding ML code, pipelines and artifacts to detect threats like data poisoning or model inversion. Embedded in DevSecOps tools, security-tuned DSLMs enable real-time inspection, AI BOM generation and vulnerability detection. As AI powers critical systems, securing the ML layer is essential. Security-tuned DSLMs support both protection and compliance with standards like NIST AI, ISO 42001 and the EU AI Act.

Recommended Actions

- Prioritize and map high-value AI use cases by analyzing domain-specific workflows, identifying areas where large reasoning models (for offline analysis) and compact DSLMs (for real-time enforcement) can enhance decisions in security, compliance and operations.
- Prioritize security-tuned DSLMs over LLMs by selecting models that are fine-tuned on threat telemetry, log data and workflows to reduce false positives and enable real-time, context-rich automation.
- Select vertically aligned DSLMs by matching models to operational domains and compliance needs by choosing security-tuned DSLMs trained on telemetry and threat patterns specific to domains like cloud, identity, OT or data protection.

Gartner Recommended Reading

Emerging Tech: Top Emerging Use Cases of Domain-Specific Language Models

Top 7 Forces Driving Domain-Specific Language Models

Sovereign AI

Analysis By: Mark Wah, Evan Zeng

Definition:

Sovereign AI is the effort by nation-states to invest in and progress their own development and use of AI to advance their unique sovereign objectives. A primary tenet of sovereign AI is a nation-state's desire to control its own development, modeling and use of AI systems and techniques, with less dependence on other countries' innovation and talent and less reliance on global vendors. The goal is twofold: global AI leadership and advancement of unique sovereign objectives.

Sample Vendors

China Mobile; Google; NVIDIA; Oracle

Range

The range for sovereign AI is three to six years because of regional requirements, including AI regulations and regional model providers.

Sovereign AI awareness is increasing due to the global race for AI leadership and related investments, national security concerns, evolving standards and governance, compliance frameworks, and diverse deployment options. Governments are shifting policies to promote domestic AI development and infrastructure, reducing external dependency. For instance, the U.S. AI Action Plan and related executive orders emphasize national sovereign AI acceleration. European regulations like the GDPR, the AI Act and the Digital Services Act, alongside national cloud initiatives, are driving specific countries' compliant GPU cloud services. Sovereign AI initiatives are underway in Canada, the European Commission, India, New Zealand, the United Kingdom, China and the United States (see How National AI Sovereignty Will Impact Your Enterprise for details).

Cloud adoption preferences vary, with some markets favoring cloud technology stacks, while others, like China and emerging Asia/Pacific (APAC), prefer on-premises or private/hybrid cloud for data and intellectual property control that aligns with country-specific sovereign AI initiatives.

Sovereign AI ambition is driven by national security, economic sovereignty, data sovereignty, localization and cultural context. An increasing number of countries are actively building their own AI infrastructure for competitiveness and future safeguarding. Data sovereignty, including data localization, is critical because data serves as the fuel for training and fine-tuning AI models, requiring sovereign control over every component of the AI stack — from computing capacity to storage, human resources and proprietary knowledge. Nation-states are now building national AI platforms to address biases and conflicts of interest as sovereign entities seek tighter control over LLM data governance and aim to reduce dependency on third-party AI providers in public decision making (see How Global AI Policy and Regulations Will Impact Your Enterprise for further details).

Mass

The mass of sovereign AI is medium because of broad regional awareness and emerging regional initiatives and policies.

Sovereign AI impacts nearly all aspects of government and its interacting enterprises, including sectors such as IT, banking, finance and insurance. The breadth of its impact is significant in European and APAC markets that show rapid growth in sovereign cloud adoption driven by government initiatives to establish regulatory frameworks and national AI capabilities. International markets are developing multilingual LLMs to capture local nuances and improve contextual understanding. Gartner search analytics also indicate that Asia/Pacific is the top region interested in sovereign AI. AI technology sovereignty is another reason for government-initiated sovereign AI, as it enables better control over the security of the AI technology supply chain, balancing the adoption of AI technology providers from multiple countries without heavily relying on open-source technologies.

Sovereign AI represents a revolutionary and transformative force for both adopters and providers, fundamentally reshaping how AI is developed, controlled and deployed. It will be challenging for providers to build and for users to adopt and deploy without a tailored approach, making it difficult to sell and scale. This arises from regional specifics, including preferences for on-premises deployment to achieve greater control over data and intellectual property. Varying AI maturity and a scarcity of AI specialists in some regions impede adoption across government functions. Addressing these obstacles is crucial and includes public-private collaborations. These partnerships accelerate data and analytics, AI upskilling and technology adoption. They also enable AI chip providers to leverage government subsidies, becoming key players in new sovereign AI value and supply chains. Engaging local providers and system integrators is vital to ensure alignment with specific requirements.

Recommended Actions

- Avoid a “one size fits all” approach, especially for global reach, by engaging in partnerships to address localization requirements.
- Learn from sovereign AI already underway by monitoring progress in jurisdictions such as China, India, Singapore, the European Commission and the United States.
- Differentiate from global tech service providers by deepening collaboration with regional or country-specific AI model providers and regional AI engineering tool vendors.

Gartner Recommended Reading

How Sovereign AI Impacts Enterprise Opportunity and Creates Nonobvious Threat

How Global AI Policy and Regulations Will Impact Your Enterprise

Synthetic Data

Analysis By: David Senf

Definition:

Synthetic data is artificially generated to fill the incompleteness, privacy concerns, bias and limitations of real data. It can easily achieve scale to benefit a long tail of requirements to provide extensive coverage for edge cases and future scenarios. This presents numerous benefits for cybersecurity. It is both an important input and a key output of intelligent simulation to analyze extensive "what if" scenarios for predictive threat intelligence, for example. Synthetic data has broad applicability across numerous security use cases, including privacy-preserving data sharing, attack path analysis, security control validation, AI models (red team testing) and antifraud measures.

Sample Vendors

BetterData; Electric Twin; Epistemix; Fairgen; MOSTLY AI; NVIDIA; Sarus Technologies; Simudyne

Range

Synthetic data is projected to reach early majority adoption within three to six years. While demand is expected to grow significantly, current adoption remains nascent and supply-side incorporation of synthetic data into solutions is still limited.

Synthetic data is currently in the early adoption phase. Despite its wide range of applications and its critical role in advancing the continued success of generative AI model training, both its market value and availability are still emerging. The core cybersecurity value propositions driving synthetic data adoption — including privacy preservation, improved predictive threat intelligence accuracy, red teaming and control validation — will require time to resonate. It will contribute to domain-specific model training for highly customized security solutions. Managed providers, vendors and end-user organizations will learn to generate fit-for-purpose data that reflects a target environment's assets and controls for tailored attack paths, misconfiguration and vulnerability discovery.

The growth rate for synthetic data is medium because its addressable market is broad, but it will need to be incorporated into security tools and workflows to prove effective. Overall, investments into synthetic data are growing rapidly, with venture capital investment increasing nearly 400% in the last three years (as noted in Emerging Tech Impact Radar: Intelligent Simulation). As these initial startups mature, undergo acquisitions and achieve greater market visibility, the market is expected to experience sustained growth and the emergence of new use cases. However, end-user adoption has lagged behind the robust forward-looking investments in the sector.

Mass

The mass for synthetic data is high because its value extends across all industries and it will spur important innovations in cybersecurity.

The value of synthetic data extends across all industries and organization sizes, with particular interest in larger, heavily regulated firms in finance, for example. Additionally, security service providers benefit from the capability to simulate a wide range of threat actor techniques, both at a general level and tailored to specific customer environments. As a key input for intelligent simulations, synthetic data provides rich context to provide forward-looking scenarios, rather than lagging indicators from existing telemetry data. Moreover, synthetic data serves as a crucial output of intelligent simulation, providing security operations and related tools with more accurate and actionable data.

Multiple cybersecurity technologies and services benefit from the addition of synthetic data. It helps address the limitations of real-world data by enabling data sharing without the associated privacy risks. Moreover, it enables the exploration of questions such as how threat actors evolve new attack techniques. Furthermore, synthetic data broadens the scope of AI and other data-driven solutions by enabling the training of domain-specific models for security operations and related tasks.

Recommended Actions

- Product leaders for security solutions and services should explore using synthetic data to train more robust and accurate domain-specific models for enhanced exposure validation, threat intelligence, anomaly detection and a myriad of other security use cases.
- Leverage synthetic data as a critical input for intelligent simulations to conduct advanced “what if” analysis for forward-looking insights beyond historical data. Moreover, incorporate synthetic data generation from intelligent simulations to provide tailored, fit-for-purpose datasets that reflect specific customer environments, attack paths and control effectiveness.
- Forge technology partner alliances actively to help educate customers and service providers on the tangible benefits of synthetic data, emphasizing its capacity to protect privacy, improve security tool accuracy and overcome data availability challenges.

Gartner Recommended Reading

Adopt Synthetic Data to Boost Your Innovation Ecosystems

Emerging Tech: Top-Funded Startups in Synthetic Data

Emerging Tech: Tech Innovators for Intelligent Simulation — Technology and Platform Innovation

6 to 8 Years

Guardian Agents

Analysis By: Alfredo Ramirez IV, Mark Wah

Definition:

Guardian agents are AI-based technologies designed to support trustworthy and secure interactions with AI. They function both as AI assistants, supporting users with tasks like content review, monitoring, and analysis, and as evolving, semiautonomous or fully autonomous agents, capable of formulating and executing action plans, as well as redirecting or blocking actions to align with predefined agent goals. Guardian agents leverage a broad spectrum of agentic AI capabilities and AI-based deterministic evaluations to oversee and manage the full range of agent capabilities, balancing runtime decision making with risk management.

Sample Vendors

Airrived; Amazon Web Services (AWS); Google; Lyzr; Microsoft; Relevance AI

Range: 6 to 8 years

The range for Guardian agents is six to eight years because their necessity becomes more pronounced as AI agents scale, and the vast majority of organizations are not ready for this level of deployment.

A number of things need to occur in the market before guardian agents can become a more common feature. Enterprises need to determine where agentic AI has an optimal use-case fit. Next, agentic platforms must be procured or built. Following these decisions, scaling agentic workflows must become prevalent. At a certain level of scale, it will become necessary to formalize an approach to guardian agents to oversee agentic activities at a speed and scale that humans cannot match. For example, these agents can function in two primary ways: As Sentinels, which review AI's integrity, risk posture, and policies, and as Operatives, which manage and control actions during execution.

The adoption of guardians will trail the adoption of AI agents in general. Some organizations will scale far quicker than others. This pace could potentially accelerate, depending on how the market embraces agents acting on behalf of customers, employees, and the organization at large.

Mass: Medium

The mass of guardian agents is medium because of the uncertainty of adoption across markets.

Agentic AI offers compelling functionalities to more than one industry and potentially to many business functions within a given industry. However, not all business functions and industries will necessarily require large-scale agentic operations. Enterprises that are not AI-ready, and that lack the necessary resources and AI literacy, will be slower in adopting agentic AI capabilities. Demand for guardian agents will be tied to the overall scale of agentic AI deployment, leading to adoption among the largest organizations in any given vertical and AI-native technology companies.

Guardian agents may well represent some advancement over other agentic capabilities, but they are unlikely to be fundamentally more technically advanced than the agents they oversee. Similar to a human team working with their manager, and individual contributors whose work is overseen by legal or compliance officers, teams of agents at scale will need to have their work overseen by guardian agents. The difference between the two lies in the goals set for each to accomplish, the expertise in which they are grounded, and the tools available to them.

Recommended Actions

- Product leaders involved in agent orchestration need to stay ahead of future competition by designing their systems and insisting on AI evaluations, which include a functional guardian agent design pattern.
- Capitalize on the future market for guardian agents by defining industry-specific agentic templates for your clients to use as part of their agentic workflows to align the guardian agent's goals with broad regulatory and internal AI governance requirements.

Gartner Recommended Reading

Guardians of the Future: How CIOs Can Leverage Guardian Agents for Trustworthy and Secure AI

Podcast: Guardian Agents — The AI Protecting Us From AI

About the Impact Radar

This Emerging Tech Impact Radar analyzes and illustrates two significant aspects of impact — when we expect a technology to have a significant impact on the market (namely, the range); and how big of an impact it has on relevant markets (specifically, mass). Each emerging technology or trend profile analysis is composed of these two aspects. See Note 1 for a complete description of our approach to this research.

In this document, profiles are organized by range and mass. Impact Radar range starts with the center and moves to the outer rings of the radar. The emerging technology's position on the impact radar represents when it will cross the chasm from early adopter to early majority. The rings represent one to three years, three to six years and six to eight years from crossing the chasm.

Mass is rated from very high to very low. The higher the mass score, the more broadly the emerging technology or trend is predicted to be adopted, and the more revolutionary the innovation is expected to be.

The objective of this research is to guide product leaders on how emerging technologies and trends are evolving and impacting areas of interest. Providers can leverage this knowledge to determine which technologies or trends are most important to the success of their business and when it makes sense to advance their products and services by investing in them. Technology vendors should use this Emerging Technologies and Trends Impact Radar to:

- Identify emerging technologies and trends that are important to the success of their business
- Determine when to act upon those trends and technologies based on business strategy
- Begin formulating a response to the technology or trend's evolution

Note 1:

The Emerging Tech Impact Radar content analyzes and illustrates two significant aspects of impact:

- When we expect it to have a significant impact on the market (specifically, range)
- How big an impact it will have on relevant markets (namely, mass)

Analysts evaluate range and mass independently and score them each on a 1 to 5 Likert-type scale:

- For range, this scoring determines in which radar ring the emerging technologies and trends will appear.
- For mass, the score determines the size of the radar point.

In the Emerging Tech Impact Radar, the range estimates the distance (in years) that the technology, technique or trend is from crossing over from early adopter status to early majority adoption. This indicates that the technology is prepared for and progressing toward mass adoption. So at its core, range is an estimation of the rate at which successful customer implementations will accelerate. That acceleration is scored on a five-point scale with one being very distant (beyond eight years) and five being very near (within a year). Each of the five scoring points corresponds to a ring of the Emerging Tech Impact Radar graphic (see Figure 1). Those emerging technologies and trends with a score of one (beyond eight years) do not qualify for inclusion on the radar. When formulating scores for range, Gartner analysts consider many factors, including:

- The volume of current successful implementations
- The rate of new successful implementations
- The number of implementations required to move from early adopter to early majority
- The growth of the vendor community
- The growth in venture investment

Mass in the Emerging Tech Impact Radar estimates how substantial an impact the technology or trend will have on existing products and markets. Mass is also scored on a five-point scale — with one being very low impact and five being very high impact. Emerging technologies and trends with a score of one are not included in the radar. When evaluating mass, Gartner analysts examine the breadth of impact across existing products (specifically, sectors affected) and the extent of the disruption to existing product capabilities. It should be noted that an emerging technology or trend may be expressed in different positions on different Emerging Tech Impact Radars. This occurs when the maturity of emerging technologies and trends varies based on the scope of radar coverage.

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Hype Cycle for AI and Cybersecurity, 2025

Emerging Tech: Intelligent Simulation Accelerates Proactive Exposure Management

Cool Vendors in Agentic AI TRiSM

Emerging Tech: Agentic AI Integration Will Separate TDIR Platform Winners and Losers

Table 1: Priority Matrix for AI Cybersecurity Ecosystem

Mass	*Range*			
↓	***Now (0 to 1 Year)*** ↓	***1 to 3 Years*** ↓	***3 to 6 Years*** ↓	***6 to 8 Years*** ↓
Very High		AI Governance AI Usage Control		
High		Agentic AI Ecosystem Security Agentic AI for Security AI Runtime Defense AI Security Posture Management	AI Code Security Assistant AI Supply Chain Security Information Governance Synthetic Data	
Medium		Cybersecurity AI Assistants	AI Security Testing Cybersecurity Agent Builders Intelligent Simulation for Security Multimodal AI Protection Security-Tuned DSLM Sovereign AI	Guardian Agents
Low				

Source: Gartner