Gartner

Hype Cycle for AI in ITSM, 2026

2 June 2026 - ID G00846344 - 86 min read

By: Chris Matchett, Chris Laske, Siddharth Shetty

Initiatives: Delivery of Functional Responsibilities

AI is reshaping IT service management, offering I&O leaders opportunities to boost efficiency and user experience. This Hype Cycle showcases innovations that AI can provide, but valuable outcomes depend on proper governance and ITSM foundational data.

Strategic Planning Assumptions

By 2027, 50% of AI projects at IT service desks will be abandoned due to unforeseen costs, risks or an inability to achieve the projected return on investment.

By 2027, generative AI (GenAI) will create more IT support and knowledge-based articles than humans will.

By 2027, IT service desk analysts will spend as much time interacting with AI as with business consumers.

In 2028, agentic ITSM actions will cause at least 2,000 incidents per medium organization due to the unreadiness of I&O teams and AI software immaturity.

By 2028, more than 30% of ITSM platform costs will be for AI capabilities, up from less than 15% in 2025.

By 2030, AI and automation will enable 80% of ITSM core workflows to run autonomously, reserving human action only for exceptions.

Analysis

What You Need to Know

This Hype Cycle for AI in ITSM highlights 23 innovative technologies, including ITSM implementations of machine learning and GenAI, although market attention is often drawn to the promise of autonomous, agentic automation. However, the most meaningful progress is being made through practical, incremental automations and AI applications that improve operational efficiency and business consumer experience.

In 2026, the effectiveness of AI initiatives is increasingly dependent on the quality and governance of underlying ITSM practices and underlying ITSM, business and technology data. As adoption broadens, the most successful heads of I&O are those who focus on data quality, standards adherence, invest in pilots with clear ROI, and remain pragmatic — embracing proven AI capabilities while maintaining a critical perspective on overhyped, immature innovations.

The Hype Cycle

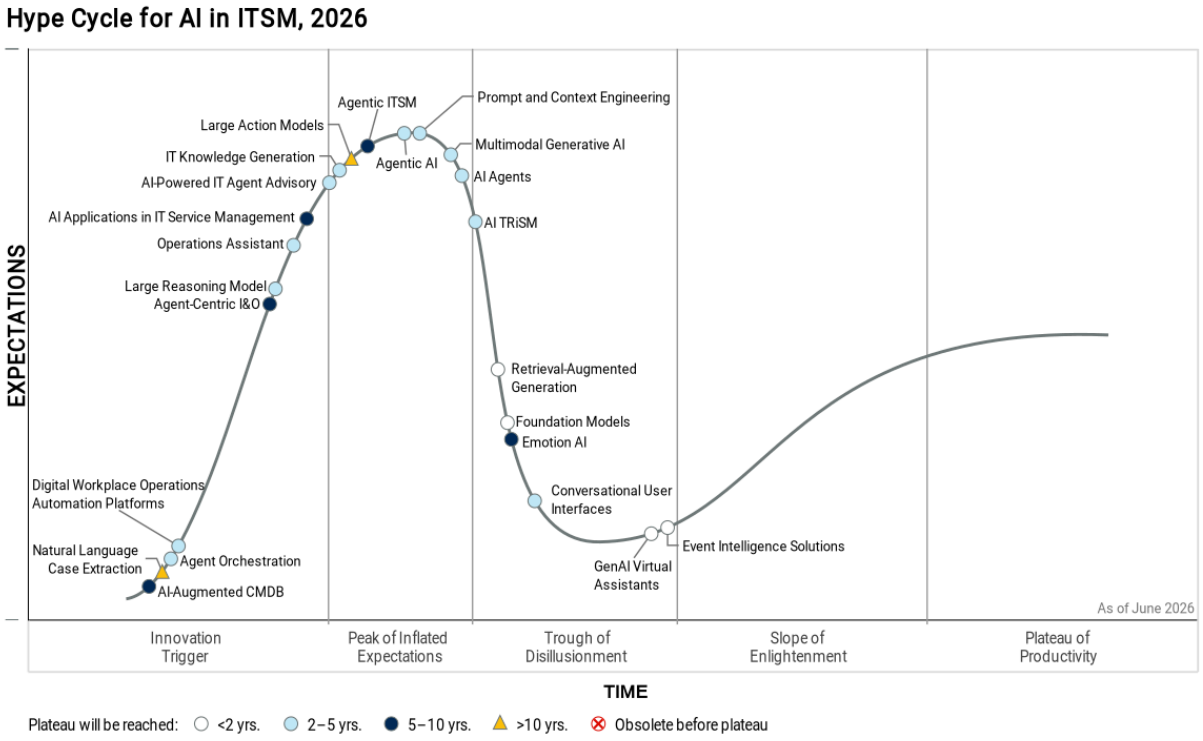
AI in ITSM continues to evolve, with heads of I&O facing a dynamic mix of opportunity and challenge. The widespread adoption of generative AI and foundation models is raising expectations for automation, knowledge management and user experience. The integration of large language models, retrieval-augmented generation (RAG), and GenAI virtual assistants into ITSM platforms and AI applications in ITSM is driving developments in case handling, knowledge generation, and support automation.

Even as generative AI is still maturing, the market is seeing growing interest in agentic ITSM and multiagent orchestration, though these remain largely aspirational, with practical deployment limited by organizational readiness and the immaturity of vendor offerings. Digital workplace operations automation platforms is a new market that offers AI and agent-based automation for IT support and digital employee experience (DEX) workflows, as an emerging alternative to AI applications in ITSM.

A defining trend for 2026 is the heightened emphasis on the quality and reliability of underlying data, especially configuration and knowledge base content. Organizations are recognizing that effective AI in ITSM depends on accurate, current and well-governed data. This is driving investment in AI-augmented CMDB, IT knowledge generation, and data curation tools, as well as renewed focus on integration between discovery, observability, and ITSM platforms.

Meanwhile, agentic ITSM and large action models are attracting significant market attention, but most organizations are waiting for proven, practical use cases before committing to large-scale adoption. As the market matures, heads of I&O are prioritizing solutions that deliver measurable ROI, while remaining cautious about vendor hype and the risks of immature automation.

Figure 1: Hype Cycle for AI in ITSM, 2026



The Priority Matrix

The Priority Matrix arranges each innovation profile by its level of benefit attainable versus its predicted time to plateau. The darker the color, the more immediate attention heads of I&O should pay to those innovations.

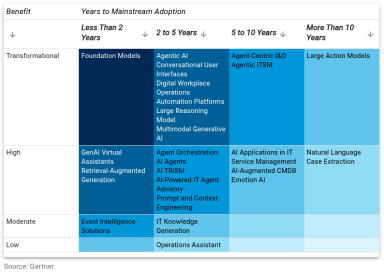
Over the next two years, heads of I&O should fast-track three high-impact technologies that promise near-term maturity by doing the following:

- Deploying AI-driven event intelligence correlation and noise reduction across monitoring and observability tools. Begin with pilots on high-volume alert streams, integrating outputs into existing incident management processes to reduce noise and accelerate triage.

- Deploying foundation models with prompt and context engineering to support a range of text, code, image or multimodal tasks. Focus pilots on rapid-value scenarios — such as automated summarization or template generation — while applying prompt and context engineering to align outputs with business requirements.
- Using RAG for proprietary AI search and IT knowledge workflows. RAG enhances generative AI by retrieving relevant information at query time. It assists in establishing index update processes and access controls to ensure accuracy and security.

Table 1: Priority Matrix for AI in ITSM, 2026

(Enlarged table in Appendix)



Benefit	Years to Mainstream Adoption			
	Less Than 2 Years	2 to 5 Years	5 to 10 Years	More Than 10 Years
Transformational	Foundation Models	Agentic AI Conversational User Interfaces Digital Workplace Operations Automation Platforms Large Reasoning Model Multimodal Generative AI	Agent-Centric I&O Agentic ITSM	Large Action Models
High	GenAI Virtual Assistants Retrieval-Augmented Generation	Agent Orchestration AI Agents AI TRISM AI-Powered IT Agent Advisory Prompt and Context Engineering	AI Applications in IT Service Management AI-Augmented CMDB Emotion AI	Natural Language Case Extraction
Moderate	Event Intelligence Solutions	IT Knowledge Generation		
Low		Operations Assistant		

Source: Gartner

Off the Hype Cycle

- **AI-focused problem management:** This profile is no longer tracked as a stand-alone innovation in 2026. Its core capabilities have been incorporated into AI-powered IT agent advisory, reflecting the market's shift toward integrated advisory and automation features for ITSM teams.
- **Virtual support agents:** VSAs have been removed from the Hype Cycle in 2026, as they are now considered a mature, mainstream capability. Virtual agents have become a standard feature in ITSM platforms and are no longer a source of market hype or differentiation.

On the Rise

AI-Augmented CMDB

Analysis By: Ankita Hundal

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Definition:

A configuration management database (CMDB) with AI uses machine learning, graph analytics and telemetry data to enhance service configuration data and infer component relationships. These capabilities improve dependency mapping, detect drift and maintain service topology with less manual effort. Although discovery and service mapping tools increasingly embed these capabilities, evidence of consistent improvements in CMDB accuracy and overall success rates remains limited.

Why This Is Important

Infrastructure changes faster than CMDBs can track, as cloud scale, ephemeral containers and continuous deployment drive ongoing configuration drift. AI-assisted CMDBs use telemetry and discovery data to keep relationships current and improve dependency insight. However, results remain inconsistent and depend on data quality management, integration maturity and operating model alignment.

Business Impact

Improved service configuration context enables better operational decision making. Faster understanding of dependencies during incidents and changes reduces the time teams spend diagnosing impact. AI-assisted CMDB capabilities can accelerate these outcomes in dynamic environments, but they complement rather than replace existing approaches. Over time, they can help align infrastructure, security and cost insights around a shared operational view.

Drivers

- Ongoing challenges with CMDB accuracy and ownership are driving interest in AI-assisted approaches that improve dependency mapping and reduce manual effort when traditional models cannot keep pace.

- Observability platforms are raising expectations by deriving real-time topology from telemetry, which pressures CMDB approaches to deliver more current, behavior-based dependency insight.
- The shift to infrastructure as code redefines how teams model and deploy environments, increasing demand for CMDB approaches that align with continuously changing states.
- Growing hybrid complexity across cloud services, containers, virtual machines and legacy systems is increasing the need for unified dependency visibility across platforms.
- The rising volume of signals from discovery, monitoring, cloud APIs and automation tools is driving demand to correlate fragmented data into a coherent view of system relationships.
- Expanded automation increases reliance on accurate dependency context to safely execute remediation, scaling and policy enforcement actions.
- Higher release frequency and continuous change increase the need to detect and manage configuration drift more effectively over time.
- Infrastructure scale continues to outpace team capacity, driving demand for approaches that reduce dependence on manual relationship tracking.
- Security and regulatory requirements are increasing the need for clear visibility into dependencies, drift and change history across environments.

Obstacles

- Vendors promote AI-enabled CMDB capabilities, but production adoption remains limited and results are inconsistent.
- Many organizations fail to define what “correct” means. Without this definition, neither CMDBs nor AI can reliably detect meaningful drift, and current AI capabilities do not close this gap.
- Fragmented ownership across teams and tools limits integration, governance and configuration data consistency.
- Poor data quality, duplication and insufficient context undermine trust in outputs and limit adoption.

- Black-box AI decisions face increased scrutiny in regulated environments that require explainability.
- Skills gaps and organizational resistance to change slow the adoption of AI-assisted approaches.
- Integrating telemetry, cloud data, infrastructure as code and observability into a usable model requires significant effort and cross-team coordination.

User Recommendations

- Address specific operational challenges first, such as incident investigation, change impact analysis and configuration drift detection, where dependency insights improve decision-making.
- Establish reliable discovery and service mapping before introducing AI, as trustworthy configuration data forms the foundation for meaningful AI analysis.
- Use AI-assisted CMDB capabilities selectively in dynamic environments, alongside observability and infrastructure as code as primary sources of truth.
- Integrate all primary sources of infrastructure information — discovery tools, cloud inventories and infrastructure-as-code repositories — into a unified configuration view.
- Evaluate vendors based on their ability to connect data sources and maintain accurate service relationships, rather than the sophistication of AI features alone.
- Track progress using practical indicators, such as faster incident analysis, clearer change impact visibility and reduced manual effort to reconcile configuration data.

Sample Vendors

BMC Helix; Dynatrace; Faddom; Freshworks (Device42); Ivanti; ServiceNow; Splunk

Gartner Recommended Reading

Drive CMDB Transformation Through Continuous Discovery and Service Mapping

How to Avoid the Top Mistakes With ITAM and CMDB Programs

Ignition Guide to Creating an Effective CMDB

Natural Language Case Extraction

Analysis By: Chris Matchett

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Natural language case extraction uses natural language technologies and LLMs to automatically recognize incidents and requests from organic conversations and create associated tickets in a case management system (e.g., an ITSM platform or a CRM customer engagement center).

Why This Is Important

Natural language case extraction (NLCE) is an essential capability for a functional collaborative support hub. It scans collaboration channels and support forums for conversations, data, knowledge and metrics for support activities that have not gone through a case management system and pulls them into the platform. NLCE agentically automates the slow, labor-intensive work previously done by IT service desk agents, who would manually search for incidents to add into the ITSM platform.

Business Impact

- Provides visibility of support effort and impact outside of formal contact centers.
- Provides advance warning of service issues that employees are discussing.
- Automatically creates incident tickets without requiring a service desk contact.
- Improves knowledge generation and maintenance by finding undocumented solutions used in peer support.
- Enables reporting of time spent by IT and business experts on the collaborative support hub and measures the impact of peer IT support.

Drivers

- NLCE is technically feasible using generative AI large language models (LLMs), which are well-suited to extracting context and data from unstructured content such as conversations.
- Vendors in AI applications for ITSM have plans for this capability within their roadmaps.
- NLCE appeals to employees and customers who prefer peer support and discussion channels.
- Organizations that implement a collaborative support hub for IT seek NLCE solutions to connect the hub to the case management system, such as their ITSM platforms.

Obstacles

- Software vendors don't have functional implementations of NLCE, despite the technical feasibility, resulting in its low availability.
- Organizations must build their own AI application with integrations to their case management system(s) to be able to use this functionality today.
- Employees may resist or mistrust AI features that monitor conversations.

User Recommendations

- Add a conversational agent to communal channels as a participant to offer assistance alongside formal support channels and human moderators.
- Enable virtual support agents to mark conversations as complete once successfully resolved in order to manage the clutter of common and repeat issues that have known solutions. This allows human expert responders and community moderators to focus on the other conversations that a conversational AI agent is unable to deal with.
- Use NLCE on conversations that are instantiated by swarming, as well as on more permanent collaborative support hub channels.
- Ask your case management platform vendor about these capabilities and how to activate them. The roadmaps are prioritized on customer demand.

Gartner Recommended Reading

Innovation Insight: Collaborative Support Hubs

How to Operate and Manage a Collaborative Support Hub

Critical Capabilities for AI Applications in IT Service Management

Agent Orchestration

Analysis By: Anushree Verma, Alastair Woolcock

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Definition:

Agent orchestration represents the next stage of agentic AI evolution as networks of diverse, specialized agents interact dynamically to solve multifaceted problems, adapt to environmental changes and continuously optimize their collective performance. These can be agents interacting with multiple applications to achieve a particular goal for the user.

Why This Is Important

The current AI agent implementations are mostly built with individual, task-specific agents built for specific tasks, which are very focused and provide incremental benefits. Therefore, this creates a value gap for enterprise-scale AI adoption. This is leading to a technical debt and a rise in security and risk issues due to agent sprawl. Therefore, without orchestration, AI agents will sprawl across the enterprise and become chaotic and unmanageable, limiting business impact.

Business Impact

Agentic orchestration represents the rise of a new control plane from record keeping to delivering business outcomes. It enables transparent economics of AI agents — tracking automation mix, SLA attainment, and cost per decision — and aligning monetization to proven outcomes. Furthermore, it can bridge gaps between disparate systems, such as ERP, CRM, HCM and more, allowing agents to access, analyze, and act on data across the enterprise securely and work on complex tasks at scale.

Drivers

- **Technological advancement:** Agentic orchestration represents the next stage of agentic AI evolution as networks of diverse, specialized agents interact dynamically to solve multifaceted problems, adapt to environmental changes and continuously optimize their collective performance. As these AI agents evolve, they will need a control plane to generate workflows, an execution plane and a governance engine to enforce SLAs and policies.
- **Vendor investment:** Many vendors have accelerated capabilities for developing agentic orchestration in the last three to six months, contributing to the acceleration of the trend.
- **Standards and protocols:** Development will accelerate agentic orchestration adoption and enable the evolution of AI agents to collaborate between multiple agentic systems. Currently, agentic AI platforms have enabled large language model (LLM)-based applications to access APIs and data sources, but integrations are often unique or proprietary. This fragmentation increases complexity of orchestrating agents and highlights the lack of industry standards, adding to user confusion. Standards and protocol development is still in its early days. Certain standards may evolve or become obsolete over time as well.
- **The unification promise:** Agent sprawl is creating unmanaged security and risk issues for the enterprise. Moreover, there is a value gap created by the current deployments due to the misalignment with the business outcomes. Agentic orchestration holds the promise of unifying these task-specific agents and provides a mechanism to align them to the business outcomes.

Obstacles

- Most products in the market are AI assistants, not AI agents. AI assistants do not typically engage in self-directed actions like AI agents do. Therefore, in such cases agentic orchestration would not be needed which limits market demand, unless there has been agent washing where an assistive implementation has been packaged as an AI agent.
- The core proposition of agentic orchestration should ideally go beyond one vendor's environment to maximize the transformative potential of such an offering. However, most of the offerings in the market limit it to a single development environment or limited to one's own development platform.
- There are fragmented pieces of agentic orchestration in the name of management plane, agentic fabric, control tower, trust center and governance, which needs to be brought together in a unified manner.

User Recommendations

- Do not overinvest in AI agents deployment for every small task. Start evaluating agentic orchestration platforms by your AI agent vendors to unify your implementation which will be more beneficial from a long term capability of your company.
- Look at ways to shift to agentic orchestration platforms that can verify outcomes end-to-end, with transparent evidence of performance, compliance, and cost control.
- Assess the capabilities of your vendors to determine whether they have orchestration offerings or their products can be integrated with other agentic orchestration — and to what extent — to determine their long term viability with these products.

Sample Vendors

Capgemini; GlobalLogic; IBM, Kore.ai; OneReach.ai, Pipefy; Ravical; SME UP; XMPro

Gartner Recommended Reading

Tech FutureSight: Agentic Orchestration Emerges As $550B AI Control Plane By Redefining Value Creation

Digital Workplace Operations Automation Platforms

Analysis By: Tom Cipolla

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

Definition:

Digital workplace operations automation (DWOA) platforms enable end-user services and digital workplace teams to automate, orchestrate and continuously optimize day to day operations by coordinating actions across multiple management systems. DWOA platforms are positioned above siloed endpoint management, IT service management (ITSM), digital employee experience (DEX), identity, asset, and security tools, using their data, integrations and capabilities to intelligently execute operational tasks.

Why This Is Important

The DWOA platforms market is an emerging category of tools that automate actions of, and utilize the data within, digital workplace (DW) management tools. DWOA platforms are positioned one level above the DW management tool layer and are capable of orchestrating complex workflows across many tools, which significantly reduces the cost and complexity associated with digital workplace operations as well as increases DEX. DWOA platforms simplify complex logistics within digital workplaces.

Business Impact

- **Reduced cost of operations:** Automation of common tasks reduces manual labor.
- **Operational consistency:** Repeatable automation enables consistent and predictable results.
- **Improved DEX:** Increased contextual data from multiple management systems powers accurate and intelligent automation aimed at increasing DEX.
- **Multisystem workflow orchestration:** Complex workflows such as onboarding and offboarding can be orchestrated across multiple systems (e.g., ITSM, endpoint management, identity and access management).

Drivers

- Vendors that currently operate at the utility layer (endpoint management, DEX management, desktop as a service (DaaS), managed device lifecycle services) are seeking opportunities to expand their customer base and renewal stickiness in the enterprise.
- The limited ability to innovate within the management tool silos has resulted in a higher degree of commoditization, which threatens vendor pricing and revenue strategies.
- AI capabilities are rapidly advancing, enabling domain-specific advancements.
- Digital workplace leaders now expect higher levels of automation to reduce costs and release resources from operations activities to transformational activities.
- The increased costs of endpoints due to memory shortages are driving a requirement to better utilize assets by assembling insights from data within ITSM, endpoint management and DEX platforms.
- Expectations associated with AI-driven headcount reduction have not been realized, generating demand for tools and platforms in new areas.
- Digital workplace leaders are simultaneously pressured to improve DEX, increase operational velocity and reduce the total cost of operations. This violates the triple constraint of quality, speed and cost. AI-powered platforms promise to break this constraint.
- Non-AI-based automation platforms currently exist, but many smaller organizations do not adopt them because of their complexity and requirement for advanced skills.
- IT leaders are searching for successful automation and AI use cases to demonstrate success.

Obstacles

- Unlike many IT services the digital workplace is closer to the human, as automation is exploited in employee facing services, the potential for error is magnified and the impact radius is increased. This is fundamentally different from other IT domains, where human sentiment and experience dynamics are not as prevalent, making digital workplace automation inherently riskier.
- Digital workplace leaders with low risk tolerance will not quickly adopt DWOA platforms out of fear of end-user disruption.
- Many agentic AI and workflow automation tools exist, but few are focused on the digital workplace. This results in a “build your own” model, which digital workplace teams will struggle to find success with due to limited time and skills.
- DWOA platforms must be operated with humans in the loop, progressively moving to humans on the loop, but never with humans out of the loop.
- Skills are not available and processes not designed for high levels of AI-enabled automation within the digital workplace.

User Recommendations

- Deploy a governance framework that will allow prioritization and monitoring of automations and their effectiveness, which becomes more critical when the platform leverages AI in autonomous and semiautonomous operations.
- Select a DWOA platform based on specific workflow orchestration capabilities for common digital workplace tasks, such as onboarding and offboarding, self-healing, software installation, DaaS provisioning and performance tuning, PC refresh, desktop incident triage.
- Dedicate digital workplace personnel to monitor and operate DWOA platform automations.
- Control risk by starting with shadow workflows — which mirror the logic of known good processes but only simulate the actions rather than directly taking action. Analyze simulation output to assess readiness for deployment.
- Gain leadership and organizational consensus on automation risk and secure executive buy-in on risk tradeoffs with speed and reduced cost of operations.

Sample Vendors

Aisera, Console, Efficient Ether, HP, Juriba, ReadyWorks, Rezolve.ai, Serval, ServiceNow

Gartner Recommended Reading

Innovation Insight for Digital Workplace Operations Automation Platforms

Agent-Centric I&O

Analysis By: Roger Williams

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Agent-centric infrastructure and operations (agent-centric I&O, previously referred to as agent-native I&O) is a pattern for organizing infrastructure and IT operations designed to exploit agentic AI in making autonomous and semiautonomous decisions to act toward organizational goals. Humans and machines work together to understand the current and future needs of key stakeholders and adapt systems and practices to meet those needs.

Why This Is Important

Agentic AI is poised to be the biggest disruptor of I&O since the introduction of the cloud. Agent-centric I&O provides the technical and organizational environments required for AI agents and multiagent systems to fulfill their potential. This includes seamless delivery of infrastructure platforms and services that minimize waste (such as overprovisioning) when responding to demand spikes, and observability of current versus expected states of I&O offerings to secure and improve system performance.

Business Impact

Agent-centric I&O enables AI agents to create the following competitive advantages:

- AI agents can self-serve infrastructure for experimental offerings within human-defined limits.
- People and AI agents can use simulated environments to explore disruptive possibilities.

- AI agents can administer systems and practices to create the resilience and adaptiveness required to meet business needs now and in the future.

Drivers

- AI agent capabilities have already started to drive significant differences in work adjacent to I&O such as software engineering, making them a topic of great interest.
- Heads of I&O are interested in AI agents to address chronic challenges with complexity, talent shortages and expensive cost structures, yet are unclear on what they need to do to best obtain those results.
- Agent-centric I&O provides a compelling vision for how I&O can play a critical role in business success, which addresses perennial I&O concerns about being viewed as a strategic value contributor.
- Agentic AI poses a dilemma for I&O leaders regarding how much of their effort should be devoted to preparing for agentic AI versus addressing current needs. Agent-centric I&O goes beyond just the introduction of AI agents to address this dilemma by providing a comprehensive vision that builds future readiness for agentic AI adoption while addressing current challenges.
- Agentic AI needs trustworthy data about the entire infrastructure estate and the IT operating model to make accurate, timely decisions with minimal costs and risks. The intelligent simulation element of agent-centric I&O addresses this by giving AI agents needed feedback on actions in complex environments to support them at the speed required for effective experimentation.
- I&O staff are at risk of not developing vital infrastructure expertise as AI agents take on easier tasks, which agent-centric I&O accommodates through its use of intelligent simulation.

Obstacles

- **Inadequate governance of I&O:** AI agents need explicit guidance about goals, boundaries and rules that are often not documented, and organizational staff need ways to measure and assess the performance of AI agents to improve results.
- **Staff readiness for AI agents:** Agent-centric I&O requires people to have the motivation, expertise and time to understand how best to work with AI agents.
- **Lack of agentic AI standardization:** Differences in vendor offerings will require trade-offs in how AI agents are matched to work that needs to be done.
- **Lack of I&O effectiveness in managing infrastructure platforms:** The ability for I&O to effectively manage infrastructure platforms is mandatory for obtaining full value from the use of AI agents. Gaps frequently exist in the infrastructure estate, operating model and documented knowledge about the organization that limits I&O effectiveness in managing infrastructure platforms and, in turn, their ability to adopt agent-centric I&O.

User Recommendations

- Create a vision for how agentic AI would change I&O by identifying where I&O is constrained due to limited resources for decision making and taking action, and imagining what would be different if these constraints were broken by agentic AI.
- Provide clarity about the mission and the constraints that AI agents must adhere to in order to fulfill it.
- Use Gartner's IT Score for Infrastructure and Operations to understand and close the most significant gaps.
- Require new infrastructure investments to be based on intelligent infrastructure and platforms that AI agents can then use to perform tasks and augment their capabilities.
- Address quality gaps in technology asset data required for effective agentic AI use by strengthening ITAM, service management, observability and digital employee experience practices.
- Sponsor staff development vital skills for working with AI agents and instill confidence that they will still be treated as valuable members of the organization as AI agent use increases.

Gartner Recommended Reading

Blueprint for Platform-Centric I&O

The Future of I&O 2030: The Impact of AI

The Future of I&O Automation Is Agentic So Begin Piloting Now

Large Reasoning Model

Analysis By: Tong Zhang, Leinar Ramos

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

A large reasoning model (LRM) is an advanced type of large language or multimodal model designed to address complex problems that necessitate deliberate, multistep logical processes. It boosts problem-solving capabilities by eliciting reasoning chains, utilizing reinforcement learning, conducting extensive multistep attempts and verifying its own logic during inference before delivering a final output.

Why This Is Important

Standard LLMs provide instant response. However, they lack the capability to solve complex problems by generating intermediate steps, exploring potential solutions, verifying the correctness and backtracking from dead ends before coming to the solution. LRMs are designed to spend more effort during inference tackling complex tasks that require multiple logical pivots beyond simple knowledge extraction.

Business Impact

- LRMs incorporate structured reasoning, iterative reflection and verification mechanisms to perform business tasks that require logical reasoning and problem-solving skills.
- LRMs refine the reliability by trading compute time for accuracy. Through automated backtracking, they significantly reduce hallucinations in complex workflows.
- LRMs are considered the core AI model of various AI agents, enhancing their planning capabilities for better agentic solutions with minimal human intervention.

Drivers

- Diminishing returns from scaling up pretraining in standard LLMs highlight the need for a paradigm shift toward LRMs. Shifting computing resources from training phases to inference phases allows models to spend more time refining their responses. Techniques such as generating chains of thought, revising answers, employing external verifiers, backtracking and sampling multiple times to select the best answer significantly improve model performance.
- Increased interest in AI agents drives the need for LRMs, which are widely expected to become a core AI agent component, to improve their ability to tackle problems involving complex business processes.
- Use of model distillation drives LRM adoption. Training distilled smaller models from LRMs achieves performance that surpasses that of small language models while requiring fewer computing resources than the original LRMs. This lowers inference and training costs, making LRMs more accessible to organizations.
- LRMs produce explicit, multistep chains of thought, providing a built-in audit trail for compliance regulation use cases.
- Large action models (LAMs) are foundation models trained and optimized to identify and generate an action or set of actions that can be used to impact a target environment to meet a goal. LAMs can augment reasoning models by improving their action-oriented capabilities. This integration allows LRMs to not only process information, but also to execute tasks, bridging the gap between understanding and action.

Obstacles

- LRMs have higher cost and latency, limiting their adoption in many enterprise use cases.
- Little evidence suggests LRMs possess a genuine understanding of concepts, meaning they may fail on edge cases where historical data patterns no longer apply.
- Despite LRMs showing an improved chain of thought over LLMs, their black box nature challenges businesses needing reliable, explainable solutions.
- Enterprises find effective evaluation of LRMs is challenging, which hampers adoption, and current benchmarks face limitations such as data contamination and narrow scope.
- LRMs may enter repetitive loops, wasting tokens and time without guaranteeing superior results to LLMs.
- Doubts remain about applying LRM reasoning beyond math and science.
- LRMs may behave inappropriately if they find shortcuts to satisfy their reward functions.
- LRMs may use their powerful logic to build an internally consistent, but completely false, reasoning, making errors harder to spot.

User Recommendations

- Pilot LRMs in areas where complex capabilities for decision making, strategic planning or problem solving are required and LLMs failed to achieve the desired outcome.
- Provide employees foundational training about LRMs and appropriate use cases.
- Experiment with LRMs as the core agent orchestrator for AI agent solutions, replacing traditional LLMs to check whether performance can be improved.
- Create nuanced evaluation methods for LRMs, including quality of outcome, token consumption and inferencing latency.
- Acknowledge the inherent black box nature of LRMs and mandate appropriate levels of validation, testing and human oversight to ensure deep interpretability.
- Implement strict ROI criteria for LRM utilization. Queries should be directed to standard models by default, reserving LRM only for complex use cases.

Sample Vendors

Alibaba Cloud; Anthropic; Baidu; DeepSeek; Google; MoonShot AI; OpenAI; xAI; Z.ai

Operations Assistant

Analysis By: Chris Matchett, Matt Crossley

Benefit Rating: Low

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Operations assistants are conversational interfaces that provide human IT agents with data-driven insights and also execute ITSM actions to help them carry out their role. They provide access to AI and GenAI capabilities, such as case summarization and automatic communications.

Why This Is Important

While traditional virtual support agents have focused primarily on helping business consumers, there is a growing demand for solutions tailored specifically to the needs of human IT agents. This has driven interest in operations assistant solutions, which offer user experiences similar to other AI assistants but are tailored toward helping I&O workers increase their productivity.

Business Impact

- Optimized and improved IT support processes through insight and automation
- Faster resolutions and improved accuracy in triage, categorization and expert identification
- Easy access to deeper insight by analyzing processes, tickets and workflows to identify opportunities for improvement
- Improved experience for human IT agents, leading to fewer frustrations and improved morale

Drivers

- While business-consumer-facing virtual support agents have previously been the main focus, heads of I&O are now prioritizing AI assistant solutions that deliver similar user experiences for human IT agents, aiming to enhance their productivity.
- Some organizations have purchased or are piloting solutions such as Microsoft Copilot or Anthropic's Claude Cowork and hope to use similar technology within I&O.
- Increasing numbers of vendors in IT operations management (ITOM) and AI applications in ITSM markets have branded part of their offering as AI assistants.
- Some heads of I&O believe that conversational interfaces will become a primary method of agent interaction with IT operations and service management software.
- Heads of I&O believe that AI-enabled assistant technology can help them meet the challenges of macroeconomic conditions to find cost savings through increases in productivity.

Obstacles

- General-purpose conversational AI assistants primarily designed for meeting transcription and summarization use cases are not suited to be an operations assistant to facilitate I&O workflows.
- Operations assistants are too new to be considered as a must-have. Consequently, the high cost of current commercial solutions is hard to justify for experimental use.
- Human IT agents may not want a simplified conversational interface to carry out their tasks, instead preferring fully featured forms or command line user interface (UI).
- IT agent advisory capabilities do not require an operations assistant; instead, they more commonly show up as other UIs, such as the classic web form UI of ITSM platforms.
- The term "copilot" has seen reduced use by vendors from 2025 to 2026, as it has become more associated with Microsoft's offering than a technology category.
- Operations assistants rely on other underpinning AI capabilities, such as agent advisory.

User Recommendations

- Seek operational assistants that are designed for ITOM and ITSM use cases because general-purpose AI assistants would need extensive customization.
- Confirm that existing IT agents can benefit from conversational interfaces rather than preferring traditional user and command line interfaces to interact with tools.
- Do not rely on operations assistants to replace domain specialists. Ensure that novice or new target user groups have sufficient domain knowledge to identify erroneous generative output.
- Since stand-alone operations assistant products have not emerged into the mainstream and do not contribute to a drop in overhead, heads of I&O should wait before committing to emergent products.

Sample Vendors

Automation Anywhere; BigPanda; Freshworks; OpenText; PagerDuty, ServiceNow

Gartner Recommended Reading

Magic Quadrant for Artificial Intelligence Applications in IT Service Management

Critical Capabilities for Artificial Intelligence Applications in IT Service Management

AI Use-Case Assessment for IT Service Desk

AI Applications in IT Service Management

Analysis By: Chris Matchett

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

AI applications in IT service management are tools that augment and enhance IT service management (ITSM) workflows using AI. These analyze ITSM data and metadata to provide intelligent advice and execute actions on ITSM practices and workflows, such as IT service desk and support activities. The software can be delivered as an incumbent ITSM platform with integrated AI features, a native AI add-on, or a third-party product that integrates with an ITSM platform.

Why This Is Important

Heads of I&O seek the benefits of AI, but they are not replacing ITSM platforms to obtain AI capabilities. Instead, they choose between incumbent ITSM platform capabilities and third-party, specialist AI solutions. The market is continually evolving, driven by advances in AI and natural language technologies. AI applications in ITSM are distinguished by their reliance on ITSM data and metadata as the primary input, ensuring alignment with ITSM practices.

Business Impact

- Faster resolution and improved accuracy in triage, categorization and expert identification.
- Tangible cost reductions and labor savings through automation of support issues and requests.
- Improved employee-facing user experience and stronger business consumer relationships.
- Deeper insight into service, practice and staff performance.
- AI advisory, automation and agentic features to help avoid disruptions, reduce downtime and improve IT service reliability.

Drivers

- The popularity of GenAI is prompting heads of I&O to have higher expectations of conversational AI platforms and look beyond previous-generation chatbot features in ITSM platforms.
- Buyer desire for agentic automation and “zero-touch” IT support has accelerated, and vendors are promoting their AI applications in ITSM products that offer agentic ITSM features.
- Buyer requirements have expanded beyond virtual support agents (VSAs) for self-service to include IT agent advisory, with a dual focus on improving both employee experience and ITSM team efficiency.
- ITSM platforms provide a variety of AI features, but heads of I&O often purchase add-ons or integrate stand-alone products from third-party vendors when they require more extensive capabilities.
- GenAI capabilities are increasingly sought by heads of I&O to automate content generation and improve communications. Examples include summarizing information, such as knowledge base articles or case work-log updates, and generating major incident notifications.
- This market saw several significant acquisitions in late 2025, which reflect a trend of vendors seeking to acquire missing capabilities and strengthen their ITSM automation and workflow orchestration solutions.

Obstacles

- AI applications in ITSM products have room for significant improvement, notably in machine learning analytics and agentic ITSM.
- I&O groups lack skills in areas such as prompt engineering to accelerate deployments using LLMs.
- I&O organizations lack sufficient foundational data to train their AI capabilities — whether consistently categorized incidents or good-quality knowledge for their virtual support agent.
- As heads of I&O must choose between incumbent ITSM platforms and add-on products specialized in AI, several smaller vendors have been acquired and merged into larger offerings in this market, reducing choice for buyers.
- Most AI applications in ITSM have yet to deliver on agentic ITSM capabilities beyond marketing hype.
- Costs associated with acquiring and implementing AI applications in ITSM are generally considered high and unpredictable.

User Recommendations

- Use the AI Use-Case Assessment for IT Service Desk to determine which AI use cases will provide the most transformational value, while being sufficiently feasible to implement quickly.
- Invest in AI applications for ITSM to achieve significant enhancements across a broad range of ITSM practices.
- Consider general-purpose conversational AI solutions for commodity features, such as using a virtual support agent to deflect calls from the service desk.
- Verify whether the incumbent ITSM platform can already meet these needs. If an add-on or upgrade is required, then assess the licensing and implementation costs before comparing with third-party solutions.

Sample Vendors

Atlassian; Automation Anywhere; BMC Helix; ServiceNow

Gartner Recommended Reading

Market Overview for AI Applications in IT Service Management

Magic Quadrant for AI Applications in IT Service Management

Critical Capabilities for AI Applications in IT Service Management

AI Use-Case Assessment for IT Service Desk

Infographic: Choose an AI Solution for ITSM

At the Peak

AI-Powered IT Agent Advisory

Analysis By: Chris Matchett, Siddharth Shetty

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Definition:

AI-powered IT agent advisory is the use of AI to analyze ITSM data and metadata and generate recommendations that accelerate and enhance IT agent activities. Example features include intelligent categorization, triage, routing, escalation, swarming, risk advisory and sentiment analysis.

Why This Is Important

IT agent advisory features are a core capability of AI applications for ITSM that differentiate them from business consumer-facing conversational assistants such as virtual support agents. This expands capabilities to meet the needs of IT agents through advice and actions embedded in the UI of an ITSM platform or via an operations assistant conversational interface.

Business Impact

- Enhances decision making, helping with IT support tasks
- Reduces costs through labor savings and faster resolutions
- Improves accuracy in triage, classification, swarming and routing
- Provides change-risk advisory to avoid disruptions and ensure reliability
- Sentiment analysis improves service experiences
- Frees up human IT agents' time from routine tasks
- Optimizes ITSM practices through actionable insights

Drivers

- Heads of I&O believe that AI can help them find cost savings through automation, leading to interest in AI for IT agent use cases and capabilities such as agent advisory.
- AI can be used to improve the operational efficiency of I&O teams using agent advisory features from ITOM vendors.
- AI technologies, such as GenAI and LLMs that enable AI-powered IT agent advisory features, are continually evolving.

Obstacles

- Products on the market have significant room for improvement, particularly in areas such as intelligent escalation, which is part of agent advisory (see Critical Capabilities for AI Applications in IT Service Management).
- I&O organizations often lack sufficient foundational data (e.g., consistently categorized incidents or knowledge) to train their AI capabilities.
- AI skills such as prompt engineering are needed to maximize the benefits of agent advisory, but are not yet sufficiently developed within I&O.
- AI-powered IT agent advisory often requires an upgrade or additional software purchase that heads of I&O can struggle to justify without evidence of proven ROI.
- Many AI applications in ITSM rely on LLMs for AI but lack machine learning techniques like cluster analysis and pattern recognition, which are needed for the full range of IT agent advisory features.

User Recommendations

- Use AI Use-Case Assessment for IT Service Desk to determine which AI use cases will provide the most transformational value and are feasible for quick implementation.
- Invest in AI applications for ITSM to gain AI-powered IT agent advisory features. Pair this capability with an operations assistant to use a conversational interface to leverage the advisory.
- Verify whether your incumbent ITSM platform already meets these needs. If an add-on or upgrade is required, assess the licensing and implementation costs before comparing them with third-party solutions.

- Optimize existing processes and remove bottlenecks before using AI-powered IT agent advisory to optimize ITSM practices, as AI will not fix a fundamentally broken process.
- Treat IT agent advisory as suggestions and maintain a human-in-the-loop verification of the results to ensure quality and reliability.

Sample Vendors

Automation Anywhere; BMC Helix; ServiceNow

Gartner Recommended Reading

AI Use-Case Assessment for IT Service Desk

Magic Quadrant for AI Applications in IT Service Management

Critical Capabilities for AI Applications in IT Service Management

How to Achieve Success With AI at the Service Desk

IT Knowledge Generation

Analysis By: Rich Doheny

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

IT knowledge generation creates knowledge articles with GenAI by summarizing attempted and successful solutions from sources, formatted in accordance with enterprise style guidelines and templates. These sources can include organic conversations, documents, and historical cases. These articles can then be leveraged in a management system (e.g., ITSM platform or conversational AI tool) for both IT and end-user consumption.

Why This Is Important

An effective knowledge base enhances IT self-service adoption and IT support staff efficiency, saving I&O time and money. Most development and ongoing maintenance of the knowledge assets is manual, relying on time-intensive human effort after an issue is resolved. Therefore, many organizations struggle to deliver an effective and sustainable knowledge management (KM) practice. By leveraging GenAI techniques, heads of I&O can automate the creation and maintenance of knowledge assets.

Business Impact

An effective KM system improves IT support quality and self-service. Heads of I&O can adopt GenAI for knowledge creation to:

- Accelerate the authoring and publication process
- Identify IT knowledge gaps
- Increase the number of useful knowledge assets
- Deliver knowledge-article content and format standards
- Reduce IT support overhead

Drivers

- Heads of I&O have been looking to automate the knowledge creation process for years, and the proliferation of GenAI now makes that possible.
- Given the manual nature of building and maintaining knowledge articles, I&O leaders find it challenging to keep IT staff engaged..
- Heads of I&O want more responsive KM practices to effectively support the proliferation of technology caused by digital business initiatives.
- Product teams have led to more distributed and federated support models. Automated knowledge generation can reduce the risk of knowledge silos by simplifying the process of knowledge authoring.
- AI applications in ITSM increasingly provide support for knowledge generation.
- Deep research techniques offer the promise of drafting articles sourced from multiple sources with full attribution, without relying on extensive case documentation.
- ITSM platform vendors, AI apps in ITSM, and conversational AI providers are increasingly embedding knowledge-generation features in their products.

Obstacles

- IT support teams often lack ticket documentation standards, which will reduce the value of knowledge generation.
- Automating the production of knowledge can increase the risk of knowledge bloat through low-value or redundant knowledge due to poor management controls on newly generated knowledge.
- Few knowledge-generation solutions automate the publishing pipeline and life cycle management for knowledge, as most solutions only focus on creating the knowledge asset text.
- LLMs risk misinterpreting the data they are fed, causing hallucinations (i.e., making up facts).
- There are data privacy, data ownership and security concerns about LLM data usage.
- Solutions offering this capability are often bundled with other GenAI features that are priced at a premium, requiring a broader strategy to drive the ROI.
- Many organizations use third-party support services that may not be incentivized for ticket documentation and/or KB creation.

User Recommendations

- Select an AI-enabled knowledge-generation solution that is integrated into the publishing life cycle and knowledge base and provides domain-trained LLMs.
- Have a human review and validate any AI-generated content, even previously published information that is only being updated.
- Investigate how GenAI can flag related knowledge and automate the maintenance of existing knowledge to keep knowledge assets current.
- Identify areas to automate the capture of information for ticket documentation, such as call transcripts or log data, which can be fed into LLMs to build knowledge drafts.
- Avoid dissolving the existing KM practice; instead, evolve it to account for system-generated knowledge.
- Encourage knowledge managers to work closely with the AI platform teams.
- Ensure that contracts with external support partners include terms to support knowledge capture and knowledge portability at the termination of the agreement.

Gartner Recommended Reading

Revitalize IT Support With GenAI-Powered Knowledge Management

Large Action Models

Analysis By: Frank O'Connor

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Definition:

Large action models (LAMs) are foundation models trained and optimized to identify and generate an action or set of actions that can be used to impact a target environment to meet a goal. For example, a large language model (LLM) solution can recommend a good restaurant to you; a LAM can book it for you.

Why This Is Important

CEOs and other business leaders are bullish about the potential of LAM-based applications to deliver new kinds of interactions — intuitive, highly personalized, flexible and efficient interactions that can create new value streams. Leaders should prepare by investing in LAM-powered AI agents that execute tasks to achieve desired goals.

Business Impact

Industries likely to benefit from LAMs are automotive, education, healthcare, professional services, technology and software, with impacts including:

- Enhanced customer service experiences with more natural interactions, improving user satisfaction.
- Guided intelligent decision support solutions across different business functions.
- Delivery of personalized content and experiences as well as curated learning journeys.
- Better patient monitoring and diagnostic support, streamlining healthcare workflows.

Drivers

- **Agentic workflows:** Agentic workflows represent a shift from traditional, linear processes to dynamic, AI-driven systems where humans and machines collaborate seamlessly. Unlike fully autonomous AI, agentic workflows highlight the complementary strengths of humans and AI, with AI acting as a cognitive catalyst and humans providing context, intuition and oversight.
- **AI breakthroughs:** Recent advances in generative pretrained transformers and large language models (LLMs) have greatly improved how AI understands human intent and plans actions. Reasoning models now enable deliberate, multistep planning. Test-time compute reasoning models represent a shift from pattern-matching to active inference, allowing these models to "think" before acting — a prerequisite for reliable, multistep task execution.
- **Action-centric training:** While LLMs are trained to predict the next word in a sequence of text, LAMs are trained to predict the next action in a sequence of tasks. This action-centric approach lets LAMs move beyond language modeling to support more complex reasoning and planning. With reinforcement learning, LAMs combine pattern recognition with advanced decision making for challenging tasks.
- **Reinforced learning integration:** LAMs also excel at multimodal understanding. They process vision, audio and language to create more flexible AI assistants. They can adapt to changes in workflows or interfaces and generate advanced automation from high-level instructions, reducing development time and effort.
- **Open-ended task capabilities:** While current AI models struggle with open-ended tasks, LAMs help agents determine the best actions for handling complex, wide-ranging goals. In autonomous systems, like self-driving cars, drones and space exploration, LAMs may enhance the ability of agents to execute complex objectives reliably.
- **Robotics:** In robotics, LAMs could empower robots to learn complex actions from data rather than relying on manual programming, with applications in manufacturing, logistics, healthcare and home assistance.

Obstacles

- **Lack of trust:** Organizations are unsure if the human customer can trust the technology to accurately predict and execute tasks, and if the machine customer can trust the organization offering the service.
- **Data scarcity:** Training LAMs requires vast datasets of mapped action sequences and their corresponding goals across diverse environments.
- **Computational demands:** LAMs are computationally intensive to train due to the complexity of modeling action sequences.
- **Interpretability and oversight:** Action policies learned by LAMs may be opaque and have poor explainability, requiring mechanisms for human interpretability, oversight and control.
- **Accessibility:** While LAMs and AI agents have the potential to transform access for disabled people, care must be taken to ensure their interfaces are accessible and inclusive.

User Recommendations

- Start with focused, constrained applications and environments rather than aiming for general capabilities initially.
- Invest heavily in high-fidelity simulation infrastructure to enable scalable training and safe iterative development before real-world deployment.
- Build interpretability, human oversight and control mechanisms from the ground up to ensure LAM actions remain transparent and aligned with intended goals.

Sample Vendors

Covariant; MultiOn; Rabbit; Salesforce; Toloka; Uniphore

Gartner Recommended Reading

Research Roundup for Generative AI

Emerging Tech: From Language to Action — Are Large Action Models the Future?

Forecast: Generative AI Models, Worldwide, 2024-2030, 1Q26

Agentic ITSM

Analysis By: Chris Matchett

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Agentic IT service management (ITSM) systems automate ITSM practices and complex tasks, with goal seeking, nondeterministic planning and decision making at runtime. This is accomplished through AI agents, which are autonomous or semiautonomous software entities that use AI techniques to perceive, make decisions, take actions and achieve goals in their digital or physical environments.

Why This Is Important

As IT environments grow more complex, heads of I&O seek solutions that transcend reactive, rule-based automation. Genuine agentic ITSM promises proactive, autonomous operations — unlike many vendor offerings that merely “agentic-wash” processes. While still near the Innovation Trigger, true agentic AI does have the potential to drive efficiency and cost savings.

Business Impact

- Streamlines IT service workflows by autonomously adapting to dynamic conditions and reducing manual intervention, resulting in faster incident resolution and improved operational agility
- Optimizes costs by intelligently automating complex ITSM tasks, reducing reliance on labor and lowering overall expenses
- Boosts service reliability with proactive diagnosis and self-healing for faster resolutions
- Enhances knowledge management with deep research and agentic search for accurate insights

Drivers

- Rapid progress in generative and agentic AI is creating opportunities for more sophisticated, goal-driven ITSM functionalities that transcend traditional automation.
- Heads of I&O are interested in AI agents to address chronic challenges such as increasing complexity, talent shortages and high operational costs.
- Rising end-user curiosity, combined with strong vendor promotion of “agentic” features in marketing, is fueling interest in autonomous AI applications for ITSM.
- Heads of I&O now demand ITSM solutions that go beyond rule-bound automation toward true autonomous capabilities.
- Expectations for rapid incident resolution and proactive support are pushing heads of I&O to explore innovative, self-optimizing agentic AI solutions.
- Heads of I&O increasingly want to leverage deep research and adaptive incident management to transform and optimize IT operations.

Obstacles

- Many vendor solutions overstate their agentic ITSM capabilities with semiautomated, rule-based approaches that fall short of true autonomy.
- While agentic AI is garnering hype among vendors and early inquiries from heads of I&O, without demonstrable ROI, true agentic ITSM risks becoming stuck in the Trough of Disillusionment.
- Implementing agentic AI systems requires robust foundational ITSM data and teams skilled in working with AI, both of which are lacking in many I&O organizations.
- Concerns persist that fully agentic AI IT systems could inadvertently cause unplanned disruptions to IT services unless robust safeguarding and governance mechanisms are in place.
- Real-world ITSM practices involving nondeterministic reasoning remain largely theoretical, making agentic ITSM primarily a “build” rather than a “buy” proposition.
- Vendors seeking to monetize agentic capabilities are increasingly turning to consumption-based pricing models, which are difficult for heads of I&O to plan for.

User Recommendations

- Use the “fauxgentic” checklist in How to Assess Agentic ITSM to systematically assess agentic capabilities and avoid investing in scripted automation solutions.
- Focus investments on AI applications that have demonstrated success in ITSM, such as automated change management, self-healing beyond scripted actions and deep research for knowledge generation, noting that such out-of-the-box capabilities are rare.
- Pause investment in agentic ITSM capabilities until they become mainstream if you are not prepared to build these use cases from foundational components.

Gartner Recommended Reading

Innovation Insight: Agent-Native I&O

Magic Quadrant for AI Applications in IT Service Management

Critical Capabilities for AI Applications in IT Service Management

How to Assess Agentic ITSM

Agentic AI

Analysis By: Erick Brethenoux, Pieter den Hamer

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Definition:

Agentic AI is an approach to building AI solutions that uses one or multiple software entities classified wholly or partly as AI agents. AI agents are autonomous or semiautonomous software entities that apply AI techniques to perceive, decide, act and pursue goals across digital or physical environments.

Why This Is Important

Agentic AI remains at the peak of the Hype Cycle, driven by rapid interest and ongoing confusion about its capabilities. It continues to ride both the generative AI and emerging multiagent hype waves. Yet, for the right use cases, agentic AI can deliver real value through deeper AI integration. AI agents are ushering in new software practices built on highly distributed decision-making systems.

Business Impact

Decades of AI agent systems, often based on embedded systems, show that agentic AI can generate significant business value when applied appropriately. Agentic AI creates this value through goal-driven systems that offer more flexibility, adaptability and higher levels of automation and, more importantly, augmentation — further bridging the gap between humans and machines.

Drivers

- **Multiagent systems:** Beyond hype around AI agents, coordination across multiple agents is also overhyped. Agentic AI often inherits misconceptions from interest in multiagent systems, which combine several agents to work toward shared goals.
- **Technological advancement:** Agentic AI draws on rapid advances in composite (hybrid) AI, decision intelligence and large action models.
- **Market momentum and hype:** Agentic AI enjoys strong market interest, with many organizations experimenting and investing in early pilots.
- **Vendor investment:** Vendors are accelerating agentic capabilities, which amplifies demand and accelerates the trend.
- **Advanced automation promise:** Agentic AI enables less brittle, more resilient, and more contextual process flows than traditional workflow automation or RPA, opening flexible levels of automation.
- **Complex use cases:** As business environments, goals, and execution grow more complex and dynamic, organizations need more distributed, less deterministic approaches, a fit for multiagent systems.

Obstacles

- Market hype is diluting the meaning of AI agents, with vendors engaging in “agent washing” by rebranding AI assistants, RPA tools and chatbots to attract buyers without delivering true agentic capabilities. This fuels false expectations about the technology’s maturity.
- Predictability is a core limitation for AI agents. Many AI agents can perform specific tasks, but they lack the reliability needed for consistent execution, making them unsuitable for full automation.
- Greater autonomy introduces new risks, exceeding those associated with stand-alone AI models or GenAI assistants. As agents take more independent actions, the potential impact of errors grows significantly.
- Most market offerings are still AI assistants, not true agents. Assistants rarely take self-directed actions or manage multi-step goals, yet users are becoming increasingly confident without understanding the technical and operational commitments required to build real agents.

User Recommendations

- Do not take vendor promises at face value. Assess whether their offerings truly qualify as AI agents or agentic AI, to what extent they do so, and whether they can deliver the expected benefits — not all use cases require these capabilities.
- Determine your actual need for agentic capabilities by evaluating each use case’s complexity, dynamics and other relevant characteristics.
- Evaluate alternative delivery approaches, such as conventional workflow automation or robotic process automation, to avoid deploying AI agents where simpler solutions are more effective.
- Prepare for the future of AI agents by building the foundations of application composability, governance, and data management. The agentic AI approach demands additional scrutiny and stronger security controls.

Sample Vendors

CrewAI; Dust; Epic Games; LangChain; Unity; XMPro

Prompt and Context Engineering

Analysis By: Frances Karamouzis

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

Definition:

Prompt and context engineering is the discipline of providing inputs, in the form of text or images, to GenAI models to specify the request. Prompt engineering is about how you phrase the instruction — word choice, structure, constraints and format of the request itself. Context engineering is bigger — it's about everything surrounding the prompt: background documents, examples, memory, retrieved data and system instructions.

Why This Is Important

Prompt and context engineering are the foundational anchors and linchpins of business alignment for desired outcomes. The quality (inclusive of the context) of the request is directly aligned to the quality, usability and impact of the results. Each model has its own sensitivity level, and the discipline of prompt and context engineering is critical to long-term value creation.

Business Impact

Context engineering is the practice of deliberately designing and structuring the information you provide to an AI model in order to get the best possible outputs. While prompt engineering focuses narrowly on crafting the right question or instruction, context engineering is broader — it's about shaping everything the model "sees" when generating a response.

LLMs have no persistent memory — they only know what's in their context window at inference time. The quality of their response is almost entirely determined by what you put in that window. Garbage in, garbage out; great context in, great output out.

Business impact:

- **Higher quality outputs:** Well-crafted prompts lead to more precise and useful outputs, which can improve the quality, usability and business impact of the results.

- **Optimized marginal cost of delivery:** Time to market, quality, efficiency and effectiveness are all key business impacts. By improving the efficiency and efficacy through reduced latency, speed and quality of results, the enterprise is impacting the marginal cost of delivery.
- **Performance:** Helps improve model performance and reduce hallucinations.
- **Business alignment:** Allows various roles to steer foundation models, which are general-purpose in nature, to align to the business, domain and industry.
- **Rapid prototyping:** Businesses can quickly test out different approaches to prompting, allowing for agile experimentation and adaptation to new requirements or user needs.
- **Broader applicability:** By fine-tuning prompts, companies can adapt their AI applications to diverse functions — from marketing and sales to technical support — without extensive retraining.

Drivers

- **Balance efficiency, efficacy and agility:** The fundamental driver for both prompt and context engineering is it allows organizations to strike a balance between efficiency, efficacy and business agility premised on their enterprise-specific business demands.
- **Process or task-specific customizations or new use cases:** The insertion of context and patterns a model uses to influence the output generated allows for customizations for a particular enterprise or domain, or for regulatory items. Prompts are created to help improve the quality for different use cases — such as domain-specific question answering, summarization and categorization — with or without the need for fine-tuning a model, which can be expensive or impractical. This would also apply to creating and designing new use cases that utilize the model's capability for image and text generation.
- **Validation and verification:** It is important to test, understand and document the limits and weaknesses of the models to ensure a reduced risk of hallucination and unwanted outputs.

Obstacles

- **Approaches and techniques:** A unified approach to performing prompt engineering does not exist. Complex scenarios need to be broken down into smaller elements. It is challenging to debug complex prompts. Understanding how specific prompt elements influence the logic of the LLM is vital.
- **Dynamic vendor (provider) landscape:** Providers are constantly fine-tuning and tweaking techniques and approaches that influence behavior within their offerings. Organizational teams creating prompts must iteratively adapt and pivot to ensure scalable and maintainable methods of prompt engineering.
- **Scalability:** Scaling prompt engineering for large projects or across multiple teams can be difficult due to the manual and iterative nature of the process. Scalable and maintainable methods of prompt engineering are still a work in progress for most organizations.

User Recommendations

- Recognize and embrace the underlying premise that strong prompt and context engineering is directly correlated to quality results and outputs. And this impacts cycle time and speed to outcome.
- Invest in the skill of prompt and context engineering as a “must-have” for a variety of roles in the organization, such as data scientists, business users, domain experts, software engineers and citizen developers.
- Elevate the team with the myriad of prompt optimization tools that will diminish (or at the very least shift) the need for manual engineering.
- Communicate and cascade the message that prompt engineering is not foolproof. Enterprise teams apply rigor and diligence to permeate and work to ensure successful solutions.

Sample Vendors

AIMon; Airia; Anthropic; FlowGPT; Google; HoneyHive; LangChain; Microsoft; PromptBase; PromptLayer; Zep AI

Gartner Recommended Reading

How to Engineer Effective Prompts for Large Language Models

Prompt Engineering With Enterprise Information for LLMs and GenAI

Guide to Designing the Context Engineer Job in the Age of AI

Guide to Redesigning the Machine Learning Engineer Job in the Age of AI

Innovation Insight: Context Engineering

Multimodal Generative AI

Analysis By: Danielle Casey, Roberta Cozza

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Emerging

Definition:

Multimodal generative AI is the ability to use multiple types of data inputs and outputs in generative models, such as images, videos, audio, text, numerical data, maps and biometric information. Multimodality refers to both data input and output, in which multiple data formats can be used or consumed during an interaction. Multimodality augments the usability of AI by allowing models to interact with and create outputs across various modalities.

Why This Is Important

Multimodal GenAI is important because real-world data is multimodal. Multimodality helps capture the relationships between different data streams and scales the generation and automation benefits of GenAI across applications and workflows. Multimodal functionality will increasingly be present in tech offerings as foundation models lower the technical barrier to multimodal functionality, and as users demand associated performance.

Business Impact

Multimodal GenAI is increasingly having a transformational impact on enterprise applications by enabling the addition of new features and functionality otherwise unachievable. The impact of multimodality is not limited to specific industries or use cases and can be applied at any touchpoint between AI and humans. Today, text, voice and image are common modalities. Over the next few years, models will evolve to support additional modalities that are more complex and domain-specific. As multimodal GenAI matures, multimodal understanding will evolve into multimodal actioning, where agents automate tasks from multimodal data inputs.

Drivers

Multimodal GenAI is being driven by:

- **Model innovation:** GenAI models are lowering the technology barrier for providers to offer multimodal functionality in their offerings. The emergence of domain-specialized models will also drive multimodal innovation. Models trained on industry-, role- or use-case-specific data will be increasingly multimodal (e.g., legal, finance, healthcare, marketing and multimedia). World models will also create new multimodal opportunities in physical AI.
- **Improved user experience (UX):** Multimodality improves UX by enabling richer experiences as it meets users with the data that is available. The ability to interact with both unstructured and structured data is supporting additional use cases within the organization.
- **Extensibility of automation:** Multimodality supports new tasks, such as data extraction, data manipulation (converting one data type to another) and creating new data outcomes. Multimodal applications will have a higher automation potential, and the use of multimodality with agentic automation will create new use case opportunities. By 2030, multimodality will increase AI automation potential by 50% with new functionality and use cases.

Obstacles

Multimodal GenAI models are primarily inhibited by:

- **Training challenges:** Multimodal data has varying degrees of quality and formats compared to unimodal data, amplifying challenges with training costs and time, as well as potential inaccuracies and biases in output from multimodal data sources.

- **Lack of data:** Data availability may be limited in some modalities (e.g., large-scale audio datasets or healthcare imagery), which impacts model training and accuracy.
- **Data governance:** Multimodal GenAI increases the exposure to a wider range of sensitive data. Examples of particularly sensitive data types include maps or geolocation data, biometric data or health data. Enterprise governance standards and the lack of regulations and standards are also inhibiting capability advancement.
- **Lack of application experience:** Best practices, testing methods, software libraries and supporting tools have not been fully developed to enable multimodal applications.

User Recommendations

- Identify two or three modalities that are most important to your organization when choosing which GenAI model to use for which use case.
- Evaluate and explore use cases where multimodal capabilities may offer differentiating capabilities or new opportunities that are out of reach of unimodal offerings.
- Assess the technical complexities of processing and integrating data inputs and outputs from diverse multimodal sources, and the potential risks and controls required for using agents to automate tasks based on multimodal inputs.
- Create specific cross-modal data policies and tools aimed at protecting privacy, detecting bias and ensuring compliance with emerging AI regulations.

Sample Vendors

Alibaba Cloud; Anthropic; Archetype AI; DeepSeek; Google; Microsoft; Mistral AI; OpenAI; Stability AI; TwelveLabs

Gartner Recommended Reading

Emerging Tech: Multimodal Generative AI Interfaces Transform User Experiences

Emerging Tech Impact Radar: Generative AI

Emerging Tech: Data Fabrics With Multimodal Data Focus for Generative AI-Enabled Applications

AI Agents

Analysis By: Tom Coshow, Haritha Khandabattu

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

AI agents are autonomous or semiautonomous software entities that use AI techniques to perceive, make decisions, take actions and achieve goals in their digital or physical environments.

Why This Is Important

AI agents have the ability to make decisions and take action in their target environment to achieve organizational goals. By using AI practices and techniques such as LLMs, organizations are creating and deploying AI agents to achieve complex tasks.

Business Impact

AI agents have the potential to:

- Revolutionize a broad range of industries and environments with their ability to automate tasks from consumer, industrial, data analytics, content creation and logistics.
- Make informed decisions and interact intelligently with their surroundings.

Drivers

- **Generative AI breakthroughs:** Reasoning models and LAMs advance the ability to plan a complex series of actions.
- **Multimodal understanding:** The ability to use diverse modalities like vision, audio and language enables more general and flexible AI agents. This allows automatic adaptation to changes in the workflow, user interface or API. With this, one can create advanced workflows without explicit programming, significantly reducing the development time and effort for automation.
- **Increased decision-making complexity:** AI is increasingly used in real-world engineering problems containing complex systems, where large networks of interacting parts exhibit emergent behavior that cannot be easily predicted. AI agents can learn, plan and execute in complex environments.
- **Composite AI, including neurosymbolic models:** Advances in models that improve planning and problem solving are enabling more complex AI agents. AI agents can utilize a wide variety of AI practices to forecast, make decisions and plan.

Obstacles

- **Vulnerabilities:** Due to the complexity of the AI agent system, all components face various potential vulnerabilities such as access security, data security and governance.
- **Lack of trust:** Users are unsure whether they can trust the technology to accurately predict and execute tasks independently. Without a human in the loop, agents may take multiple consequential actions in rapid succession and bring about significant impacts before a human notices.
- **Interpretability and oversight:** Action policies may be opaque and have poor explainability, requiring mechanisms for human interpretability, oversight and control.
- **Pace of change:** The technologies deployed, from models to tooling, and the framework options available are changing rapidly, making it difficult for organizations to define their roadmap.

User Recommendations

- Incorporate AI agents into strategic planning by investing in understanding their capabilities and potential applications in various environments, considering their increasing autonomy and wide-ranging usability.
- Investigate the possibilities of utilizing multiagent systems, collectives of AI agents, that can operate both collaboratively and independently, enhancing adaptability and flexibility in response to different tasks and scenarios.
- Promote the development and integration of the use of a variety of AI practices, enabling learning, negotiation and decision-making capabilities.

Sample Vendors

Amazon; Anthropic; CrewAI; Google; LangChain; Maisa; Microsoft; OneReach.ai; OpenAI; Salesforce

Gartner Recommended Reading

Innovation Insight: AI Agents

Build AI Agent Services to Revolutionize Client Operations

Innovation Insight for the AI Agent Platform Landscape

AI Agents: Are You Ready to Set Your AI Free?

Emerging Patterns for Building LLM-Based AI Agents

Sliding into the Trough

AI TRiSM

Analysis By: Avivah Litan, Jeremy D'Hoinne, Bart Willemsen, Lauren Kornutick

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Definition:

AI trust, risk, and security management (AI TRiSM) comprises four layers of technical capabilities that support enterprise policies for all AI use cases and help assure AI governance, trustworthiness, fairness, safety, reliability, security, privacy, and data protection. The top two layers — AI governance, and AI runtime inspection and enforcement — are purposefully directed to support AI. The bottom two layers represent information governance and traditional technology in the context of AI.

Why This Is Important

AI brings new trust, risk, and security management challenges that conventional controls do not address. Top concerns for enterprises include data compromise, third-party risks, undesired outcomes, and the need to ensure enterprise AI actions align with deployment purposes. Organizations must retain independence from any single AI model or hosting provider to ensure scalability, flexibility, and trust (and cost control as a derivative benefit) as AI markets rapidly change.

Business Impact

Those failing to manage AI risks experience project failures, underperformance, and compromised data. Inaccurate, unethical/unintended AI outcomes, and interference from malicious actors can result in financial and reputational loss, liability, and social harm. AI underperformance can also lead to poor business decisions and uncontrolled costs. Applying AI TRiSM improves project timelines, operational precision, product durability, and enhanced customer trust and overall AI investment ROI.

Drivers

- The increasing use of AI, GenAI, and AI agents is limited by a lack of trust in AI as a safe and ethical option for supporting critical business processes. Fewer than 5% of respondents fully trust their vendor's hallucination safeguards, AI security, and governance controls according to the 2025 GenAI and Agentic AI in Enterprise Apps Survey (see Assessing the Impact of Generative AI and Agentic AI In Enterprise Applications).
- Enterprises face multiple AI risks and are most concerned with data compromise, third-party risks, and inaccurate or unwanted output.
- Large language models are nondeterministic, and their output and behavior are unpredictable.
- Regulations for AI risk management (such as the AI Act in Europe or Local Law 144 in New York) are driving businesses to institute measures for managing AI risk. Such regulations define new compliance requirements that organizations will have to meet on top of existing ones, like those pertaining to privacy protection.
- Malicious hacks against enterprise AI are still uncommon, while incidents of unconstrained harmful chatbots are well-documented, and internal oversharing data compromises are prevalent.
- AI agents pose new risks of aberrant behavior deviating from human instructions and greatly expand the attack surface.
- The rapid proliferation of AI agents will create more need for governance than human-in-the-loop oversight can fulfill alone.
- User demand for AI TRiSM solutions is increasing, and providers of all sizes are competing for this new enterprise business. Many AI TRiSM startups have been acquired by large security vendors.
- Some organizations are mostly concerned with security and risk mitigation, while others also focus on supporting ethical or safe practices and regulatory compliance.
- AI trust, risk, and security issues surface organizational silo issues, pushing teams to realign to solve problems that cross departmental boundaries and to implement technical measures that address them.

Obstacles

- Adopting AI TRiSM technology is often an afterthought. Many organizations don't consider it until AI applications or agents are in production, when it becomes challenging to retrofit.
- Many enterprises are resource-constrained and don't have the skills or capacity to implement AI TRiSM.
- Enterprises often rely on their incumbent vendors to provide AI TRiSM capabilities, although they often lack it and must rely on vendor licensing agreements to ensure their confidential data remains private in the host environment.
- Most embedded software and SaaS services use AI but they often do not support APIs to third-party AI TRiSM products that can enforce enterprise policies.
- AI TRiSM requires a cross-functional team, including legal, compliance, cybersecurity, IT, and data analytics staff, to establish common goals and use common frameworks. Coordination between these teams might be lacking, leading to competing or overlapping technologies. This fragmentation of TRiSM efforts hinders AI progress.

User Recommendations

- Set up an organizational unit to manage AI TRiSM and include members with a vested interest in AI projects.
- Discover and inventory all AI used in the organization, leveraging the capabilities of TRiSM vendors who support this.
- Define acceptable use policies that are flexible with just the right level of granularity for agile enforcement.
- Revisit and implement data classification, protection, and access management across all enterprise information that can potentially be used by AI. Collaborate across the teams involved in governance.
- Work with legal and compliance to contractually define accountability for unacceptable AI use or behavior in third-party-embedded AI applications.
- Obtain vendor attestation to meet legal requirements.

- Evaluate and implement layered AI TRiSM technology to support and enforce policies across all AI in use. This includes enterprise-owned services, along with controls offered by incumbent model and platform providers. However, do not solely rely on the latter.

Gartner Recommended Reading

Market Guide for AI Trust, Risk and Security Management

Top Strategic Technology Trends for 2024: AI Trust, Risk and Security Management

The Cybersecurity Leader's Playbook: Navigating the EU AI Act

How to Build a Lightweight Organizational Structure for AI Governance

Market Guide for Guardian Agents

Retrieval-Augmented Generation

Analysis By: C.A. Swan, Afraz Jaffri

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Definition:

Retrieval-augmented generation (RAG) is an architecture design pattern that uses search functionality to retrieve relevant data and add it to the prompt of a generative AI (GenAI) model in order to ground the generative output with information. RAG can be used for retrieving both public internet data and data from private organizational knowledge bases.

Why This Is Important

The majority of available foundation large language models (LLMs) are trained on general public information. For LLM completions to be contextual and trustworthy for enterprise use cases, they need to be informed by retrieved information from organizational knowledge bases. A correct RAG implementation reduces hallucinations and provides the auditable, source-cited outputs increasingly required by enterprise governance and emerging AI regulations.

Business Impact

Workers often struggle to find the information required to complete business tasks. RAG can enhance enterprise search and how information is summarized and created, reduce hallucinations, and accelerate productivity. The applications for self-service content consumption for employees and for customer-facing applications have a significant impact on employee and customer satisfaction, brand equity, and workforce productivity.

Drivers

- Searching and finding current information and knowledge documented in enterprises can be frustrating for workers. With RAG, they benefit from a better information retrieval and synthesis approach.
- The advances of GenAI applied toward content summarization and contextualization have enabled search applications to achieve a more robust and focused content consumption experience.
- GenAI service vendors, either via a hyperscale marketplace or via specialty model APIs, are making it easier for organizations to configure the RAG pattern on their knowledge bases. Many more pilot projects are moving to production.
- Competitive pressure to adopt the latest innovations in order to increase productivity and maintain competitive edge drives organizations to use their knowledge bases to support the RAG pattern for both internal- and external-facing applications.
- Availability of technologies for building RAG, as well as RAG being baked into GenAI assistant experiences, is penetrating enterprises.
- Multiple vendors are offering RAG integrated into their search and research tools; examples include deep research from tools like OpenAI ChatGPT, Anthropic Claude, Perplexity, and Google NotebookLM. These are driving the adoption of RAG by end users.
- AI agents are increasingly using RAG to get grounded in organizational knowledge in order to take the correct actions in achieving their intended goals.

Obstacles

- Scaling RAG requires robust data governance, quality metadata, secure system connectors, and ongoing index updates. Without this foundation, organizations encounter increasing challenges as they move beyond pilot stages.
- Configuration of knowledge base access controls for the RAG pattern is not a trivial task and can prevent widespread adoption across all organizational knowledge.
- Advanced RAG architectures — such as agentic retrieval, multimodal RAG, and knowledge-graph-based retrieval (GraphRag) — are now essential for a wide range of enterprise use cases. This has led to architectural decision fatigue, as organizations face challenges in selecting and managing the optimal solutions for diverse queries and data types.
- Concerns about IP protection and responsible AI in the use of LLMs are still an obstacle.
- Vendor fragmentation and evolving enterprise RAG use cases are placing increased engineering and architectural demands on technical teams.
- Context engineering, including the use of larger context windows, is seen as an alternative to RAG.

User Recommendations

IT and D&A leaders looking at adopting GenAI capabilities on top of private and public corporate data should:

- Prioritize RAG investments by assessing the maturity and readiness of the enterprise in relation to knowledge management and quality and information retrieval for employees and customers.
- Pilot applications using the RAG pattern on a well-known knowledge base to assess the lift in the content consumption experience and to gain buy-in for further investment.
- Plan for investment in filling any skills gaps that exist in internal knowledge engineering capabilities, either through upskilling or external hiring.

- Engage with technology vendors or service providers that offer vertical or domain-specific RAG platforms with enterprise-grade governance built in, including role-based access controls, audit logging, personally identifiable information (PII) masking, and compliance certifications (SOC 2, HIPAA, GDPR). Prioritize vendors that bundle agentic capabilities alongside retrieval to avoid fragmented architecture.

Sample Vendors

Amazon Web Services; Coveo; Elastic; Glean; Google; Graphwise; Microsoft; Squirro; Vectara; WRITER

Gartner Recommended Reading

Getting Started With Retrieval-Augmented Generation

Selecting Retrieval-Augmented Generation Architectures for Generative AI Applications

Emotion AI

Analysis By: Annette Zimmermann, Roberta Cozza

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Definition:

Emotion AI technologies, also called affective computing, use various models (NLU, GenAI, ML, CNN and RNN) to analyze the emotional state of a user via computer vision, audio/voice input, sensors and/or software logic. Emotion AI can facilitate responses by performing specific, personalized actions to fit the mood of the customer.

Why This Is Important

Emotion AI is considered high impact as it turns human behavioral attributes into data that will have a large impact on human-machine interfaces. Machines will become more “humanized” as they can detect sentiments in many different contexts. Furthermore, the emergence of multimodal AI models trained with text, audio and images can be used to provide emotional insights.

Business Impact

The past 12 months witnessed a shift in the market wherein a number of contact center voice analytics providers moved away from NLU-based systems and toward LLM-based technology. Emotion AI based on computer vision has already been used for more than a decade in market research with neuromarketing platforms that test users' reactions toward products. We see the technology expanding to other verticals, including healthcare (diagnostics), sales enablement and employee wellness.

Drivers

- Vision-based emotion AI systems are driving the majority of new use cases at this time. Some new use cases have emerged. For example, emotion analysis embedded in video-call software and integrated in telemedicine applications in healthcare to monitor patient well-being.
- Another driver for detecting emotions/states is the need for a system to act more sympathetically. For instance, emotion AI creates anthropomorphic qualities that will play an important role in physical AI, specifically for humanoid robots. This "emotional capability" is an important element in enhancing the communication and interaction between users and humanoid robots.
- Emotion AI vendors are starting to combine their technology with GenAI; for example, by creating automated reporting functionalities after the emotion analysis.
- Multimodal GenAI models that support multiple data input and output types (e.g., images, video, audio and numerical data) are advancing emotional understanding because they can leverage large multimodal datasets for emotion indicators. This means multimodal GenAI is improving emotion understanding for personalization of services. Examples are conversational AI avatars and customer support agents, as well as use cases in marketing and customer service to generate emotionally compelling campaigns, content and interactions.
- Market research and neuromarketing tools are continuously leveraging emotion detection in various user scenarios, including focus groups and product testing.

Obstacles

- The advancement of GenAI technology has turned out to be a retractor for emotion AI in certain areas, specifically in contact centers. Vendors have replaced emotion AI models and natural language processing (NLP) algorithms in some instances with LLMs for simple text-based sentiment analysis.
- Privacy concerns are the main obstacle to rapid adoption in the enterprise. Privacy is especially a concern in real-life environments (versus lab environments) for both consumer-facing situations, like monitoring emotions in a retail environment via cameras and employee-facing situations.
- The EU AI Act prohibits computer vision-based emotion detection systems in education. The ban has brought a few projects to a stop in this region.
- Emotion AI may be suspected of bias and lacking nuance. When using facial expression analysis, models are likely to be retrained in different geographies and ethnicities to get the system to detect the different nuances present due to different cultural backgrounds.

User Recommendations

- Review vendors' capabilities and reference cases carefully. As the market is currently immature, most vendors are focused on limited use cases in two or three industries.
- Monitor the development around multimodal GenAI models, as multimodality can increase output accuracy.
- Enhance your customer analytics and behavioral profiling by applying emotion AI technologies, bringing your customer experience strategy to the next level.
- Be use-case-driven. The use case will determine vendor selection and the emotion AI technology to be used.
- Appoint responsibility for data privacy in your organization — a chief data privacy officer or equivalent.
- Work with your vendor on change management to avoid user backlash due to collection of sensitive data.

Sample Vendors

Emotion Logic; Hume AI; Intelligent Voice; kama.ai; MetaSoul; MorphCast; Soul Machines; Stern Tech

Gartner Recommended Reading

Emerging Tech: AI Vendor Race: Emotion AI Adoption Advances Through Computer Vision

Emerging Tech: Sensors and GenAI Unlock Innovations in Advanced Wearables

Foundation Models

Analysis By: Arun Chandrasekaran

Benefit Rating: Transformational

Market Penetration: More than 50% of target audience

Maturity: Mature mainstream

Definition:

Foundation models are large-parameter models that are trained on a broad gamut of datasets in a self-supervised manner. They are mostly based on transformer or diffusion deep neural network architectures and are increasingly becoming multimodal. They are called foundation models because of their critical importance and applicability to a wide variety of downstream use cases. This broad applicability is due to the pretraining and versatility of the models.

Why This Is Important

Foundation models are an important step forward for AI due to their massive pretraining and wide use-case applicability. They can deliver state-of-the-art capabilities with higher efficacy than their predecessors. They've become the go-to architecture for natural language processing and have also been applied to computer vision, audio and video processing, and software engineering use cases.

Business Impact

With their potential to enhance applications across a broad range of enterprise use cases, foundation models are having a wide impact across vertical industries and business functions. Their impact has accelerated, with a growing ecosystem of AI-native startups building enterprise applications and AI agents on top of them. The recent improvements in reasoning and tool usage are enabling more autonomous and task-specific workflows across a broad set of use cases.

Drivers

- **Quicker time to value:** Foundation models can effectively deliver value through prebuilt APIs, prompt engineering, retrieval-augmented generation or further fine-tuning. While fine-tuning may enable more customization, the other three options are less complex, quicker and cheaper.
- **Superior performance across multiple domains:** The difference between foundation models and prior neural network solutions is stark. The large pretrained models can produce coherent text, code, images, speech and video at a scale and accuracy not possible before and are increasingly becoming better at reasoning tasks.
- **Fast-paced innovation:** The past year has seen an influx of foundation models with better reasoning and tool use, along with smaller, pretrained domain-specific models built from them. Most of these are available as cloud APIs or open-source projects, further reducing the time and cost to experiment and driving quicker enterprise adoption.
- **Productivity gains:** Foundation models are having an impact across broad swaths of enterprise business functions as their ability to automate tasks gets wider. Business functions such as marketing, customer service and IT (especially software engineering) are areas where clients have witnessed initial gains.

Obstacles

- **Flawed results:** Although a significant advance, foundation models still require careful training and guardrails. Because of their training methods and black-box nature, they can deliver unacceptable results or hallucinations. They also can propagate downstream any bias or copyright issues in the datasets.
- **Require appropriate skills and talent:** As with all AI solutions, the end result depends on the skills, knowledge and talent of the trainers and users, particularly for prompt engineering and fine-tuning.
- **Data integration:** Adapting foundation models to specific enterprise use cases often requires building RAG pipelines, context graphs, fine-tuning or significant prompt engineering automation. While the know-how of how to implement this is growing, highly effective techniques to ground these models remain both technically complex and expensive.

User Recommendations

- Start with models that have superior ecosystem support and adequate enterprise guardrails around security and privacy and are more widely deployed.
- Be objective about the adequate balance between accuracy, costs, security and privacy, and time to value when selecting foundation models to determine the appropriate model needed. Beware of building models from scratch, given the complexity and steep costs.
- Educate developers and data and analytics teams on prompt engineering and other advanced techniques needed to steer these models.
- Designate an incubation team to monitor industry developments, communicate the “art of the possible,” experiment with business units and share valuable lessons learned companywide.
- Build a composable AI platform architecture to leverage an ecosystem of models based on adequate trade-offs across accuracy, latency, cost and safety.
- Always consider pairing foundation models with rich context layers for grounding and guardrailing the reasoning and planning stages of foundation models in order to better serve the use cases.

Sample Vendors

Alibaba Group; Anthropic; Cohere; DeepSeek; Google; IBM; Meta; Microsoft; Mistral AI; OpenAI

Gartner Recommended Reading

Innovation Guide for Generative AI Models

Answering IT Leaders' Top 10 Questions on Open Generative AI Models

Explore Small Language Models for Specific AI Scenarios

Conversational User Interfaces

Analysis By: Gabriele Rigon, Adrian Lee, Bern Elliot

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Conversational user interfaces (CUIs) are human-computer interfaces that enable natural language interactions via multimodal inputs for the purpose of fulfilling a request, such as answering a question or completing a task. CUIs' sophistication varies from understanding basic queries, such as FAQs, to handling complex dialogues. CUIs are the frontend of tools, ranging from simple Q&A chatbots and conversational search to advanced voice AI use cases, AI assistants and conversational AI agents.

Why This Is Important

In CUIs, users control interactions through text or voice inputs using natural language, which can be combined with images and video.

Unlike graphical user interfaces (GUIs), CUIs remove the need for users to learn application-specific commands, thereby reducing friction and cognitive load. CUIs increase accessibility and user satisfaction, and allow organizations to deliver seamless, context-aware experiences across multiple platforms and channels.

Business Impact

CUIs can:

- Enhance the usability of enterprise applications across business functions facilitating access to enterprise knowledge, improving efficiency and reducing cost to serve.
- Augment customer experience by automating support through conversational AI assistants and AI agents.
- Allow organizations to scale support while governing consistent communication across channels and business units.
- Support multiple languages, allowing organizations to serve a global workforce and customer base.

Drivers

- **Users' expectations, accelerated by GenAI and agentic AI:** Users expect to converse with applications and interact in natural language. CUIs, in the form of AI assistants or conversational AI agents, are complementing or even replacing traditional interfaces in some applications, such as search, business intelligence platforms and productivity software, including document and spreadsheet applications.

- **Conversational AI solutions:** The underlying technology supporting custom-developed CUIs built on conversational AI platforms (CAIPs) has matured significantly in recent years, with vendors embracing GenAI and agentic AI trends. More application-level solutions have also emerged to compete with platforms, including stand-alone GenAI-native applications, such as ChatGPT Enterprise or Amazon Q, and GenAI-enabled extensions of platforms for productivity suites, CRM, ERP, ITSM and contact center technology. AI engineering environments designed for AI engineers and software developers are increasingly positioned to compete with CAIPs as the preferred deployment model for high-maturity organizations.

- **Multimodal interactions:** Multimodal models increase the feasibility of multimodal interactions, such as inputting images, videos, audio or other sensory data, or receiving them as output. Beyond text-based interfaces, voice continues to represent a key mode of interaction with CUIs. In some use cases, such as in customer service and support, voice-based interaction significantly enhances communication.

- **Accessibility requirements:** The global legislative efforts and contractual obligations related to accessibility are creating a demand for digital services that can meet these requirements. CUIs, particularly multimodal CUIs, can help fulfill these needs by enhancing accessibility to applications that may have been previously hard to use or inaccessible to hearing or visually impaired users.

Obstacles

- Ungoverned and tactical sprawl of AI-enabled CUIs, such as conversational AI Agents, leads to increased risks and abandoned projects.
- CUIs are not the best approach when tight control over input or output or intricate data handling is required, and GUIs should be preferred.
- Developing multimodal CUIs is complex and requires more effort than GUIs. Significant attention should be dedicated to GenAI-specific guardrails, as well as controls around AI agent components.
- UX may be negatively affected by content anomalies and poor conversational design, and the uncertainties of emerging agentic AI methods.
- Understanding the benefits and limitations of CUI solutions is complex, and market fragmentation and silos may lead buyers to choose suboptimal solutions.
- Accelerated offerings misaligned with enterprise-grade requirements and information governance gaps may delay or block deployments.
- Measuring the value of CUIs may be problematic and ambiguous, especially in employee-facing scenarios.

User Recommendations

- Design a strategy for consolidation upon one or a few CAI solutions or approaches, avoiding technical debt from the proliferation of CUIs deployed by different business units in different regions.
- Determine whether the CUI needs to incorporate any GenAI, agentic AI or multimodal capabilities, as these may not only deliver additional value but also introduce new risks and necessitate the implementation of enhanced controls.
- Prioritize preliminary information governance, as well as accuracy, privacy and security when evaluating design requirements of your custom CUIs. Then incorporate sophistication, integration flexibility and continuous monitoring and improvement.
- Define metrics to calculate the expected ROI at the time of designing the CUI use case.

Sample Vendors

Amazon Web Services; Boost.ai; Genesys; Google; Kore.ai; Microsoft; NiCE Cognigy; Omilia; OpenAI; SoundHound

Gartner Recommended Reading

Market Guide for Conversational AI Solutions

Generative AI Brings Opportunity and Risks to the Conversational AI Market

Emerging Tech: Market Risk Projection of Generative AI on Conversational AI

Emerging Tech Impact Radar: Conversational Artificial Intelligence

Emerging Tech: AI Vendor Race: GenAI Is Driving Investment for Conversational AI

Event Intelligence Solutions

Analysis By: Martin Caren

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

Event intelligence solutions (EISs) apply AI and data analytics to system and service events to augment, accelerate, and automate manual efforts in the event management process. EISs are defined by the key characteristics of cross-domain event ingestion, topology assembly, event correlation and reduction, pattern recognition, and remediation augmentation. They help reduce event noise and improve productivity by processing event streams into actionable insights and enabling proactive responses.

Why This Is Important

Rising application complexity and event volumes from multiple tools and platforms overwhelms manual processes. EISs use AI to classify and cluster cross-domain events in near real time, delivering insights that speed up incident response, reduce risk, and help ensure business continuity. EISs help keep IT operations running more efficiently in large, heterogeneous IT environments.

Business Impact

EISs deliver value through:

- **Improved agility and productivity:** Operators focus on critical issues, resolve incidents faster, and boost efficiency.
- **Enhanced service availability and reduced triage costs:** Faster root cause identification and automated remediation minimize downtime and lower operational expenses.
- **Maximized return on monitoring investments:** Unified events across tools enable informed decisions, rapid resolution, and better team collaboration.

Drivers

Demand for EIS capabilities is accelerating and fueled by:

- **Increasing complexity:** Organizations rely on a growing portfolio of monitoring and observability solutions to ensure the reliability and resilience of increasingly complex and distributed, hybrid and multicloud workloads.
- **Increasing monitoring expectations:** Investments and improvements in monitoring, and the pursuit of observability, generate more data from more sources. This increase in data presents operators with extremely detailed views into their applications, business services, and the end-user experience. Effective use of this additional data requires near-real-time analysis and correlation of events from related assets and services.
- **Demand for reliability:** Shifts in roles and responsibilities are driven by modern operating models, like DevOps and site reliability engineering (SRE), in the pursuit of greater availability and faster incident resolution. EISs enable agility by offloading some mechanical tasks of event triage, root cause analysis, and solution identification, accelerating response to common issues and freeing up human creative capacity for novel events and business priorities.

Obstacles

- **Unrealistic expectations and ROI:** Hype makes it hard for clients to separate AI claims from reality, making ROI difficult to demonstrate. EISs are often seen as operator tools, not as strategic assets, so they are vulnerable during budget cuts.
- **Maturity and complexity:** EIS success typically requires high maturity in dependencies like change management and automation. Without this, it adds a layer of complexity that can outweigh the benefits. Furthermore, fragmented data access remains a persistent bottleneck.
- **Market overlap and AI SRE:** EISs overlap with observability and IT service management (ITSM), creating a crowded and sometimes confusing market. Newer, AI-driven SRE models may also leapfrog traditional EISs, which must evolve beyond correlation to avoid becoming a legacy utility.

User Recommendations

- Layer the cross-domain analysis of an EIS with maturing your monitoring and observability strategy. This approach creates a solid foundation of valuable data for ingestion and analysis, and the surfacing of insights across domains.
- Do not focus solely on automated remediation; this is rarely achieved at scale. Accelerating response and augmenting human decision making have tremendous value. These approaches often avoid the challenge of the probabilistic uncertainty, combined with automated change in production environments.
- Explore EISs that make use of agentic AI for enhanced diagnostic capabilities and the possibility of accelerated remediation and root cause analysis (RCA).

Sample Vendors

BigPanda; BMC Software; Cisco; GreySkies; IBM; Interlink Software; PagerDuty; ServiceNow; Virtana; Vitria

Gartner Recommended Reading

Market Guide for Event Intelligence Solutions

Observability Tool Consolidation Must Begin With Event Intelligence

Choosing the Right Observability Architecture: Blueprints and Decision Criteria

GenAI Virtual Assistants

Analysis By: Danielle Casey, Bern Elliot

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Definition:

GenAI virtual assistants (VAs) represent a newer generation of VAs that leverage LLMs to deliver functionality not obtained with traditional conversational AI technology. GenAI supports better intent understanding and dynamic responses, offers content generation features (create new data assets) and the ability to understand multiple data types (text, image, video). These capabilities allow GenAI VAs to support more advanced Q&A and complex task automation to deliver better outcomes.

Why This Is Important

All leading VA providers have productized their large language models (LLMs) and GenAI capabilities. LLMs including multimodal variations can be embedded into a VA or can be accessed via an API call. There are multiple approaches to using an LLM — deployed out of the box, fine-tuned or retrained for customization, chained with additional LLMs across various tasks, or contextualized via prompt engineering. Each option must be evaluated against performance requirements and costs.

Business Impact

- Improve performance and flexibility in conversational dialogue, enabling greater contextuality, reducing development effort and operational support costs, and improving self-service.
- Extend automation by expanding from Q&A support to content discovery and creation features.
- Improve user interaction and engagement by supporting multimodal interactions (beyond text and voice, for example, via images and videos).
- Provide better outcomes without the perceived risk of more-advanced AI agent automation.

Drivers

- **Demand for GenAI value:** GenAI-enabled VAs are a popular use case, as the use cases and key performance indicators of VA implementations are well-defined and measurable, and the need to improve customer or employee support may already be recognized. This makes VAs a good option for organizations piloting GenAI.
- **GenAI-enabled use cases:** Organizations looking for promised value outcomes, such as productivity improvement and cost reduction from Q&A support for customers, information discovery and generation features for employees, and agent assist for call center operators. GenAI VAs are also becoming specialized and offering deeper automation by use case, business role, or industry domain.
- **Rapid time to market:** GenAI VA offerings are focused on lowering the barrier to adoption via easy-to-use platforms where users can select their preferred LLMs or prepackaged use case integrations.
- **Model innovation:** Vendor investments into domain specialized models, reasoning models, and multimodal functionality is enabling GenAI VAs to tackle more complex tasks and use cases.

Obstacles

- **AI agent hype:** AI agents present a competitive challenge to GenAI VAs. Most VA providers are investing in agentic AI, which is a more advanced and autonomous version of GenAI VAs. This is creating customer confusion on the capability differences between assistants versus agents.
- **Explainability:** Explainable GenAI output is key to addressing accuracy and reliability concerns. Providers are working to embed outcome explainability and performance reliability — often via combining GenAI with rule-based tech.
- **Solution performance:** Many GenAI VAs struggle to deliver strong value outcomes using general-purpose LLMs. Addressing this challenge requires significant model or data customization via model fine-tuning or context engineering requiring customer education, budget, skills, and data readiness. Domain-specialized models will also help here, but remain scarce.
- **Governance and security:** GenAI makes governance and security more challenging, and the technology to address this lags GenAI VAs in maturity.
- **ROI:** Tokenization and generative inferences are changing VAs' cost structure. Difficulty in proving ROI and high cost will slow adoption.

User Recommendations

- Define and prioritize use cases. GenAI VAs offer low-risk, measurable use cases that are ideal for piloting and scaling GenAI within the organization.
- Consider solutions that offer hybrid options — both LLM and rule-based flows — for better reliability, explainability, and compliance.
- Measure implementation success by clearly identifying the business case and supported value outcomes. Regularly assess performance to identify model accuracy issues or business misalignment.
- Improve GenAI VA performance and drive deeper automation by customizing prebuilt VAs (for different use cases, industries, or business roles), building an enterprise context layer and using model fine-tuning on enterprise data when required for that specialized support.
- Prepare for agentic automation. Start with GenAI VAs before progressing to agentic automation and the technology, business and change management challenges this emerging tech introduces.

Sample Vendors

Amazon; Google; IBM; Kore.ai; Leena AI; OneReach.ai; Openstream.ai; Orby AI; Salesforce; SoundHound AI

Gartner Recommended Reading

Emerging Tech: Market Risk Projection of Generative AI on Conversational AI

Market Guide for Conversational AI Solutions

Emerging Tech: Conversational AI Differentiation in the Era of Generative AI

Appendixes

See the previous Hype Cycle: Hype Cycle for AI in ITSM, 2025

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 2: Hype Cycle Phases

(Enlarged table in Appendix)

Phase	Definition
Innovation Trigger	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
Peak of Inflated Expectations	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers.
Trough of Disillusionment	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
Slope of Enlightenment	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
Plateau of Productivity	The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.

Table 3: Benefit Ratings

Benefit Rating	Definition
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner

Table 4: Maturity Levels

(Enlarged table in Appendix)

Maturity Levels	Status	Products/Vendors
Embryonic	In labs	None
Emerging	Commercialization by vendors Pilots and deployments by industry leaders	First generation High price Much customization
Adolescent	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization
Early mainstream	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies
Mature mainstream	Robust technology Not much evolution in vendors or technology	Several dominant vendors
Legacy	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner

Document Revision History

Hype Cycle for AI in ITSM, 2025 - 19 June 2025

Recommended by the Authors

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

Understanding Gartner's Hype Cycles

Tool: Create Your Own Hype Cycle With Gartner's Hype Cycle Builder

Hype Cycle for ITSM, 2025

Hype Cycle for Generative AI, 2026

Magic Quadrant for AI Applications in IT Service Management

AI Use-Case Assessment for IT Service Desk

Predicts 2026: AI's Transformative & Disruptive Path in ITSM

Table 1: Priority Matrix for AI in ITSM, 2026

Benefit ↓	*Years to Mainstream Adoption*			
	Less Than 2 Years ↓	*2 to 5 Years* ↓	*5 to 10 Years* ↓	*More Than 10 Years* ↓
Transformational	Foundation Models	Agentic AI Conversational User Interfaces Digital Workplace Operations Automation Platforms Large Reasoning Model Multimodal Generative AI	Agent-Centric I&O Agentic ITSM	Large Action Models
High	GenAI Virtual Assistants Retrieval-Augmented Generation	Agent Orchestration AI Agents AI TRiSM AI-Powered IT Agent Advisory Prompt and Context Engineering	AI Applications in IT Service Management AI-Augmented CMDB Emotion AI	Natural Language Case Extraction
Moderate	Event Intelligence Solutions	IT Knowledge Generation		
Low		Operations Assistant		

Source: Gartner

Table 2: Hype Cycle Phases

Phase	Definition
Innovation Trigger	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
Peak of Inflated Expectations	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the innovation is pushed to its limits. The only enterprises making money are conference organizers and content publishers.
Trough of Disillusionment	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
Slope of Enlightenment	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the innovation's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
Plateau of Productivity	The real-world benefits of the innovation are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
Years to Mainstream Adoption	The time required for the innovation to reach the Plateau of Productivity.

Source: Gartner

Table 3: Benefit Ratings

Benefit Rating	Definition
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner

Table 4: Maturity Levels

Maturity Levels	Status	Products/Vendors
Embryonic	In labs	None
Emerging	Commercialization by vendors Pilots and deployments by industry leaders	First generation High price Much customization
Adolescent	Maturing technology capabilities and process understanding Uptake beyond early adopters	Second generation Less customization
Early mainstream	Proven technology Vendors, technology and adoption rapidly evolving	Third generation More out-of-box methodologies
Mature mainstream	Robust technology Not much evolution in vendors or technology	Several dominant vendors
Legacy	Not appropriate for new developments Cost of migration constrains replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner