

# Hype Cycle for Artificial Intelligence, 2024

17 June 2024 - ID G00809438 - 126 min read

By Analyst(s): Afraz Jaffri, Haritha Khandabattu

Initiatives: [Artificial Intelligence](#); [Analytics and Artificial Intelligence](#); [Evolve Technology and Process Capabilities to Support D&A](#); [Generative AI Resource Center](#)

Investment in AI has reached a new high with a focus on generative AI, which, in most cases, has yet to deliver its anticipated business value. This research helps AI leaders identify other techniques worthy of investment by profiling a wide range of AI innovations, many of which are advancing fast.

## Analysis

### What You Need to Know

Generative AI (GenAI) has passed the Peak of Inflated Expectations, although hype about it continues. In 2024, more value will derive from projects based on other AI techniques, either stand-alone or in combination with GenAI, that have standardized processes to aid implementation. To deliver maximum benefit, AI leaders should base future system architectures on composite AI techniques by combining approaches from innovations at all stages of the Hype Cycle.

As the volume and scale of AI projects have increased, second-order effects have come into play. Increasing attention is therefore being paid to governance, risk, ownership, safety and mitigation of technical debt. These factors are being addressed at national, enterprise, team and individual practitioner levels, but, even with regulations reaching advanced stages, maturity is far from being achieved.

### The Hype Cycle

The two biggest movers on this year's Hype Cycle, AI engineering and knowledge graphs, highlight the need for means of handling AI models at scale in a robust manner. AI engineering is a fundamental requirement for delivery, at scale, of enterprise AI solutions that demand new team topologies. Knowledge graphs provide dependable logic and explainable reasoning, in contrast to the fallible, yet powerful, predictive capabilities of the deep-learning techniques used by GenAI.

Innovations at the Innovation Trigger include composite AI, AI-ready data, causal AI, decision intelligence, AI simulation and multiagent systems. These reflect the growing need to advance process and decision automation beyond single-model outputs into orchestrated multiturn composite services.

At the Peak of Inflated Expectations, responsible AI, AI TRiSM, prompt engineering and sovereign AI point to increasing concerns about the governance and safety aspects of the rapidly expanding use of AI by enterprises and individuals.

Soon to leave the peak or already in the Trough of Disillusionment are synthetic data, ModelOps, edge AI, neuromorphic computing and smart robots. These innovations still have momentum, but levels of implementation vary, and they are frequently used incorrectly or subject to inflated expectations of business value. Neuromorphic computing and smart robots have advanced significantly in the past year, indicating the potential for rapid progression through the rest of the Hype Cycle.

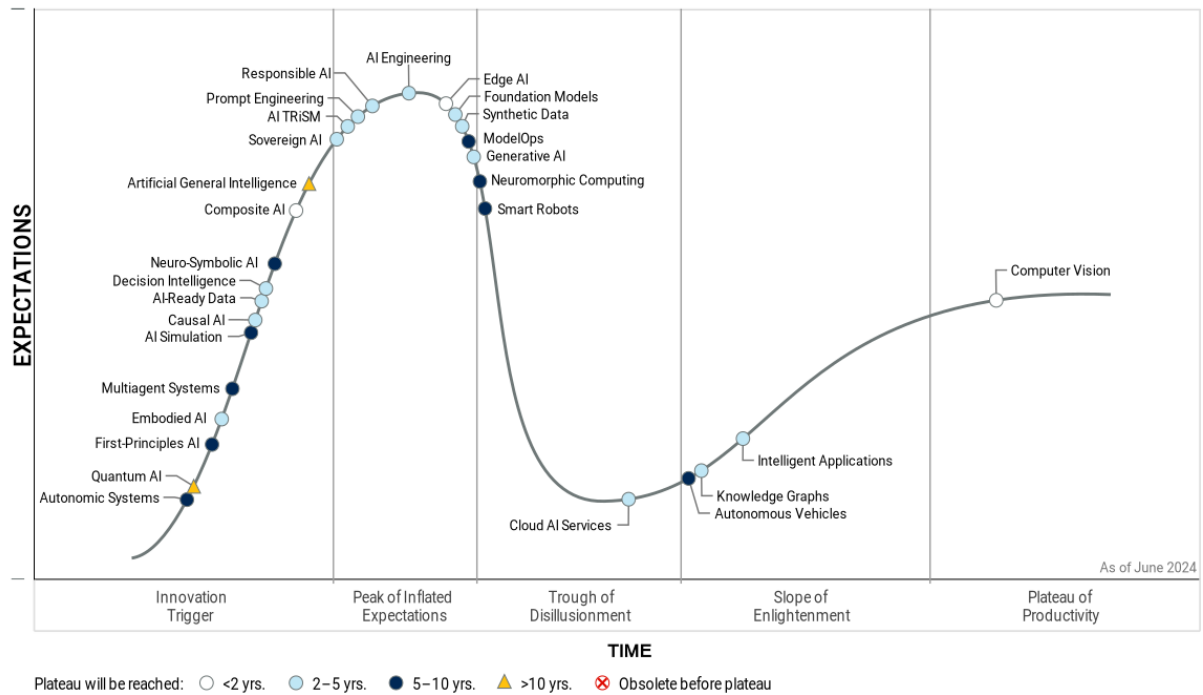
Cloud AI services have regressed on the Hype Cycle since last year, due to the number of GenAI-based cloud AI services that have come to market. Vendors and end users of these services have experienced problems with service capacity, reliability, model update frequency and cost fluctuation, which may, however, be considered growing pains.

On the Slope of Enlightenment are AI technologies that have many years of innovation behind them and are getting nearer to mainstream adoption. Usage of autonomous vehicles has increased in some locations, despite severe skepticism in certain quarters, the imposition of restrictions and the withdrawal of some operating licenses. Intelligent applications, now powered by GenAI, have entered the workforce, but more time is needed to objectively quantify their impact on productivity.

New entries on this year's Hype Cycle include quantum AI, embodied AI and sovereign AI, as companies and governments are starting to respond to the potential, and dangers, of an AI-dominated future.

Figure 1: Hype Cycle for Artificial Intelligence, 2024

## Hype Cycle for Artificial Intelligence, 2024



Gartner

## The Priority Matrix

Compared with many other Hype Cycles, this one is unusual in having so many innovations of transformational or high benefit, none of moderate benefit, and only one of low benefit.

Gartner expects that, within two years, composite AI will be the standard methodology for developing AI systems, and to be widely adopted. Another transformational innovation, computer vision, is already the subject of mass consumer adoption through smart devices.

Innovations two to five years away from mainstream adoption that merit particular attention include decision intelligence, embodied AI, foundation models, GenAI, intelligent applications and responsible AI. Early adoption of these will lead to significant competitive advantage and ease the problems associated with using AI models within business processes.

Among the innovations five to 10 years away from mainstream adoption, neuromorphic computing could open doors to novel AI architectures. An influx of new ideas and entrepreneurial ventures will be essential for further development of this technology.

AI leaders should balance strategic exploration of potentially transformative or highly beneficial innovations with investigation of innovations that do not require extensive proficiency in engineering or data science, and that have been commoditized both as stand-alone applications and as components of packaged business solutions.

Table 1: Priority Matrix for Artificial Intelligence, 2024

(Enlarged table in Appendix)

Benefit ↓	Years to Mainstream Adoption			
	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational	Composite AI Computer Vision	Decision Intelligence Embodied AI Foundation Models Generative AI Intelligent Applications Responsible AI	Autonomic Systems Autonomous Vehicles First-Principles AI Neuromorphic Computing	Artificial General Intelligence
High	Edge AI	AI Engineering AI-Ready Data AI TRISM Causal AI Cloud AI Services Knowledge Graphs Prompt Engineering Sovereign AI Synthetic Data	AI Simulation ModelOps Multiagent Systems Neuro-Symbolic AI Smart Robots	
Moderate				
Low				Quantum AI

Source: Gartner (June 2024)

Off the Hype Cycle

The following innovations have been dropped from this year’s Hype Cycle:

- Operational AI systems: Subsumed by AI engineering.
- Data labeling and annotation: Dropped because it is more relevant to the [Hype Cycle for Data Science and Machine Learning, 2024](#).
- AI maker and teaching kits: Dropped due to a lack of hype.

## On the Rise

### Autonomic Systems

Analysis By: Erick Brethenoux, Nick Jones

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

#### Definition:

Autonomic systems are self-managing physical or software systems, performing domain-bounded tasks, that exhibit three fundamental characteristics: autonomy (execute their own decisions and tasks autonomously without external assistance); learning (modify their behavior and internal operations based on experience, changing conditions or goals); and agency (have a sense of their own internal state and purpose that guides how and what they learn and enables them to act independently).

#### Why This Is Important

Autonomic systems are emerging as an important trend as they enable levels of business adaptability, flexibility and agility that can't be achieved with traditional AI techniques alone. Their flexibility is valuable in situations where the operating environment is unpredictable and real-time monitoring and control aren't practical. Their learning ability is valuable in situations where a task can be learned even though there is no well-understood algorithm (composite AI) to implement it.

#### Business Impact

Autonomic systems excel where:

- Conventional automation applying composite AI techniques is inadequate, or using fixed training data is impractical or not agile.
- It is impractical to provide real-time human guidance, or training conditions can't be anticipated.
- We cannot program the exact learning algorithm, but the task is continuously learnable.

- Continuously or rapidly changing tasks or environments make frequent retraining and testing of machine learning systems too slow or costly.

## Drivers

Autonomic systems are the culmination of a three-part trend:

- Automated systems are a very mature concept. They perform well-defined tasks and have fixed deterministic behavior (such as an assembly robot welding cars). The increasing number of use cases around automation using AI techniques is a strong base for autonomous systems.
- Autonomous systems go beyond simple automation to add independent behavior. They may exhibit some degree of adaptive behavior, but are predominantly under algorithmic control (such as self-driving cars or a Boston Dynamics' Spot robot that has its overall route and goals set by a remote human operator but has substantial local autonomy — that is, for a very specific task). Adaptive AI capabilities are a necessary foundation for autonomic systems and should accelerate the adoption of autonomic systems.
- Autonomic systems exhibit adaptive behavior through learning and self-modifying algorithms. For example, Ericsson has demonstrated the use of reinforcement learning and digital twins to create an autonomic system that dynamically optimizes 5G network performance while creating optimization rules. This trend is showing the feasibility of such systems. Early learning about carefully bounded autonomic systems will build trust in their capabilities to operate independently.

Other drivers include:

- Autonomic behavior is a spectrum. For example, chatbots learn from internet discussions; streaming services learn which content you like; and delivery robots share information about paths and obstructions to optimize fleet routes. The advantages of systems that can learn and adapt their behavior will be compelling.
- Agent-based systems are seeing an adoption renaissance fueled by the increasing complexity of existing applications and the advent of large action models.
- Substantial academic research is underway on autonemics, which will result in more widespread use.

## Obstacles

- **Nondeterminism:** Systems that continuously learn and adapt their behavior aren't predictable. This will pose challenges (such as legal) for employees and customers who may not understand how and why a system performed as it did.
- **Immaturity:** Skills in the area will be lacking until autonomics becomes more mainstream. New types of professional services may be required (like autonomous business skills).
- **Social concerns:** Misbehavior, nondeterminism or lack of understanding could generate public resistance when systems interact with people.
- **Digital ethics and safety:** Autonomic systems will require architectures and guardrails to prevent them from learning undesirable, dangerous, unethical or even illegal behavior when no human is validating the system.
- **Legal liability:** It may be difficult for the supplier of an autonomic system to take total responsibility for its behavior because that will depend on the goals it has set, its operating conditions and what it learned.

## User Recommendations

- Start by building experience with autonomous systems first to understand the constraints and requirements (legal, technical and cultural) that the organization is subjected to. Pilot autonomic technologies in cases where early adoption will deliver agility and performance benefits in software or physical systems.
- Manage risk in autonomic system deployments by analyzing the business, legal and ethical consequences of deploying autonomic systems — which are partially nondeterministic. Do so by creating a multidisciplinary task force.
- Optimize the benefits of autonomic technologies by piloting them in situations such as complex and rapidly changing environments where early adoption will deliver agility and performance benefits in either software or physical systems.

## Sample Vendors

Aspire; IBM; Latent AI; Playtika Holding; Vanti.

## Gartner Recommended Reading

[Top Strategic Technology Trends for 2022: Autonomic Systems](#)

## Quantum AI

Analysis By: Chirag Dekate, Soyeb Barot

Benefit Rating: Low

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

### Definition:

Quantum artificial intelligence is an embryonic field of research emerging at the intersection of quantum technologies and AI. Quantum AI aims to exploit unique properties of quantum mechanics to develop new and more powerful AI algorithms that deliver better than classical performance, potentially resulting in new types of AI algorithms designed to run on quantum systems.

### Why This Is Important

Quantum AI is an area of active research. Once commercialized, quantum AI could potentially help in:

- Enabling organizations to use quantum systems to address advanced AI analytics faster while using a fraction of the resources used in conventional AI supercomputing resources.
- Developing new AI algorithms that exploit quantum mechanics to deliver capabilities beyond ones that can be executed on classical systems.
- Unlocking disruptive applications that include drug discovery, energy industry and logistics.

### Business Impact

While the business impact of the embryonic quantum AI field today is low, when validated techniques mature, quantum AI will enable competitive advantage across industries; for instance:

- **Life sciences:** Transform drug discovery by shortening timelines, lowering costs and improving outcomes.
- **Finance:** Optimize portfolios, minimize risk and improve fraud detection systems.



- **Material science:** Discover new materials that revolutionize energy transportation, manufacturing and create new revenue streams.

## Drivers

- Hype around quantum technologies is driving more businesses and researchers to explore the intersection of quantum and AI.
- The accelerated pace of innovation in quantum systems (including larger volume of higher quality qubits, and greater stability and reliability of quantum systems) is driving greater interest in applicability in areas, including quantum AI.
- Access to quantum computing as a service is lowering the barrier to entry, encouraging greater collaboration among researchers and enabling exploration of new algorithms and techniques.
- Governments and enterprises globally are increasing funding for quantum (and quantum AI) research, resulting in accelerated innovation.
- The halo effect of increased hype around generative AI is driving new focus on alternative research techniques, including quantum AI, that could potentially deliver new disruptive results.
- Universities and training programs are developing programs and curricula to develop a quantum-ready workforce.

## Obstacles

- **Hardware limitations:** Current quantum systems, while getting stabler, are still error-prone and inherently noisy, limiting their utility and impact on practical quantum AI.
- **Algorithm limitations:** While several quantum AI algorithms have been proposed, very few have been vetted and proven, and they are nowhere close to being enterprise-ready.
- **Cost:** Despite their limited utility and widespread accessibility, rapidly evolving noisy intermediate-scale quantum (NISQ) systems are relatively expensive, which could inhibit research and development efforts needed to devise quantum AI algorithms.
- **Scalability of systems:** Scaling quantum systems to the level necessary for enterprise-ready quantum AI continues to be a major technical hurdle.
- **Compute paradigms:** Integrating traditional data and analytics pipelines with quantum is inherently challenging because quantum systems operate on a fundamentally different paradigm both from a data representation perspective and from a compute (non-von Neumann model) perspective.

## User Recommendations

- Prioritize investments in AI and GenAI over any quantum AI investments. Quantum AI is too nascent to warrant focused investments and unlikely to yield material gains in the next two to three years.
- Partner with local universities by sponsoring academic research as a means of derisking your quantum AI investments and create a university-to-industry talent pipeline.
- Create a quantum AI opportunity radar that enables you to track progress of underlying technologies and quantum AI algorithms, enabling you to maximize value creation as the embryonic field of quantum technologies evolves.

## Sample Vendors

Amazon; Google; IBM; IonQ; Multiverse Computing; Pasqal; SandboxAQ; Zapata AI

## Gartner Recommended Reading

[Infographic: How Use Cases Are Developed and Executed on a Quantum Computer](#)

[Predicts 2024: Emerging Defense Technology and New Domains](#)

#### First-Principles AI

**Analysis By:** Erick Brethenoux, Svetlana Sicular

**Benefit Rating:** Transformational

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Emerging

#### Definition:

First-principles AI (FPAI; aka physics-informed AI) incorporates physical and analog principles, governing laws and domain knowledge into AI models. In contrast, purely “digital” AI models do not necessarily obey the fundamental governing laws of physical systems and first principles — nor generalize well to scenarios on which they have not been trained. FPAI extends AI engineering to complex systems engineering and model-based systems (like agent-based systems).

#### Why This Is Important

As AI expands in engineering and scientific use cases, it needs a stronger ability to model problems and better represent their context. Digital-only AI solutions cannot generalize well enough beyond training, limiting their adaptability. FPAI instills a more reliable representation of the context and the physical reality, yielding more adaptive systems. This leads to reduced training time, improved data efficiency, better generalization and greater physical consistency.

#### Business Impact

Physically consistent and scientifically sound AI models can significantly improve applicability, especially in engineering use cases. FPAI helps train models with fewer data points and accelerates the training process, helping models converge faster to optimal solutions. It improves the generalizability of models to make reliable predictions for unseen scenarios, including applicability to nonstationary systems, and enhances transparency and interpretability boosting trustworthiness.

## Drivers

- **FPAI approaches instill a more flexible representation of the context and conditions in which systems operate, allowing developers to build more adaptive systems.** Traditional business modeling approaches have been brittle. This is because the digital building blocks making up solutions cannot generalize well enough beyond their initial training data, therefore limiting the adaptability of those solutions.
- **FPAI approaches provide additional physical knowledge representations, such as partial differential equations to guide or bound AI models.** Traditional AI techniques, particularly in the machine learning family, have been confronted with severe limitations — especially for causality and dependency analysis, admissible values, context flexibility and memory retention mechanisms. Asset-centric industries have already started leveraging FPAI in physical prototyping, predictive maintenance or composite materials analysis, in conjunction with augmented reality implementations.
- **Complex systems like climate models, large-scale digital twins and complex health science problems are particularly challenging to model.** Composite AI approaches provide more concrete answers and manageable solutions to these problems, but their engineering remains a significant challenge. FPAI provides more immediate answers to these problems.
- **First principles knowledge simplify and enrich AI approaches** by defining problem and solution boundaries, reducing the scope of traditionally brute force approach employed by ML; for example, known trajectories of physical objects simplify AI-enabled sky monitoring. First-principles-based semantics reveal deepfakes.
- **The need for more robust and adaptable business simulation systems will also promote the adoption of FPAI approaches.** With a better range of context modelization and more accurate knowledge representation techniques, simulations will be more reliable and account for a wider range of possible scenarios — all better anchored in reality.

## Obstacles

- The development of systematic tests and standardized evaluation for these models across benchmark datasets and problems could slow down the adoption of FPAI capabilities.
- Computationally, the scaling of the training, testing and deployment of complex FPAI models on large datasets in an efficient manner will be an issue.
- Resource-wise, collaboration across many diverse communities (physicists, mathematicians, computer scientists, statisticians, AI experts and domain scientists) will be a challenge.
- Brute force approaches are prevalent in AI, and are easy to implement for data scientists, while first principles require additional fundamental knowledge of a subject, calling for a multidisciplinary team.

## User Recommendations

- Set realistic development objectives by identifying errors that cannot be reduced and discrepancies that cannot be addressed, including data quality.
- Encourage reproducible and verifiable models starting with small-scoped problems; complex systems (in the scientific sense of the term) are generally good candidates for this approach.
- Enforce standards for testing accuracy and physical consistency for physics and first-principles-based models of the relevant domain, while characterizing sources of uncertainty.
- Promote model-consistent training for FPAI models and train models with data characteristics representative of the application, such as noise, sparsity and incompleteness.
- Quantify generalizability in terms of how performance degrades with degree of extrapolation to unseen initial and boundary conditions and scenarios.
- Ensure relevant roles and education in a multidisciplinary AI team (with domain expertise), so that the team can develop effective and verifiable solutions.

## Sample Vendors

Abzu; IntelliSense.io; MathWorks; NNAISENSE; NVIDIA; VERSES

## Gartner Recommended Reading

[Innovation Insight: AI Simulation](#)

[Innovation Insight for Composite AI](#)

[Go Beyond Machine Learning and Leverage Other AI Approaches](#)

[Innovation Insight: Causal AI](#)

[Predicts 2023: Simulation Combined With Advanced AI Techniques Will Drive Future AI Investments](#)

## Embodied AI

Analysis By: Pieter den Hamer

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

### Definition:

Embodied AI is based on the view that intelligence and embodiment in a certain context are inextricably linked — one shapes the other. It is an approach where a physical or virtual AI agent's models are trained and co-engineered with its user interface, sensors, appearance, actuators or other capabilities required to interact with a specific, real or simulated environment. This enables robust, reliable and adaptive execution of intelligent tasks.

### Why This Is Important

Embodied AI aims to create AI agents that can act autonomously or to augment humans in practical, dynamic contexts — much more so than current AI, including abstract large language models with limited reliability and effectiveness in decision-making and action-taking. This is achieved through active perception and adaptive behavior, orchestrated by an AI agent's intelligence that is in symbiosis with the capabilities and constraints of the AI agent's host or body in a certain environment.

## Business Impact

Embodied AI will further value creation with AI across various use cases. Particularly where there is a need for more practical know-how, a better representation of the physical, social or other characteristics of its environment, and a greater resilience to deal with unexpected or disruptive events. Example use cases include virtual assistants, avatars, gaming characters, autonomous vehicles and smart robots. This will pave the way toward more effective and trusted AI and more game-changing use of AI to enable new products, services and business models.

## Drivers

- The advent of generative AI (GenAI) has catalyzed AI adoption in general. Yet it has also highlighted the limitations of current AI, particularly with respect to reliability and the challenges with contextualization and grounding of AI in reality.
- Embodied AI benefits from advances in compute power and GenAI to support realistic simulations with reinforcement learning for adaptive behavior training. This also supports approaches to co-evolving baseline versions of both embodiment and intelligence of AI agents, before further improving and deploying them in the real world.
- Embodied AI is enabled by emerging approaches such as physics-informed or first principles AI (representing among others, the laws of physics or engineering heuristics), composite AI (for example, using neuro-symbolic AI for spatiotemporal reasoning) and causal AI (representing cause-and-effect relations).
- The interest for embodied AI is further fueled by innovations in areas such as virtual/augmented/mixed reality, gaming, smart robotics, autonomous systems, natural language generation and emotion AI. All of which are related to the improved design of AI agents, both physical and virtual. Physical agents also benefit from advances in sensor technology, robotics engineering, and, for example, new materials for more natural mechanics and haptic interfaces.
- Advances in embodied AI are underpinned by evolving scientific insights about intelligence, which is no longer seen as a centralized brain-only concept. Cognitive traits like perception, emotion, reasoning and behavior are actually distributed and co-evolved in multiple parts of the body. This also aligns well with distributed AI system architectures, including multiagent systems.
- Embodied AI is seen as a critical step toward possible future artificial general intelligence as it is inseparable from its operational entity that interacts with its environment. This means it is not abstracted from but grounded in reality by design, holding the promise of providing intrinsic meaning or semantics to its knowledge representations and 'native' common sense.



## Obstacles

- The world is a very complex, unpredictable and even chaotic place. Which is why the development of realistic simulations, effective robotics and – for example – truly autonomous cars has proven to be elusive.
- Real-world interaction requires real-time, highly responsive AI, even with limited energy and compute resources (for example, on mobile or edge devices). However, more lightweight and energy-efficient AI are not easily achievable.
- Embodied AI holds the promise of even more powerful and autonomous AI. Unfortunately, this may not only facilitate benevolent but also malevolent use. Effective regulation and risk management for responsible AI are, however, not a given.
- AI embodiments can be unnecessarily anthropomorphic in their design (a body with two legs and two arms), bringing in additional complexity and challenges.
- Embodied AI requires multidisciplinary collaboration between experts in areas as diverse as machine learning, graphical design, mechanical engineering and still others, depending on the use case and type of AI agent.

## User Recommendations

- Identify use cases that may benefit from applying embodied AI, both in more virtual domains, such as online customer interaction or knowledge worker augmentation, and in more physical domains, such as manufacturing or logistics.
- Explore the value that embodied AI can add by reducing the limitations of current AI in terms of better interpretation of, for example, physical constraints in a warehouse or cultural norms in client interaction. This may result in increased safety or decreased bias in the use of AI, respectively.
- Extend the mindset of how AI agents should be developed or trained. Move from a modeling-only approach toward one that considers how intelligence can be a synergy between AI models and the design of the agent's embodiment. This could, for example, relate to the facial expression of virtual agents, or the coordination of movement in physical agents.

## Sample Vendors

Amazon; Figure; Google; Hanson Robotics; Intel; Intrinsic; NNAISENSE; Qualcomm; Tesla

## Gartner Recommended Reading

[Innovation Insight: AI Agents](#)

[Innovation Insight: AI Simulation](#)

[Hype Cycle for Mobile Robots and Drones, 2023](#)

[Building a Digital Future: The Metaverse](#)

[Emerging Technologies: Introducing the Artificial Intelligence Roadmap for Virtual Assistants](#)

## Multiagent Systems

Analysis By: Leinar Ramos, Pieter den Hamer, Anthony Mullen

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

### Definition:

A multiagent system (MAS) is a type of AI system composed of multiple, independent (but interactive) agents, each capable of perceiving their environment and taking actions. Agents can be AI models, software programs, robots and other computational entities. Multiple agents can work toward a common goal that goes beyond the ability of individual agents, with increased adaptability and robustness.

### Why This Is Important

Current AI is focused on the creation of individual agents built for specific use cases, limiting the potential business value of AI to simpler problems that can be solved by single monolithic models. The combined application of multiple autonomous agents can tackle complex tasks that individual agents cannot, while creating more adaptable, scalable and robust solutions. It is also able to succeed in environments where decentralized decision making is required.

### Business Impact

Multiagent systems can be used in:

- **Generative AI:** Orchestrating AI agents for complex tasks
- **Robotics:** Swarms of robots and drones for warehouse optimization, search and rescue, environment monitoring, and other use cases
- **Energy and utilities:** Smart grid optimization and load balancing
- **Supply chain:** Optimizing scheduling, planning, routing and supply chain optimization
- **Telecom:** Network optimization and fault detection
- **Healthcare:** Using agents to model actors (individuals, households, professionals)

## Drivers

- **Generative AI agents:** Large language models (LLMs) are increasingly augmented with additional capabilities, such as an internal memory and plug-ins to external applications, to implement AI agents. An emerging design pattern is to assemble and combine these LLM-based AI agents into more powerful systems, which is increasing the feasibility of and interest in multiagent systems.
- **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. The decentralized nature of multiagent systems makes them more resilient and adaptable to complex decision making.
- **Simulation and multiagent reinforcement learning:** Advances in the realism and performance of simulation engines, as well as the use of new multiagent reinforcement learning techniques, allow for the training of multiagent AI systems in simulation environments, which can then be deployed in the real world.

## Obstacles

- **Training complexity:** Multiagent systems are typically harder to train and build than individual AI agents. These systems can exhibit emergent behavior that is hard to predict in advance, which increases the need for robust training and testing.
- **Monitoring and governing multiple agents:** Coordination and collaboration between agents is challenging. Careful monitoring, governance and a common grounding are required to ensure that the combined multiagent system behavior achieves its intended goals.
- **Limited adoption and readiness:** Despite its benefits, the application of multiagent systems to real-world problems is not yet widespread, which creates a lack of enterprise awareness and readiness to implement.
- **Specialized skills required:** Building and deploying multiagent systems requires specialized skills beyond traditional AI skills, particularly the use of reinforcement learning and simulation.
- **Fragmented vendor landscape:** A fragmented vendor landscape inhibits customer adoption and engagement.

## User Recommendations

- Use multiagent systems for complex problems that require decentralized decision making and cannot be solved by single AI agents. This includes problems with changing environments where agents need to adapt and problems where a diverse set of agents with different expertise can be combined to accomplish a goal.
- Shift to a multiagent approach gradually since this is an emerging area of research and the risks and benefits are not yet fully understood.
- Establish clear guardrails when implementing multiagent systems, including legal and ethical guidelines around autonomy, liability, robust security measures and data privacy protocols.
- Invest in the use of simulation technologies for AI training, as simulation is the primary environment to build and test multiagent systems.
- Educate your AI teams on multiagent systems, how they differ from single-agent AI design, and some of the available techniques to train and build these systems.

## Sample Vendors

Alphabet; Ansys; Cosmo Tech; FLAME GPU; MathWorks; Microsoft; OpenAI; The AnyLogic Company

## Gartner Recommended Reading

[Innovation Insight: AI Agents](#)

[Innovation Insight: AI Simulation](#)

[AI Design Patterns for Large Language Models](#)

## AI Simulation

**Analysis By:** Leinar Ramos, Anthony Mullen, Pieter den Hamer, Jim Hare

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Emerging

### Definition:

AI simulation is the combined application of AI and simulation technologies to jointly develop AI agents and the simulated environments in which they can be trained, tested and sometimes deployed. It includes both the use of AI to make simulations more efficient and useful, and the use of a wide range of simulation models to develop more versatile and adaptive AI systems.

### Why This Is Important

Increased complexity in decision making is driving demand for both AI and simulation. However, current AI faces challenges, as it is brittle to change and usually requires a lot of data. Conversely, realistic simulations can be expensive and difficult to build and run. To resolve these challenges, a growing approach is to combine AI and simulation: Simulation is used to make AI more robust and compensate for a lack of training data, and AI is used to make simulations more efficient and realistic.

### Business Impact

AI simulation can bring:

- Increased value by broadening AI use to cases where data is scarce, using simulation to generate synthetic data (for example, synthetic data for generative AI [GenAI])
- Greater efficiency by leveraging AI to decrease the time and cost to create and use complex and realistic simulations
- Greater robustness by using simulation to generate diverse scenarios, increasing AI performance in uncertain environments
- Decreased technical debt by reusing simulation environments to train future AI models

## Drivers

- **Limited availability of AI training data is increasing the need for synthetic data techniques, such as simulation.** Simulation techniques, like physics-based 3D simulation, are uniquely positioned to generate diverse AI training datasets. This is increasingly important for GenAI as training data becomes more scarce.
- **Advances in capabilities are making simulation increasingly useful for AI.** Simulation capabilities have been rapidly improving, driven both by increased computing performance and more-efficient techniques.
- **The growing complexity of decision making is increasing interest in AI simulation.** Simulation is able to generate diverse “corner case” scenarios that do not appear frequently in real-world data, but that are still crucial to train and test AI so it can perform well in uncertain environments.
- **Increased technical debt in AI is driving the need for the reusable environments that simulation provides.** Organizations will increasingly deploy hundreds of AI models, which requires a shift in focus toward building persistent, reusable environments where many AI models can be trained, customized and validated. Simulation environments are ideal since they are reusable, scalable and enable the training of many AI models at once.
- **The growing sophistication of simulation drives the use of AI, making it more efficient.** Modern simulations are resource intensive. This is driving the use of AI to accelerate simulation, typically by employing AI models that can replace parts of the simulation without running resource-intensive, step-by-step numerical computations.
- **Research in learned simulations (known as “world models”) is driving interest in AI simulation:** Research is increasing on training world models that can learn to predict how the environment will evolve, based on its current state and agents’ actions. These learned simulations could make AI simulation more feasible by not having to directly specify simulation parameters.

## Obstacles

- **Gap between simulation and reality:** Simulations can only emulate — not fully replicate — real-world systems. This gap will reduce as simulation capabilities improve, but it will remain a key factor. Given this gap, AI models trained in simulation might not have the same performance once they are deployed; differences in the simulation training dataset and real-world data can impact models' accuracy.
- **Complexity of AI simulation pipelines:** The combination of AI and simulation techniques can result in more-complex pipelines that are harder to test, validate, maintain and troubleshoot.
- **Limited readiness to adopt AI simulation:** A lack of awareness among AI practitioners about leveraging simulation capabilities can prevent organizations from implementing an AI simulation approach.
- **Fragmented vendor market:** The AI and simulation markets are fragmented, with few vendors offering combined AI simulation solutions, potentially slowing down the deployment of this capability.

## User Recommendations

- Complement AI with simulation to optimize business decision making or to overcome a lack of real-world data by offering a simulated environment for synthetic data generation or reinforcement learning.
- Complement simulation with AI by applying deep learning to accelerate simulation, and generative AI to augment simulation.
- Create synergies between AI and simulation teams, projects and solutions to enable a new generation of more-adaptive solutions for ever-more-complex use cases. Incrementally build a common foundation of more-generalized and complementary models that are reused across different use cases, business circumstances and ecosystems.
- Prepare for the combined use of AI, simulation and other relevant techniques —such as graphs, natural language processing or geospatial analytics — by prioritizing vendors that offer platforms that integrate different AI techniques (composite AI), as well as simulation.



## Sample Vendors

Altair; Ansys; The AnyLogic Company; Cosmo Tech; Epic Games; MathWorks; Microsoft; NVIDIA; Rockwell Automation; Unity

## Gartner Recommended Reading

[Innovation Insight: AI Simulation](#)

[Predicts 2023: Simulation Combined With Advanced AI Techniques Will Drive Future AI Investments](#)

## Causal AI

Analysis By: Pieter den Hamer, Ben Yan, Leinar Ramos

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

### Definition:

Causal AI identifies and utilizes cause-and-effect relationships to go beyond correlation-based predictive models and toward AI systems that can prescribe actions more effectively and act more autonomously. It includes different techniques, such as causal graphs and simulation, that help uncover causal relationships to improve decision making.

### Why This Is Important

AI's ultimate value comes from making better decisions and taking effective actions. However, the current correlation-based approach has its limitations. It may be fine for prediction, assuming that past and future do not deviate too much, but predicting an outcome is not the same as understanding what causes it and how to improve it. Causal AI is crucial when we need to be more robust in forecasting and more prescriptive to determine the best actions to influence specific outcomes.

### Business Impact

Causal AI leads to:

- Greater decision augmentation and autonomy in AI systems by estimating intervention effects.
- Greater efficiency by adding domain knowledge to bootstrap AI models with smaller datasets.
- Better explainability by capturing easy-to-interpret cause-and-effect relationships.
- More robustness and adaptability by leveraging causal relationships that remain valid in changing environments.
- The ability to extract causal knowledge with less costly and time-consuming experiments.
- Reduced bias in AI systems by making causal links more explicit.

## Drivers

- **Analytics demand is shifting from predictive to more prescriptive capabilities.** Making accurate predictions will remain key, but a causal understanding of how to affect predicted outcomes is increasingly important.
- **AI systems increasingly need to act autonomously,** particularly for time-sensitive and complex use cases where human intervention is not feasible. This will only be possible by AI understanding what impact actions will have and how to make effective interventions.
- **Limited data availability for certain use cases require more data-efficient techniques like causal AI.** Causal AI leverages human domain knowledge of cause-and-effect relationships to bootstrap AI models in small-data situations.
- **Growing complexity and dynamics of business requires more robust AI techniques.** The volatility of the last few years has exposed the brittleness of correlation-based AI models across industries. Causal structure changes much more slowly than statistical correlations, making causal AI more robust and adaptable in fast-changing environments.
- **The need for greater AI trust and explainability is driving interest in models that are more intuitive to humans.** Causal AI techniques, such as causal graphs, make it possible to be explicit about causes and explain models in terms that humans understand.
- **Generative AI (GenAI) can accelerate causal AI implementation.** GenAI is emerging as an aid to explore documents and other data sources for existing causal knowledge. This can then be used to generate candidate causal graphs, which, while still requiring human validation or completion, may reduce time-consuming manual work.
- **The next step in AI requires causal AI.** Current deep learning models and, in particular, GenAI, have limitations in terms of their reliability and ability to reason. A composite AI approach that complements GenAI with causal AI — in particular, causal knowledge graphs — offers a promising avenue to bring AI to a higher level.

## Obstacles

- **Causality is not trivial.** Not every phenomenon is easy to model in terms of its causes and effects. Causality might be unknown, regardless of AI use.
- **The quality of a causal AI model depends on its causal assumptions and on the data used to build it.** This data is susceptible to bias and imbalance, and may be incomplete in terms of representing all causal factors, known or unknown.
- **Causal AI requires technical and domain expertise to properly estimate causal effects.** Building causal AI models is often more difficult than building correlation-based predictive models, requiring active collaboration between domain experts and AI experts.
- **AI experts might be unaware of causality methods.** If AI experts are overly reliant on data-driven models like machine learning (ML), organizations could get pushback when looking to implement causal AI.
- **The vendor landscape is nascent, and enterprise adoption is currently low.** This represents a challenge when organizations run initial causal AI pilots and identify specific use cases where causal AI is most relevant.

## User Recommendations

- Acknowledge the limitations of correlation-based AI and ML approaches which focus on leveraging correlations and mostly ignore causality. These limitations also apply to most GenAI techniques, including foundation models and large language models.
- Use causal AI when you require more augmentation and automation in decision intelligence, that is, when AI is needed not only to generate predictions but also to understand how to affect the predicted outcomes. Examples include customer retention programs, marketing campaign allocation and financial portfolio optimization, as well as smart robotics and autonomous systems.
- Select different causal AI techniques depending on the complexity of the specific use case. These include causal rules, causal graphs and Bayesian networks, simulation, and ML for causal learning.
- Educate your data science teams on causal AI. Explain the difference between causal and correlation-based AI, and cover the range of techniques available to incorporate causality.

## Sample Vendors

Actable AI; causaLens; Causality Link; Geminis Software; IBM; Microsoft; Parabole.ai; Scalnyx; Vizuro; Xplain Data

## Gartner Recommended Reading

[Innovation Insight: Causal AI](#)

[Innovation Insight for Composite AI](#)

[Innovation Insight for Decision Intelligence Platforms](#)

[Case Study: Causal AI to Maximize the Efficiency of Business Investments \(HDFC Bank\)](#)

## AI-Ready Data

Analysis By: Roxane Edjlali, Svetlana Sicular, Mark Beyer

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Embryonic

### Definition:

The ability to prove the fitness of data for the specific AI use case determines if data is AI-ready data. Proof of readiness comes from the ability to continuously meet AI requirements by assessing its alignment to the use case, enabling data qualification and ensuring data governance. As a result, AI-ready data can only be determined contextually to the use case and AI technique used, which forces new approaches to data management.

### Why This Is Important

With the rise of pretrained off-the-shelf models and hype from generative AI, organizations and their data management leaders are at the forefront of creating data strategies for AI to ensure that their data is ready to serve AI and underpin data-driven applications. Chief data and analytics officers and data management leaders need to be able to quickly respond to AI-ready data demands. It all starts by delivering AI-ready data to support AI use cases.

## Business Impact

The ability to deliver AI-ready data to support enterprises' AI strategies will be critical to delivering on the business value of AI. As a result, all organizations that invest in AI at scale will need to evolve their data management practices and capabilities not only to preserve the evergreen classical ideas of data management but also to extend them to AI. It will be critical to provision AI-ready data iteratively to cater to existing and upcoming demands of the business, ensure trust, preserve intellectual property (IP) and reduce bias and hallucinations.

## Drivers

- Models, especially for generative AI, increasingly come from the vendors rather than being delivered in-house. Data is becoming the main means for enterprises to get value from these pretrained models.
- Most commonly delivered AI solutions depend on data availability, quality and understanding, not just AI model building. Many enterprises attempt to tackle AI without considering AI-specific data management issues. The importance of data management in AI is often underestimated, so data management solutions must now be adjusted for AI needs.
- Classical data management is ripe for disruption to support AI efforts. Rapid progress of AI poses new challenges in organizing and managing the data for AI. We expect a cycle of augmented data management techniques that are better suited for meeting the data requirements of AI. Data ecosystems on the foundation of data fabric indicate the beginning of this new cycle.
- Data management capabilities, practices and tools greatly benefit AI development and deployment. The AI community invents new data-centric approaches, such as federated machine learning (ML) and retrieval-augmented generation (RAG), which benefit from data management innovations like data fabric and lakehouse. For example, implementing a knowledge graph as part of the data fabric allows for a large-language-model-led query to benefit from the context provided by the graph, which will increase the accuracy of the code generated.
- New data management solutions mitigate AI-amplified bias originating in data interpretation, labeling and human actions recorded in the data. Bias mitigation and hallucinations are an acute, AI-specific problem that forces data management to determine how to structure, analyze and prepare data.
- Generative AI is removing the distinction between structured and unstructured data, thereby requiring data management to adapt to these new uses.

## Obstacles

- AI is disconnected from data management. The AI community remains mostly unaware of data management capabilities, practices and tools that can greatly benefit AI development and deployment, which can lead to challenges when scaling prototypes in production. Traditional data management also ignores the AI-specific considerations, such as data bias, labeling and drift; this is changing, but slowly.
- Even though the data side of AI is essential, it is underestimated. It includes tasks such as preparing datasets and developing a clear understanding of why the data was collected a certain way, what the data means and what biases exist in the data.
- Responsible AI requires new governance approaches of both the data and AI model. These are AI-specific data practices that many enterprises want to solve through tooling rather than governance.
- Data management activities don't end once the model has been developed. Deployment considerations and ongoing drift monitoring require dedicated data management activities and practices.

## User Recommendations

- Formalize AI-ready data as part of your data management strategy. Implement active metadata management, data observability and data fabric as foundational components of this strategy. Combine foundational and new capabilities to meet AI needs. Establish roles and responsibilities to manage data in support of AI.
- Approach AI model development in a data-centric way due to the dependency of AI models on representative data. Diversify data, models and people to ensure AI value and ethics.
- Utilize data management expertise, AI engineering, DataOps and MLOps approaches to support the AI life cycle. Include data management requirements when deploying models. Develop data monitoring and data governance metrics to ensure that your AI models produce the correct output continuously.
- Enforce policies on data fitness for AI. Define and measure minimum data standards for AI readiness of data early on for each use case and continuously prove data fitness when taking AI to scale. These include checking lineage, quality and governance assessment, versioning and automated testing.
- Investigate those data management tools that are rich in augmented data management capabilities and can integrate well with the AI tools that have created disruptive data-centric AI capabilities like federated ML and RAG.

## Sample Vendors

Databricks; Explorium; Google; illumex; Landing AI; LatticeFlow; Microsoft; MOSTLY AI; Protopia AI; YData

## Gartner Recommended Reading

[Quick Answer: What Makes Data AI-Ready?](#)

[Innovation Insight: How Generative AI Is Transforming Data Management Solutions](#)

[Quick Answer: Options for Using Your Data With Generative AI Models](#)

[Successful Generative AI Projects Require Better Metadata Management](#)

[When Not to Use Generative AI](#)



## Decision Intelligence

Analysis By: Erick Brethenoux, David Pidsley, Pieter den Hamer

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

### Definition:

Decision intelligence (DI) is a practical discipline that advances decision making by explicitly understanding and engineering how decisions are made, and how outcomes are evaluated, managed and improved via feedback.

### Why This Is Important

The current hype around automated decision making and augmented intelligence, fueled by AI techniques in decision making (including generative AI [GenAI]), is pushing DI toward the Peak of Inflated Expectations. Recent crises have revealed the brittleness of business processes. Reengineering those processes to be resilient, adaptable and flexible will require the discipline of methods and techniques. The fast-emerging decision intelligence platforms (DIPs) market is starting to provide resilient solutions for decision makers.

### Business Impact

- **DI provides better, more timely and optimized decision making by making decisions explicit and transparent.** It reduces the unpredictability of decision outcomes by capturing the business context.
- **DI reduces technical debt and increases visibility.** It improves the impact of business processes by materially enhancing the consistency of decision models based on the power of their relevance and the quality of their transparency, making decisions transparent and auditable.

## Drivers

- **A dynamic and complex business environment with an increasingly unpredictable and uncertain pace of business:** Two forces are creating a new market around DIPs. First is the combination of AI techniques such as rules, knowledge graphs and machine learning (ML). Second is the confluence of technology clusters around composite AI, smart business processes, real-time event processing, insight engines, decision management and advanced personalization platforms.
- **The need to curtail unstructured, ad hoc decisions that are siloed and disjointed:** Often uncoordinated, such decisions promote local optimizations at the expense of global efficiency.
- **Expanding collaboration between humans and machines:** This collaboration, supplemented by a lack of trust in technologies, is increasingly replacing tasks and promoting uneasiness from a human perspective. DI practices promote transparency, interpretability, fairness, reliability and accountability of decision models, critical for the adoption of business-differentiating techniques.
- **Tighter regulations making risk management more prevalent:** From privacy and ethical guidelines to new laws and government mandates, organizations are facing difficulty in understanding the risk impacts of their decisions. DI promotes explicit decision models, reducing the risk.
- **Uncertainty regarding decision consistency across the organization:** Lack of explicit representation of decisions prevents proper harmonization of collective decision outcomes; DI remedies this issue.
- **Emergence of software tools in the form of DIPs:** DIPs will enable organizations to practically implement DI projects and strategies.
- **GenAI and its synergy with existing DI techniques and practices:** The advent of GenAI offers more efficient and richer context to decision making. It accelerates the research and adoption of composite AI models, which are the foundation of DIPs.

## Obstacles

- **Lack of proper coordination between business units:** The inability to impartially reconsider critical decision flows within and across departments diminishes the effectiveness of early DI efforts.
- **Fragmentation:** Decision-making silos have created data, competencies and technology clusters that are difficult to reconcile and that could slow down the implementation of decision models.
- **Subpar operational structure:** An inadequate organizational structure around advanced techniques, such as the lack of an AI center of excellence, could impair DI progress.
- **Lack of modeling in a wider context:** In organizations that have focused almost exclusively on technical skills, the other critical parts of human decision making — psychological, social, economic and organizational factors — have gone unaddressed.
- **Lack of AI literacy:** Many organizations still suffer from a lack of understanding of AI techniques. This AI illiteracy could slow down the development of DI projects.

## User Recommendations

- **Promote the resilience and sustainability of cross-organizational decisions** by building models using principles to enhance traceability, replicability, pertinence and trustworthiness.
- **Improve the predictability and alignment of decision agents** by simulating their collective behavior while also estimating their global contribution versus local optimization.
- **Develop staff expertise** in traditional and emerging decision augmentation and decision automation techniques, including predictive and prescriptive (optimization and business rules) analytics. Upskill business analysts, and develop new roles such as decision engineer and decision steward.
- **Tailor the choice of decision-making technique** to the particular requirements of each decision situation by collaborating with subject matter experts, AI experts and business process analysts.
- **Accelerate the development of DI projects** by encouraging experimentation with GenAI and expediting the deployment of composite AI solutions.

## Gartner Recommended Reading

[Innovation Insight for Decision Intelligence Platforms](#)

[Predicts 2024: How Artificial Intelligence Will Impact Analytics Users](#)

[Reengineer Your Decision-Making Processes for More Relevant, Transparent and Resilient Outcomes](#)

[Emerging Tech: Venture Capital Growth Insights for Decision Intelligence Platforms](#)

[Video: How Decision Intelligence Improves Business Outcomes](#)

## Neuro-Symbolic AI

Analysis By: Erick Brethenoux, Afraz Jaffri

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

### Definition:

Neuro-symbolic AI is a form of composite AI that combines machine learning (ML) methods and symbolic systems (for example, knowledge graphs) to create more robust and trustworthy AI models. This fusion enables the combination of probabilistic models with explicitly defined rules and knowledge to give AI systems the ability to better represent, reason and generalize concepts. This approach provides a reasoning infrastructure for solving a wider range of business problems more effectively.

### Why This Is Important

Neuro-symbolic AI is important because it addresses limitations in current AI systems, such as incorrect outputs, lack of generalization to a variety of tasks and an inability to explain the steps that led to an output. This leads to more powerful, versatile and interpretable AI solutions and allows AI systems to tackle more complex tasks with humanlike reasoning.

## Business Impact

Neuro-symbolic AI will have an impact on the efficiency, adaptability and reliability of AI systems used across business processes. The integration of logic and multiple reasoning mechanisms brings down the need for ever larger AI models and their supporting infrastructure. These systems will rely less on the processing of huge amounts of data, making AI agile and resilient. Neuro-symbolic approaches can augment and automate decision making with less risk of unintended consequences.

## Drivers

- Limitations of AI models based exclusively on ML techniques that focus on correlation over understanding and reasoning. The newest generation of large language models is well-known for its tendency to give factually incorrect answers or produce unexpected results. Neuro-symbolic AI addresses these limitations.
- The need for explanation and interpretability of AI outputs that are especially important in the regulated industry use cases and in systems that use private data.
- The need to prioritize understanding the meanings behind words, not just their arrangement (semantics over syntax) in systems that deal with real-world entities to ground meaning to words and terms in specific domains.
- The set of tools available to combine different types of AI models is increasing and becoming easier to use for developers, data scientists and end users. The dominant approach is to chain together results from different models (composite AI) rather than using single models.
- The integration of multiple reasoning mechanisms necessary to provide agile AI systems eventually leads to adaptive AI systems, notably through blackboardlike mechanisms.

## Obstacles

- Most neuro-symbolic AI methods and techniques are being developed in academia or industry research labs. Despite the increased availability of tools, there are still limited implementations in business or enterprise settings.
- There are no agreed-upon techniques for implementing neuro-symbolic AI, and disagreements continue between researchers and practitioners on the effectiveness of combining approaches, despite the emergence of real-world use cases.
- The commercial and investment trajectories for AI startups allocate almost all capital to deep learning approaches, leaving only those willing to bet on the future to invest in neuro-symbolic AI development.
- Currently, popular media and academic conferences do not give as much exposure to the neuro-symbolic AI movement as compared to other approaches.

## User Recommendations

- Adopt composite AI approaches when building AI systems by utilizing a range of techniques that increase the robustness and reliability of AI models. Neuro-symbolic AI approaches will fit into a composite AI architecture.
- Dedicate time to learning and applying neuro-symbolic AI approaches by identifying use cases that can benefit from these approaches.
- Invest in data architecture that can leverage the building blocks for neuro-symbolic AI techniques, such as knowledge graphs and agent-based techniques.
- Consider neuro-symbolic AI architectures when the limitations of generative AI models prevent their implementation in the organization.

## Sample Vendors

Elemental Cognition; Franz; Google DeepMind; IBM; Microsoft; RelationalAI; Wolfram|Alpha

## Gartner Recommended Reading

[Innovation Insight: AI Simulation](#)

[Go Beyond Machine Learning and Leverage Other AI Approaches](#)

[When Not to Use Generative AI](#)

## Predicts 2023: Simulation Combined With Advanced AI Techniques Will Drive Future AI Investments

### Composite AI

Analysis By: Erick Brethenoux, Pieter den Hamer

**Benefit Rating:** Transformational

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

#### Definition:

Composite AI refers to the combined application (or fusion) of different AI techniques to improve the efficiency of learning to broaden the level of knowledge representations. It broadens AI abstraction mechanisms and, ultimately, provides a platform to solve a wider range of business problems effectively.

#### Why This Is Important

Composite AI recognizes that no single AI technique is a panacea. It aims to combine “connectionist” AI approaches, like machine learning (ML) and deep learning, with “symbolic” and other AI approaches, like rule-based reasoning, graph analysis or optimization techniques. The goal is to enable AI solutions that require less data and energy to learn, embodying more abstraction mechanisms. Composite AI is at the center of the generative AI (GenAI) and decision intelligence (DI) markets.

#### Business Impact

Composite AI brings the power of AI to a broader group of organizations that do not have access to large amounts of historical or labeled data but possess significant human expertise. It helps to expand the scope and quality of AI applications (that is, more types of reasoning challenges). Other benefits include better interpretability and embedded resilience and the support of augmented intelligence. The new wave of GenAI implementations heavily relies on composite AI.

## Drivers

- The growing reliance on AI for decision making is driving organizations toward composite AI. The most appropriate actions can be further determined by combining rule-based and optimization models — a combination often referred to as prescriptive analytics.
- Small datasets, or the limited availability of data, have pushed organizations to combine multiple AI techniques. Enterprises have started to complement scarce raw historical data with additional AI techniques, such as knowledge graphs and generative adversarial networks (GANs), to generate synthetic data.
- Combining AI techniques is much more effective than relying only on heuristics or a fully data-driven approach. A heuristic or rule-based approach can be combined with a deep learning model (for example, predictive maintenance). Rules coming from human experts, or the application of physical/engineering model analysis, may specify that certain sensor readings indicate inefficient asset operations. This can be used as a feature to train a neural network to assess and predict the asset's health, also integrating causal AI capabilities.
- The proliferation of computer vision and natural language processing (NLP) solutions is used for identifying or categorizing people or objects in an image. This output can be used to enrich or generate a graph, representing the image entities and their relationships.
- Agent-based modeling is the next wave of composite AI. A composite AI solution is composed of multiple agents, each representing an actor in the ecosystem. Combining these agents into a “swarm” enables the creation of common situation awareness, more global planning optimization, responsive scheduling and process resilience.
- The advent of GenAI is accelerating the research and adoption of composite AI models through artifacts, process and collaboration generations, which are the foundation of DI platforms.



## Obstacles

- **Lack of awareness and skills in leveraging multiple AI methods:** This could prevent organizations from considering the techniques particularly suited to solving specific problem types.
- **Deploying ModelOps:** The ModelOps domain (that is, the operationalization of multiple AI models, such as optimization models, rule models and graph models) remains an art much more than a science. A robust ModelOps approach is required to efficiently govern composite AI environments and harmonize it with other disciplines, such as DevOps and DataOps.
- **Trust and risk barriers:** The AI engineering discipline is starting to take shape, but only mature organizations apply its benefits in operationalizing AI techniques. Security, ethical model behaviors, observability, model autonomy and change management practices must be addressed across the combined AI techniques.

## User Recommendations

- Identify projects in which a fully data-driven, ML-only approach is inefficient or ill-fitted. For example, in cases when enough data is not available or when the pattern cannot be represented through current ML models.
- Capture domain knowledge and human expertise to provide context for data-driven insights by applying decision management with business rules and knowledge graphs, in conjunction with ML and/or causal models.
- Combine the power of ML, image recognition or NLP with graph analytics to add higher-level, symbolic and relational intelligence.
- Extend the skills of ML experts, or recruit/upskill additional AI experts, to cover graph analytics, optimization or other techniques for composite AI. For rules and heuristics, consider knowledge engineering skills, as well as emerging skills such as prompt engineering.
- Accelerate the development of DI projects by encouraging experimentation with GenAI, which will in turn accelerate the deployment of composite AI solutions.

## Sample Vendors

ACTICO; Aera Technology; FICO; Frontline Systems; IBM; Indico Data; Peak; SAS

## Gartner Recommended Reading

[Go Beyond Machine Learning and Leverage Other AI Approaches](#)

[When Not to Use Generative AI](#)

[Top Strategic Technology Trends for 2022: AI Engineering](#)

[How to Choose Your Best-Fit Decision Management Suite Vendor](#)

## Artificial General Intelligence

Analysis By: Pieter den Hamer

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Embryonic

### Definition:

Artificial general intelligence (AGI), also known as strong AI, is the (currently hypothetical) intelligence of a machine that can accomplish any intellectual task that a human can perform. AGI is a trait attributed to future autonomous AI systems that can achieve goals in a wide range of real or virtual environments at least as effectively as humans can.

### Why This Is Important

As AI becomes more sophisticated and powerful, with recent great advances in generative AI (GenAI) in particular, a growing number of people see AGI as no longer purely hypothetical. Improving the understanding of at least the concept of AGI is critical for steering and regulating AI's further evolution. It is also important to manage realistic expectations and to avoid prematurely anthropomorphizing AI. However, if AGI becomes real, its impact on the economy, (geo)politics, culture and society cannot be underestimated.

## Business Impact

In the short term, organizations must know that the hype about AGI exists today among many stakeholders, stoking fears and unrealistic expectations about current AI's true capabilities. This AGI anticipation is already accelerating the emergence of more AI regulations and affects people's trust and willingness to apply AI today. In the long term, AI continues to grow in power and, with or without AGI, will increasingly impact organizations, including the advent of machine customers and autonomous business.

## Drivers

- Recent great advances in applications of GenAI and the use of foundation models and large language or multimodal models drive considerable hype about AGI. These advances have been enabled largely by the massive scaling of deep learning, as well as by the availability of huge amounts of data and compute power. To further evolve AI toward AGI, however, current AI will need to be complemented by other (partially new) approaches, such as knowledge graphs, multiagent systems, simulations, evolutionary algorithms, causal AI, composite AI and likely other innovations yet unknown.
- Vendors such as Google, IBM and OpenAI are openly discussing and actively researching the field of AGI, creating the impression that AGI lies within reach. However, their definitions of AGI vary greatly and are often open to multiple interpretations.
- Humans' innate desire to set lofty goals is also a major driver for AGI. At one point in history, humans wanted to fly by mimicking bird flight. Today, airplane travel is a reality. The inquisitiveness of the human mind, taking inspiration from nature and from itself, is not going to fizzle out.
- People's tendency to anthropomorphize nonliving entities also applies to AI-powered machines. This has been fueled by the humanlike responses of ChatGPT and similar AI, as well as AI being able to pass several higher-level education exams.
- Complex AI systems display behavior that has not been explicitly programmed. Among other reasons, this results from the dynamic interactions between many system components. Consequently, AI is increasingly attributed with humanlike characteristics, such as understanding. Although many philosophers, neuropsychologists and other scientists consider this attribution as going too far or being highly uncertain, it has created a sense that AGI is within reach or at least is getting closer. In turn, this has triggered massive media attention, several calls for regulation to manage the risks of AGI and a great appetite to invest in AI for economic, societal and geopolitical reasons.

## Obstacles

- Unreliability, lack of transparency and limited reasoning capabilities in current AI are not easy to overcome with the intrinsically probabilistic approach of deep learning. More data or more compute power for ever-bigger models are unlikely to resolve these issues, let alone to achieve AGI.
- The meanings of “intelligence” and related terminology like “understanding” have little scientific consensus, including the definition and interpretation of AGI. Scientific understanding about human intelligence is still challenged by the enormous complexity of the human brain and mind. Any claims about AGI – in whatever form it may emerge – are hard to validate when we don’t even understand human intelligence. However, even when AGI will be properly understood and defined, further technological innovations will likely be needed to implement AGI. Therefore, AGI as defined here is unlikely to emerge in the near future.
- If AGI materializes, it is likely to lead to the emergence of autonomous actors that, in time, will be attributed with full self-learning, agency, identity and perhaps even morality. This will open up a bevy of legal rights of AI and trigger profound ethical and even religious discussions. AGI also brings the risk of negative impacts on humans, from job losses to a new, AI-triggered arms race and more. This may lead to serious backlash, and regulations to ban or control AGI are likely to emerge in the near future.

## User Recommendations

- Engage with stakeholders to address their concerns and create or maintain realistic expectations. Today, people may be either overly concerned about future AI replacing humanity or overly excited about current AI’s capabilities and impact on business. Both cases hamper a realistic and effective approach to using AI today.
- Stay apprised of scientific and innovative breakthroughs that may indicate the possible emergence of AGI. Meanwhile, keep applying current AI to learn, reap its benefits and develop practices for its responsible use.
- Although AGI is not a reality now, current AI already poses significant risks regarding bias, reliability and other areas. Prepare for emerging AI regulations and promote internal AI governance to manage current and emerging future risks of AI.

## Sample Vendors

Aigo; Google; IBM; Microsoft; OpenAI

## Gartner Recommended Reading

[The Future of AI: Reshaping Society](#)

[Innovation Insight for Generative AI](#)

[Innovation Insight: AI Simulation](#)

[Applying AI — Key Trends and Futures](#)

[Innovation Insight for Artificial Intelligence Foundation Models](#)

## At the Peak

### Sovereign AI

Analysis By: Lydia Clougherty Jones, Clementine Valayer

**Benefit Rating:** High

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Adolescent

#### Definition:

Sovereign AI is the effort by nation-states to self-operationalize their development and use of AI with less dependence on the commercial market. It embodies political and cultural differences to advance sovereign objectives, including when developing AI strategy for value alongside sovereign-appropriate harm reduction. Given the wide variances across sovereign AI innovation to loss ratios, sovereign AI impacts international relationships, global trade and economic markets in unexpected ways.

#### Why This Is Important

Sovereign AI reflects the rapid acceleration of nation-states' adoption of AI techniques for their own use to improve alignment of nearly all of their internal government functions and activities with operational goals. While it could enhance an individual state's military defense, AI use by other nations could undermine those national security efforts. Sovereign AI aims to maximize AI value while decreasing AI risk, including for those sovereign states who collaborate to achieve common goals such as decreasing the impact of AI-generated "deepfakes" in political environments.

#### Business Impact

Sovereign AI impacts nearly all aspects of government and the enterprises with whom it interacts. It improves the effectiveness of operations by automating tasks, such as within government contact centers. Because sovereign AI modernizes the business of government, it can improve employee experience and accelerate citizen engagement. When sovereign states control their own AI systems, they reduce their dependence on other sovereign states and on the private tech market.

#### Drivers

- An increasing number of countries are actively planning and building their own AI infrastructure and capabilities to gear up their competitiveness and safeguard their future; they are developing what they consider their sustainable “sovereign AI.”
- Known and unknown risks of harms to citizens and society from irresponsible uses of AI drive sovereign states to want more control over the development of AI systems, and more so over generative AI (GenAI) government use cases.
- Increasing need for a sovereign entity to self-regulate including how its data is used to train large language models (LLMs). For example, nation-states are increasingly using AI tools to make important government decisions, but we hear from clients that these decisions are often outsourced to private companies without public input or oversight. This lack of transparency and accountability drives sovereign states to develop the AI tools themselves to address concerns about potential biases and conflicts of interest in these critical decision-making processes.
- Desire to decrease dependencies on other nations and the tech market, including when underrepresentation of cultural and linguistic inputs cannot be achieved.
- Sovereign AI plays an important role in digital sovereignty as it focuses on the sovereign control of AI data and systems, including control over computing capacity, data storage, access to human resources and proprietary knowledge for AI application development. Digital sovereignty can also significantly impact sovereign AI development, with the availability of locally stored data, for example, to train AI models.
- Sovereign AI is different from sovereign data strategies as the former’s core focus is being the developer and user of AI technologies, not the regulator of it. Sovereign data strategies reflect state efforts to regulate data and AI use by and about its citizens, private industry and its economy.
- Combating threats to political stability from the proliferation of deepfakes.
- Upskilling government workers today for a more AI-ready government workforce tomorrow.
- Meeting the increasing demand to advance local and national defense strategies.
- Progressing and maintaining leadership in the emerging technologies space.

## Obstacles

- Preparing an AI-ready IT infrastructure.



- Creating an AI-skilled government workforce.
- Modernizing the government culture to embrace advanced analytics and automation.
- Overcoming the pressures on an already-taxed IT infrastructure and fragmented business networks to develop and implement its own AI systems.
- Lack of the right data for training LLMs, resulting in AI output with varying utility.
- Lack of technically skilled humans to loop into the AI development and use life cycle, resulting in an increase of unintended negative outcomes from AI use.
- Diversity of needs and support across political and cultural variations impede nation-states from accelerating AI adoption across its government functions.
- Differences in political and cultural values will create inconsistent AI-value versus AI-harms analysis, leading to unpredictable impacts on international trade and global markets.
- Development of AI by nation-states across the world will lead to a fragmentation and possibly contradiction in the requirements for AI solutions, many of which cannot be met by either the public or private sector.

## User Recommendations

Sovereign states seeking a self-governed and controlled approach to the development of AI systems aligned with their strategic objectives should:

- Start small and prioritize the AI uses aligned with maximum advancement of your stakeholder and government business goals.
- Build an AI strategic roadmap that progresses from internal use cases to citizen-facing ones.
- Ensure that the AI strategy identifies key value opportunities and risks.
- Monitor and learn from sovereign AI already underway, including from [New Zealand](#), the [European Commission](#), [India](#) the [United States](#) and the [United Kingdom](#).
- Collaborate with (friendly) sovereign states to accelerate the learning curve, sharing failure analysis and positive narratives of unexpected success.

- Use Gartner's [The Pillars of a Successful Artificial Intelligence Strategy](#) to guide the nation toward self-governance of AI development, creating tangible value and achieving competitive national leader status.

## Gartner Recommended Reading

[Top Trends in AI Public Policy and Regulations for 2024](#)

[The Future of AI: Reshaping Society](#)

[Quick Answer: Why Is Empathy Critical for Postdigital Government?](#)

[Government Insight: U.S. Federal AI Executive Order Opportunities and Risks](#)

[The Impact of the 'U.S. Executive Order on AI'](#)

## AI TRiSM

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

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

### Definition:

AI trust, risk and security management (AI TRiSM) ensures AI governance, trustworthiness, fairness, reliability, robustness, efficacy and data protection. AI TRiSM includes solutions and techniques for model and application transparency, content anomaly detection, AI data protection, model and application monitoring and operations, adversarial attack resistance, and AI application security.

## Why This Is Important

AI models and applications should be subject to protection mechanisms during development and at runtime. Doing so ensures sustained value generation and acceptable use based on predetermined intentions. Accordingly, AI TRiSM is a framework that comprises a set of risk, privacy and security controls and trust enablers that helps enterprises govern and manage AI models and applications' life cycles — and accomplish business goals. The benefit is improved outcomes and performance and enhanced compliance with regulations such as the EU AI Act.

## Business Impact

Organizations that do not consistently manage AI risks are exponentially inclined to experience adverse outcomes such as project failures, AI misperformance and compromised data confidentiality. Inaccurate, unethical or unintended AI outcomes, process errors, uncontrolled biases, and interference from benign or malicious actors can result in security failures, financial and reputational loss, or liability and social harm. AI misperformance can also lead organizations to make suboptimal or incorrect business decisions.

## Drivers

- OpenAI's ChatGPT democratized third-party generative AI applications and transformed how enterprises compete and do work. Accordingly, the risks associated with hosted, cloud-based generative AI applications are significant and rapidly evolving.
- Democratized, third-party AI applications often pose considerable data confidentiality risks. This is partly because large, sensitive datasets used to train AI models often come from various sources, including data shared by users of these applications.
- Confidential data access must be carefully controlled to avoid adverse regulatory, commercial and reputational consequences.
- AI risk and security management imposes new operational requirements that are not fully understood and cannot be addressed by existing systems. New vendors are filling this gap.
- AI models and applications must be constantly monitored to ensure that implementations are compliant, fair and ethical. Risk management tools can identify and adjust bias controls where needed in both (training) data and algorithmic functions.
- AI outputs that are unchecked can steer organizations into faulty decision making or harmful acts because of inaccurate, illegal or fictional information driving business decisions.
- AI model and application explainability and expected behavior must be constantly tested through observation and testing of model and application outputs. Doing so ensures original explanations, interpretations and expectations of AI models and applications remain active during model and application operations. If they don't, corrective actions must be taken.
- Detecting and stopping adversarial attacks on AI requires new methods that most enterprise security systems do not offer.
- Regulations for AI risk management — such as the EU AI Act and other regulatory frameworks in North America, China and India — are driving businesses to institute measures for managing AI model and application risk. Such regulations define the new compliance requirements organizations will have to meet on top of existing ones, like those pertaining to privacy protection.

## Obstacles

- AI TRiSM is often an afterthought. Organizations generally don't consider it until models or applications are in production.
- Enterprises interfacing with hosted, large language models (LLMs) are missing native capabilities to automatically filter inputs and outputs — for example, confidential data policy violations or inaccurate information used for decision making. Also, enterprises must rely on vendor licensing agreements to ensure their confidential data remains private in the host environment.
- Once models and applications are in production, AI TRiSM becomes more challenging to retrofit to the AI workflow, thus creating inefficiencies and opening the process to potential risks.
- Off-the-shelf software that embeds AI is often closed and does not support APIs to third-party products that can enforce enterprise policies.
- Most AI threats are not fully understood and not effectively addressed.
- AI TRiSM requires a cross-functional team, including legal, compliance, security, IT and data analytics staff, to establish common goals and use common frameworks, which is difficult to achieve.
- Although challenging, the integration of life cycle controls can be done with AI TRiSM.

## User Recommendations

- Set up an organizational unit to manage AI TRiSM. Include members with a vested interest in AI projects.
- Define acceptable use policies at a level granular enough to enforce.
- Implement data classification and permissioning systems to ensure enterprise policies can be enforced.
- Establish a system to record and approve all AI-based applications and gain periodic user attestation that they are used according to preset intentions.
- Use appropriate AI TRiSM toolsets to manage AI model, application, and agent trust risk and security.
- Require vendors with AI components to provide verifiable attestations of expected AI behavior.
- Implement AI data protection solutions and use different methods for different use cases and components.
- Establish data protection and privacy assurances in license agreements with vendors hosting LLMs.
- Constantly validate and test the security, safety and risk posture of all AI used in your organization, no matter the footprint.

## Sample Vendors

Aporia; Bosch Global Software Technologies (AIShield); Harmonic; Lasso Security; ModelOp; Prompt Security; Protopia AI; TrojAI

## Gartner Recommended Reading

[Top Strategic Technology Trends for 2024: AI Trust, Risk and Security Management](#)

[Innovation Guide for Generative AI in Trust, Risk and Security Management](#)

[Market Guide for AI Trust, Risk and Security Management](#)

[Tool: Generative AI Policy Kit](#)

[Quick Answer: The EU AI Act and Its Anticipated Impact](#)

## Prompt Engineering

Analysis By: Frances Karamouzis, Jim Hare, Afraz Jaffri

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

### Definition:

Prompt engineering is the discipline of providing inputs, in the form of text or images, to generative AI (GenAI) models to specify and confine the set of responses the model can produce. The inputs prompt a set that produces a desired outcome without updating the actual weights of the model (as done with fine-tuning). Prompt engineering is also referred to as “in-context learning,” where examples are provided to further guide the model.

### Why This Is Important

Prompt engineering is the linchpin to business alignment for desired outcomes. It is important because large language models (LLMs) and GenAI models in general are extremely sensitive to nuances and small variations in input. A slight tweak can change an incorrect answer to one that is usable as an output. Each model has its own sensitivity level, and the discipline of prompt engineering is to uncover the sensitivity through iterative testing and evaluation.

### Business Impact

Prompt engineering has the following business impacts:

- **Performance:** It helps improve model performance and reduce hallucinations.
- **Business alignment:** It allows subject data scientists, subject matter experts and software engineers to steer foundation models, which are general-purpose in nature, to align to the business, domain and industry.
- **Time to market, quality, efficiency and effectiveness:** There are a number of architecture options as well as execution options that AI leaders must balance. There is also a myriad of prompt optimization tools that will diminish (or at the very least shift) the need for manual engineering.

## Drivers

- **Balance and efficiency:** The fundamental driver for prompt engineering is it allows organizations to strike a balance between consuming an “as is” offering versus pursuing a more expensive and time-consuming approach of fine-tuning. GenAI models, and in particular LLMs, are pretrained, so the data that enterprises want to use with these models cannot be added to the training set. Instead, prompts can be used to feed content to the model with an instruction to carry out a function.
- **Process or task-specific customizations or new use cases:** The insertion of context and patterns that a model uses to influence the output generated allows for customizations for a particular enterprise or domain, or regulatory items. Prompts are created to help improve the quality for different use cases — such as domain-specific question answering, summarization, categorization, and so on — 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

- **Prompt engineering is a new discipline:** The craft of designing and optimizing user requests to an LLM or LLM-based chatbot to get the most effective result is still emerging. Engineers are finding that desired outputs using GenAI can be challenging to create, debug, validate and repeat. Communities worldwide are developing new prompt engineering methods and techniques to help achieve these desirable outcomes.
- **Approaches, techniques and scalability:** 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. Scalable and maintainable methods of prompt engineering are still a work in progress for most organizations.
- **Role alignment:** Data scientists are critical to understanding the capabilities and limits of models, and to determining whether to pursue a purely prompt-based or fine-tuning-based approach (or combination of approaches) for customization. The ultimate goal is to use machine learning itself to generate the best prompts and achieve automated prompt optimization. This is in contrast to an end user of an LLM who concentrates on prompt design to manually alter prompts to give better responses.

## User Recommendations

- Build awareness and understanding of prompt engineering to quickly start the journey of shape-shifting the appropriate prompt engineering discipline and teams.
- Build critical skills among different team members that will synergistically contribute critical elements. For example, there are important roles for data scientists, business users, domain experts, software engineers and citizen developers.
- Educate the team with the myriad of options 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

FlowGPT; Google; HoneyHive; Magniv; Microsoft; PromptBase; Salesforce

## Gartner Recommended Reading

[How to Engineer Effective Prompts for Large Language Models](#)

[Prompt Engineering With Enterprise Information for LLMs and GenAI](#)

[Quick Answer: How Will Prompt Engineering Impact the Work of Data Scientists?](#)

[Generative AI Changes Software Engineering Leaders' Responsibilities](#)

## Responsible AI

**Analysis By:** Svetlana Sicular, Philip Walsh

**Benefit Rating:** Transformational

**Market Penetration:** 5% to 20% of target audience

**Maturity:** Adolescent

### Definition:

Responsible artificial intelligence (RAI) is an umbrella term for aspects of making appropriate business and ethical choices when adopting AI. These include business and societal value, risk, trust, transparency, fairness, bias mitigation, explainability, sustainability, accountability, safety, privacy, and regulatory compliance. RAI encompasses organizational responsibilities and practices that ensure positive, accountable and ethical AI development and operation.

### Why This Is Important

Early exploitation of generative AI resulted in the re-emergence of RAI as a key AI topic. As AI amplifies at a huge scale, with both good and bad outcomes, RAI enables the right outcomes by ensuring business value while mitigating risks. RAI can employ a set of tools and approaches, including industry-specific methods, adopted by vendors and enterprises. More jurisdictions introduce new regulations that drive and challenge organizations to adopt RAI practices.

## Business Impact

RAI assumes accountability for AI development and use at the individual, organizational and societal levels. If AI governance is practiced by designated groups, RAI extends its reach to all stakeholders involved in the AI process. RAI helps achieve fairness, even though biases are often baked into the data; gain trust, although transparency and explainability methods are evolving; and ensure regulatory compliance, despite the AI's probabilistic nature.

## Drivers

RAI helps AI participants develop, implement, utilize and address the various drivers they face. With further AI adoption, the RAI drivers are becoming more important and are better understood by vendors, buyers, society and legislators:

- The adoption of GenAI raises new concerns, such as hallucinations, leaked sensitive data, copyright issues and reputational risks that bring new actors in RAI (for example, in security, legal and procurement).
- Leading vendors are offering indemnification of their GenAI offerings, making customers more confident as part of their RAI approaches: although a good step, these are still incomplete.
- The organizational driver of RAI assumes the need to strike a balance between the business value and associated risks within regulatory, business and ethical boundaries. This includes considerations such as reskilling employees to adapt to AI technologies and safeguarding intellectual property.
- The societal driver includes resolving AI safety for societal well-being versus limiting human freedoms. Existing and pending legal guidelines and regulations, such as the [EU's Artificial Intelligence Act](#), make RAI a necessity.
- The customer/citizen driver is based on fairness and ethics and requires reconciling privacy with convenience. Customers/citizens may be willing to share their data in exchange for certain benefits.
- AI affects all ways of life and touches all societal strata; hence, the RAI challenges are multifaceted and cannot be easily generalized. New problems will continue to arise with rapidly evolving technologies and their uses.

## Obstacles

- Poorly defined accountability for RAI makes it look good on paper but renders it ineffective in reality.
- Organizations lack awareness of AI's unintended consequences. Many turn to RAI only after they experience AI's negative effects, whereas prevention is simpler.
- Most AI regulations are still in draft. AI products' adoption of regulations for privacy and intellectual property makes it challenging for organizations to ensure compliance and avoid all possible liability risks.
- Rapidly evolving AI technologies, including tools for explainability, bias detection, privacy protection and some regulatory compliance, lull organizations into a false sense of responsibility, while mere technology is not enough. A disciplined AI ethics and governance approach is necessary, in addition to technology.
- Measuring success is difficult. Creating RAI principles and operationalizing them without regularly measuring the progress makes it hard to sustain RAI practices.

## User Recommendations

- Publicize consistent approaches across all RAI focus areas. The most typical areas of RAI in the enterprise are fairness, bias mitigation, ethics, risk management, security, privacy, reliability, sustainability and regulatory compliance.
- Designate a champion for each use case who will be accountable for the responsible development and use of AI.
- Define the AI life cycle framework. Address RAI in all phases of this cycle. Address hard trade-off questions.
- Provide RAI training to personnel. Include AI literacy and critical thinking as part of the training.
- Operationalize RAI principles. Ensure diversity of participants and enable them to easily voice AI concerns.
- Participate in industry or societal AI groups. Learn best practices and contribute your own because everybody will benefit from this exchange. Ensure that policies account for the needs of any internal or external stakeholders.

## Sample Vendors

Adobe; Arthur; Fiddler AI; Google; H2O.ai; IBM; Microsoft; Responsible AI Institute; SolasAI; TruEra

## Gartner Recommended Reading

[Expert Insight Video: What Is Responsible AI and Why You Should Care About It?](#)

[Top Trends in AI Public Policy and Regulations for 2024](#)

[Software Engineering Leaders Must Help Drive Responsible AI](#)

[Best Practices for the Responsible Use of Natural Language Technologies](#)

[How to Ensure Your Vendors Are Accountable for Governance of Responsible AI](#)

## AI Engineering

**Analysis By:** Soyeb Barot, Anthony Mullen, Leinar Ramos, Joe Antelmi

**Benefit Rating:** High

**Market Penetration:** 1% to 5% of target audience

**Maturity:** Adolescent

### Definition:

AI engineering is the foundation for enterprise delivery of AI and generative AI (GenAI) solutions at scale. The discipline unifies DataOps, MLOps and DevOps pipelines to create coherent enterprise development, delivery (hybrid, multicloud, edge), and operational (streaming, batch) AI-based systems.

## Why This Is Important

The demand for AI solutions has dramatically increased, driven by the unrelenting hype surrounding GenAI. Few organizations have built the data, analytics and software foundations needed to move individual pilot projects to production at scale, much less operate portfolios of AI solutions at scale. There are significant engineering, process and culture challenges to address. To meet the demands for scaling AI solutions, enterprises must establish consistent AI pipelines supporting the development, deployment, reuse, governance and maintenance of AI models (statistical, machine learning, generative, deep learning, graph, linguistic and rule-based).

## Business Impact

AI engineering enables organizations to establish and grow high-value portfolios of AI solutions consistently and securely. Most AI developments are currently limited by operational and cultural bottlenecks. With AI engineering approaches — DataOps, ModelOps and DevOps — it is possible to deploy models into production in a structured, repeatable factory-model framework.

## Drivers

- DataOps, ModelOps and DevOps provide best practices for moving artifacts through the AI development life cycle. Standardization across data and model pipelines accelerates the delivery of AI solutions.
- The elimination of traditional siloed approaches to data management and AI engineering doubles the data engineering effort and reduces impedance mismatches across data ingestion, processing, model engineering and deployment, which inevitably drift once the AI models are in production.
- AI engineering enables discoverable, composable and reusable AI artifacts (data catalogs, knowledge graphs, code repositories, reference architectures, feature stores, model stores and others) across the enterprise context. These are essential for scaling AI enterprisewide.
- AI engineering makes it possible to orchestrate solutions across hybrid, multicloud, edge AI or Internet of Things.
- Broader use of foundational platforms provides initial success at scaling the production of AI initiatives with existing data, analytics and governance frameworks.
- AI engineering practices, processes and tools must be adapted to address GenAI. GenAI specific adaptations include support for prompt engineering, vector DBs/graph KBs, architecting and deploying multiagent, and interactive deployment models.
- AI engineering tools can be subdivided into model-centric and data-centric tools. Terms such as DataOps, LLMOps, LangOps or FMOps, or more broader terms such as ModelOps or MLOps, are used frequently, but we believe they are all a subset of AI engineering.

## Obstacles

- Sponsorship for foundational enterprisewide AI initiatives is unclear. The transformational promise of AI enablement has led executives to actively compete for enterprise AI responsibility.
- AI engineering needs simultaneous development of pipelines across domains and platform infrastructure maturity.
- AI engineering requires integrating full-featured solutions with specific tools, including open-source technologies, to address enterprise architecture gaps with minimal functional overlap. These include gaps around extraction, transformation and loading stores, feature stores, model stores, model monitoring, pipeline observability, and governance.
- AI engineering requires cloud maturity and possible rearchitecting, or the ability to integrate data and AI model pipelines across deployment contexts. Potential complexity and management of analytical and AI workloads alongside costs may deter organizations that are in the initial phases of AI initiatives.
- Enterprises often seek “unicorn” experts to productize AI platforms. Spot-fix vendor solutions will bloat costs and potentially complicate already intricate integration and model management tasks.



## User Recommendations

- Establish a leadership mandate for enterprisewide foundational AI initiatives.
- Maximize business value from ongoing AI initiatives by establishing AI engineering practices that streamline the data, model and implementation pipelines.
- Simplify data and analytics pipelines by identifying the capabilities required to operationalize end-to-end AI platforms and build AI-specific toolchains.
- Use point solutions sparingly and only to plug feature/capability gaps in fully featured DataOps, MLOps, ModelOps and PlatformOps tools.
- Develop AI model management and governance practices that align model performance, human behavior and delivery of business value to make it easier for users to adopt AI models.
- Leverage cloud service provider environments as foundational to build AI engineering. At the same time, rationalize your data, analytics and AI portfolios as you migrate to the cloud.
- Adopt a platform approach to GenAI by investing in centralized AI engineering tools for automation, governance and use-case enablement across a broad set of AI models and providers.
- Upskill data engineering and platform engineering teams to adopt tools and processes that drive continuous integration/continuous development for AI artifacts.

## Sample Vendors

Amazon Web Services; Dataiku; DataRobot; Domino Data Lab; Google; Microsoft; neptune.ai; OctoAI; Seldon Technologies; Weights & Biases

## Gartner Recommended Reading

[Top Strategic Technology Trends for 2022: AI Engineering](#)

[Demystifying XOps: DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI](#)

[A CTO's Guide to Top Artificial Intelligence Engineering Practices](#)

[Cool Vendors in AI Core Technologies – Scaling AI in the Enterprise](#)

[Case Study: AI Model Operations at Scale \(Fidelity\)](#)

## Edge AI

Analysis By: Eric Goodness

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

### Definition:

Edge AI refers to the use of AI techniques embedded in non-IT products (consumer/commercial), IoT endpoints, gateways and edge servers. It spans use cases for consumer, commercial and industrial applications, such as mobile devices, autonomous vehicles, enhanced capabilities of medical diagnostics and streaming video analytics. While predominantly focused on AI inference, more sophisticated systems may include a local training capability to provide optimization of the AI models at the edge.

### Why This Is Important

Many edge computing use cases are latency-sensitive and data-intensive and require autonomy for local decision making. This creates a need for AI-based applications in a wide range of edge computing and endpoint solutions. Examples include real-time analysis of edge data for predictive maintenance, inferences for decision support and video analytics. Increasingly, generative models (including smaller language models) have become an area of experimentation and investment.

### Business Impact

- Real-time data analysis and decision intelligence.
- Improved operational efficiency, such as manufacturing visual inspection systems that identify defects, wasted motion, waiting and over- or underproduction.
- Enhanced customer experience through feedback from AI embedded within products.
- Connectivity cost reduction with less data traffic between the edge and the cloud.
- Persistent functions and solution availability, irrespective of network connectivity.
- Reduced storage demand as only prioritized data is passed on to core systems.
- Preserved data privacy at the endpoint.

## Drivers

Overall, edge AI has benefited from improvements in the capabilities of AI. This includes:

- The maturation of MLOps and ModelOps tools and processes support ease of use across a broader set of features that span the broader MLOps functions. Initially, many companies came to market with a narrowcast focus on model compression.
- The improved performance of combined machine learning (ML) techniques and an associated increase in data availability (such as time-series data from industrial assets).

There is business demand for new and improved outcomes solely achievable from the use of AI at the edge, which include:

- Reducing full-time equivalents with vision-based solutions used for surveillance or inspections.
- Improving manufacturing production quality by automating various processes.
- Optimizing operational processes across industries.
- New approaches to customer experience, such as personalization on mobile devices or changes in retail from edge-based smart check-out points of sale.
- Growing interest in local deployments of generative AI.

Additional drivers include:

- Increasing number of users upgrading legacy systems and infrastructure in “brownfield” environments. By using MLOps platforms, AI software can be hosted within an edge computer or a gateway (aggregation point) or embedded within a product with the requisite compute resources.
- More manufacturers embedding AI in the endpoint as an element of product servitization. In this architecture, the Internet of Things (IoT) endpoints, such as in automobiles, home appliances and commercial building infrastructure, are capable of running AI models to interpret data captured by the endpoint and drive some of the endpoints’ functions.

- Rising demand for R&D in training decentralized AI models at the edge for adaptive AI. These emerging solutions are driven by explicit needs such as privacy preservation or the requirement for machines and processes to run in disconnected (from the cloud) scenarios.

## Obstacles

- Edge AI is constrained by the application and design limitations of the equipment deployed; this includes form factor, power budget, data volume, decision latency, location and security requirements.
- Systems deploying AI techniques can be nondeterministic. This will impact applicability in certain use cases, especially where safety and security requirements are important.
- The autonomy of edge AI-enabled solutions, built on some ML and deep learning techniques, often presents questions of trust, especially where the inferences are not readily interpretable or explainable. As adaptive AI solutions increase, these issues will increase if initially identical models deployed to equivalent endpoints subsequently begin to evolve diverging behaviors.
- The lack of quality and sufficient data for training is a universal challenge across AI usage.
- Deep learning in neural networks is a compute-intensive task, often requiring the use of high-performance chips with corresponding high-power budgets. This can limit deployment locations, especially where small-form factors and lower-power requirements are paramount.

## User Recommendations

- Determine whether the use of edge AI provides adequate cost-benefit improvements or whether traditional centralized data analytics and AI methodologies are adequate and scalable.
- Evaluate when to consider AI at the edge versus a centralized solution. Good candidates for edge AI are applications that have high communications costs, are sensitive to latency, require real-time responses or ingest high volumes of data at the edge.
- Assess the different technologies available to support edge AI and the viability of the vendors offering them. Many potential vendors are startups that may have interesting products but limited support capabilities.
- Use edge gateways and servers as the aggregation and filtering points to perform most of the edge AI and analytics functions. Make an exception for compute-intensive endpoints, where AI-based analytics can be performed on the devices themselves.

## Sample Vendors

Chooch; Edge Impulse; IFS (Falkonry); Litmus Automation; MicroAI

## Gartner Recommended Reading

[Emerging Tech Impact Radar: Edge Artificial Intelligence](#)

[Innovation Insight for Edge AI](#)

[Emerging Tech: Differentiate With an Edge AI Benchmarking Strategy](#)

[Market Guide for Edge Computing](#)

[Emerging Tech: Empower Outcome-Centric IoT With AI](#)

## Foundation Models

Analysis By: Arun Chandrasekaran

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

**Maturity:** Early 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, software engineering, chemistry, finance and legal use cases.

**Business Impact**

With their potential to enhance applications across a broad range of enterprise use cases, foundation models will have a wide impact across vertical industries and business functions. Their impact has accelerated, with a growing ecosystem of startups building enterprise applications on top of them. Foundation models will advance digital transformation within the enterprise by improving workforce productivity, automating and enhancing customer experience, and enabling rapid, cost-effective creation of new products and services.

## 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 two options are less complex, quicker and cheaper.
- **Superior performance across multiple domains** — The difference between these 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.
- **Fast-paced innovation** — The past year has seen an influx of foundation models, 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 are seeking initial gains.

## Obstacles

- **Do not deliver perfect 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.
- **Expansion to impractical sizes** — Large models are up to billions or trillions of parameters. They are impractically large to train for most organizations because of the necessary compute resources, which can make them expensive and ecologically unfriendly.
- **Concentrate power** — These models have been mostly built by the largest technology companies with huge R&D investments and significant AI talent, resulting in a concentration of power among a few large, deep-pocketed entities. This situation may create a significant imbalance in the future.

## User Recommendations

- **Create a strategy document** that outlines the benefits, risks, opportunities and execution plans for these models in a collaborative effort.
- **Plan to introduce foundation models into existing speech, text or coding domains.** If you have any older language processing systems, moving to a transformer-based model could significantly improve performance. Knowledge search, summarization, content generation are popular emerging use cases across industries.
- **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, 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.

## Sample Vendors

Alibaba Group; Anthropic; Cohere; Databricks; Google; Hugging Face; IBM; Microsoft; Mistral AI; OpenAI

## Gartner Recommended Reading

[Innovation Guide for Generative AI Models](#)

[Quick Answer: What Are the Pros and Cons of Open-Source Generative AI Models?](#)

## Synthetic Data

**Analysis By:** Arun Chandrasekaran, Anthony Mullen, Alys Woodward

**Benefit Rating:** High

**Market Penetration:** 5% to 20% of target audience



**Maturity:** Adolescent

**Definition:**

Synthetic data is a class of data that is artificially generated rather than obtained from direct observations of the real world. Synthetic data is used as a proxy for real data in a wide variety of use cases, including data anonymization, AI and machine learning (ML) development, data sharing, and data monetization.

**Why This Is Important**

A major problem with AI development today is the burden involved in obtaining real-world data and labeling it. This time-consuming and expensive task can be remedied with synthetic data, where data can be generated faster and cheaper. Additionally, for specific use cases such as training models for autonomous vehicles, collecting real data for 100% coverage of edge cases is practically impossible. Furthermore, synthetic data can be generated without personally identifiable information (PII) or protected health information (PHI), making it a valuable technology for privacy preservation.

**Business Impact**

Adoption is increasing across various industries. Gartner predicts a massive increase in adoption as synthetic data:

- Avoids using PII when training ML models via synthetic variations of original data or synthetic replacement of parts of data.
- Reduces cost and saves time in ML development.
- Improves ML performance as more training data leads to better outcomes.
- Enables organizations to pursue new use cases for which very little real data is available.
- Is capable of addressing fairness issues more efficiently.

## Drivers

- In healthcare and finance, buyer interest is growing as synthetic tabular data can be used to preserve privacy in AI training data.
- To meet the increasing demand for synthetic data for natural language automation training, especially for chatbots and speech applications, new and existing vendors are bringing new offerings to market. This is expanding the vendor landscape and driving synthetic data adoption.
- Synthetic data applications have expanded beyond automotive and computer vision use cases to include data monetization, external analytics support, platform evaluation and the development of test data.
- Transformer and diffusion architectures, the architectural foundations for generative AI (GenAI), are enabling synthetic data generation at quality and precision not seen before. AI simulation techniques are improving the quality of synthetic data by better recreating real-world representations.
- There is an expansion to other data types. While tabular, image, video, text and speech applications are common, R&D labs are expanding the concept of synthetic data to graphs. Synthetically generated graphs will resemble, but not overlap the original. As organizations begin to use graph technology more, we expect this method to mature and drive adoption.
- The growing adoption of GenAI models and future customizations of such models will drive the demand for synthetic data to pretrain these models.

## Obstacles

- Synthetic data can have bias problems, miss natural anomalies, be complicated to develop or not contribute any new information to existing, real-world data.
- Data quality is tied to the model that generates the data.
- There are no clear best practices on how to combine synthetic and real data for AI development.
- Synthetic data generation methodologies lack standardization.
- It is difficult to validate the accuracy of synthetic data. While a synthetic dataset may look realistic and accurate, it is difficult to know for sure if it accurately captures the underlying real-world environment.
- Buyers are still confused over when and how to use the technology due to the lack of skills.
- Synthetic data can still reveal a lot of sensitive details about an organization, so security and privacy are concerns. An ML model could be reverse-engineered via active learning. With active learning, a learning algorithm can interactively query a user (or other information sources) to label new data points with the desired outputs, meaning learning algorithms can actively query the user or teacher for labels.
- If fringe or edge cases are not part of the seed dataset, they will not be synthesized. This means the handling of such borderline cases must be carefully accommodated.
- There may be a level of user skepticism as data may be perceived to be “inferior” or “fake.”

## User Recommendations

- Identify areas in your organization where data is missing, incomplete or expensive to obtain, and is thus currently blocking AI initiatives. In regulated industries, such as healthcare or finance, exercise caution and adhere to rules.
- Use synthetic variations of the original data, or synthetic replacement of parts of data, when personal data is required but data privacy is a requirement.
- Educate internal stakeholders through training programs on the benefits and limitations of synthetic data. Institute guardrails to mitigate challenges such as user skepticism and inadequate data validation.
- Measure and communicate the business value, success and failure stories of synthetic data initiatives.

## Sample Vendors

Anonos (Static); Gretel; Hazy; Howso; MOSTLY AI; Parallel Domain; Rendered.ai; Tonic.ai; YData

## Gartner Recommended Reading

[Innovation Guide for Generative AI Models](#)

[Case Study: Enable Business-Led Innovation with Synthetic Data \(Fidelity International\)](#)

[Predicts 2024: The Future of Generative AI Technologies](#)

## ModelOps

Analysis By: Joe Antelmi, Soyeb Barot, Erick Brethenoux

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

**Definition:**

Model operationalization (ModelOps) is primarily focused on the end-to-end governance and life cycle management of advanced analytics, AI and decision models, such as models based on machine learning (ML), generative AI (GenAI), knowledge graphs, rules, optimization, linguistics, agents and others.

**Why This Is Important**

ModelOps helps companies in standardizing, scaling and augmenting their analytics and AI initiatives. It helps organizations to move their models from the lab environments into production. MLOps primarily focuses on monitoring and governance of ML models, while ModelOps assists with the operationalization and governance of all advanced analytics, decision and AI models, including GenAI and retrieval-augmented generation (RAG) systems.

**Business Impact**

ModelOps, as a practice:

- Provides the capability for the management and operationalization of diverse AI, analytics and decision systems.
- Enables the complex subsystems required for AI, analytics and decision system observability including versioning, monitoring, automation, data orchestration, experimentation, and explainability.
- Ensures collaboration among a wider business, development and deployment community, and the ability to associate AI, analytics and model outcomes with business KPIs.

## Drivers

- Modern AI systems are being built with a symbiotic combination of generative and classic AI models, agents and intelligent software capabilities. As the number of advanced analytics, AI and decision models at organizations increases, organizations will have to manage different types of prepackaged or custom-made models in production.
- Organizations want to be more agile and responsive to changes within their advanced analytics and AI pipelines not only with models but also with data, application and infrastructure.
- ModelOps provides a framework to separate responsibilities across various teams for how models (including GenAI, foundational models, analytics, ML, physical, simulation, symbolic, etc.) are built, tested, deployed and monitored across different environments (for example, development, test and production). This enables better productivity and collaboration, and it lowers failure rates.
- ModelOps provides tools to address model degradation via drift, and bias. In other scenarios, enabling model governance, explainability and integrity is paramount.
- The operationalization challenges of ML models are not new, but the capability to enable diverse models in production at the organization level using ModelOps is still evolving.
- Organizations don't want to deploy an unlimited number of open-source offerings to manage ModelOps, but there are few comprehensive solutions that provide end-to-end capabilities in every domain of model operationalization. Moreover, not every capability is required immediately. Often, versioning, monitoring and model orchestration precede the full implementation of feature stores, pipelines and observability.
- GenAI will require an increased focus on testing, and the introduction of capabilities to version, manage and automate prompts, routers, and retrieval-augmented generation systems. Fine-tuning will also require enhanced ModelOps capabilities to manage complex domain and function training datasets.

## Obstacles

- Organizations using different types of models often don't build the right ops, governance and management capabilities until they already have a chaotic landscape of unmanaged advanced analytic systems.
- Not all analytical techniques currently benefit from mature operationalization methods. Because the spotlight has been on ML techniques, MLOps benefits from a more evolved AI practice, but some models, like agentic modeling and optimization techniques, require more attention in ModelOps practices and platforms.
- ModelOps capabilities that help productionize GenAI are emerging but immature. Moreover, organizations are struggling to get GenAI into production, due to data, security and regulatory concerns.
- Organizations may adopt ModelOps platform capabilities that they don't immediately need. At the same time, organizations that are siloed and fail to adopt a comprehensive ModelOps strategy create redundancy in effort with respect to operationalization.

## User Recommendations

- Buy ModelOps capabilities integrated into your primary AI platforms. Enrich these capabilities with best-of-breed open-source or proprietary ModelOps offerings where unique problems, like feature stores or observability, require enhanced solutions.
- Utilize ModelOps best practices across composite AI, data, models and applications to ensure transition, reduce friction and increase value generation.
- Recruit/upskill additional engineers who can master ModelOps on AI systems that utilize unstructured data, search, graph and optimization.
- Encourage collaboration between development and deployment teams, and empower teams to make decisions to automate, scale and bring stability to the analytics and AI pipeline.
- Collaborate with software engineering teams to scale ModelOps. Offloading operationalization responsibilities to production support teams enables increased ModelOps specialization and sophistication across the ecosystem of complex AI-enabled applications.

## Sample Vendors

DataRobot; IBM; ModelOp; Modzy; Neptune.ai; OctoAI; SAS; Valohai; Verta; Weights & Biases

## Gartner Recommended Reading

[Launch an Effective Machine Learning Monitoring System](#)

[Innovation Guide for Generative AI in Trust, Risk and Security Management](#)

[The Logical Feature Store: Data Management for Machine Learning](#)

[Operationalize Machine Learning by Using Gartner's MLOps Framework](#)

[Toolkit: Delivery Metrics for DataOps, Self-Service Analytics, ModelOps and MLOps](#)

## Generative AI

Analysis By: Svetlana Sicular

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Adolescent

### Definition:

Generative AI (GenAI) technologies can generate new derived versions of content, strategies, designs and methods by learning from large repositories of original source content. Generative AI has profound business impacts, including on content discovery, creation, authenticity and regulations; automation of human work; and customer and employee experiences.

### Why This Is Important

GenAI exploration is widening:

- End-user organizations aggressively experiment with GenAI. Early adopters in most industries have had initial success with GenAI.
- Major technology vendors prioritize delivery of GenAI-enabled applications and tools.



- Numerous solutions have been emerging to innovate with foundation models, hardware and data for GenAI.
- Impacted by the GenAI hype, governments are introducing AI regulations and investing in national AI strategies.

## **Business Impact**

Business focus is shifting from excitement around foundation models to use cases that drive ROI. Most GenAI implementations are currently low-risk and internal. With the rapid progress of productivity tools and AI governance practices, organizations will be deploying GenAI for more critical use cases in industry verticals and scientific discovery. In the longer term, GenAI-enabled conversational interfaces will facilitate technology commercialization, democratizing AI and other technologies.

## Drivers

- Industry applications of GenAI are growing. GenAI reached creative work in entertainment, marketing, design, music, architecture and content generation.
- Best implementation practices are being discovered by the first enterprisewide deployments, and are fueling the top GenAI enterprise use cases: advanced chatbots, coding assistance and internal service desk. According to the 2023 Gartner AI in the Enterprise Survey, 18% of leaders highly involved in AI report that their organizations are advanced in GenAI adoption.
- GenAI is a top competitive area among major technology vendors. They compete on foundation model offerings, their enterprise readiness, pricing, infrastructure, safety and indemnification.
- New foundation models in new versions, sizes and capabilities are rapidly coming to market, making GenAI available for more use cases. Tools to improve model robustness, such as vector databases, graph technologies, LLM testing, security protection and open-source resources are making GenAI more usable.
- The progress is significant in multimodal models like Gemini or GPT4-Video, which are trained to take in both images and text; for example, they allow users to ask questions about images and receive answers via text. Models can combine concepts, attributes and styles to create original images, video and art, or translate audio to different voices and languages. Notably, text-to-image/video generation has advanced with the ability to create highly detailed and realistic visuals from textual descriptions.
- Enterprises are learning to use their own data with GenAI via prompt engineering and fine-tuning. AI-ready data and associated metadata have become central to GenAI strategies.
- Synthetic data helps enterprises to augment scarce data, mitigate bias, achieve superresolution or preserve data privacy.
- GenAI disrupts software engineering. Development automation techniques are promising to automate 5% to 10% of the programmers' work. Organizations are now willing to tackle legacy modernization with GenAI.

## Obstacles

- GenAI causes new ethical and societal concerns. Government regulations may hinder GenAI research. Pending regulations proliferate.
- Hallucinations, bias, a black-box nature and inexperience with a full AI life cycle preclude the use of GenAI in critical use cases for now.
- GenAI accountability, licensing and pricing are inconsistent among providers, and may catch customers off-guard.
- Reproducing results and finding references for generated information is challenging, but some validation solutions are emerging.
- Security professionals are new to certifying and protecting GenAI solutions; it will take time for security best practices to crystallize.
- GenAI is used for nefarious purposes. Full and accurate detection of generated content, such as deepfakes and disinformation, will remain challenging or impossible.
- The compute resources for training GenAI models are not affordable to most enterprises. Sustainability concerns about high energy consumption by GenAI are rising.

## User Recommendations

- Identify low-risk use cases where you can improve your business with GenAI by relying on purchased capabilities. Consult vendor roadmaps to avoid developing similar solutions in-house.
- Architect your GenAI solutions to be ready for near-future upgrades, as foundation models and data tooling for them are progressing swiftly.
- Pilot ML-powered coding assistants, with an eye toward fast rollouts, to boost developer productivity.
- Use synthetic data to accelerate the development cycle and lessen regulatory concerns.
- Quantify the advantages and limitations of GenAI. Issue GenAI policies and guidelines, as it requires skills, funds and caution.
- Mitigate GenAI risks by working with legal, procurement, security and fraud experts. Technical, institutional and political interventions will be necessary to fight AI's adversarial impacts.
- Optimize the cost and efficiency of AI solutions by employing composite AI approaches to combine GenAI with other AI techniques.

## Sample Vendors

Alibaba Cloud; Amazon Web Services; Anthropic; Google; Hugging Face; IBM; Meta; Microsoft; Mistral AI; OpenAI

## Gartner Recommended Reading

[Innovation Guide for Generative AI Technologies](#)

[Innovation Guide for Generative AI Models](#)

[How to Calculate Business Value and Cost for Generative AI Use Cases](#)

[When Not to Use Generative AI](#)

[10 Best Practices for Scaling Generative AI Across the Enterprise](#)

## Sliding into the Trough

### Neuromorphic Computing

Analysis By: Alan Priestley

**Benefit Rating:** Transformational

**Market Penetration:** Less than 1% of target audience

**Maturity:** Embryonic

#### Definition:

Neuromorphic computing is a technology that provides a mechanism to more accurately model the operation of a biological brain using digital or analog processing techniques. These designs typically use spiking neural networks (SNNs) rather than the deep neural networks (DNNs), feature non-von Neumann architectures and are characterized by simple processing elements, but very high interconnectivity.

#### Why This Is Important

Currently, most AI development uses parallel processing designs based on graphics processing units (GPUs). These are high-performance, but high-power-consuming, devices that are not applicable in many deployments. Neuromorphic computing utilizes asynchronous, event-based designs that have the potential to offer extremely low power operation. This makes them uniquely suitable for edge and endpoint devices, where their ability to support object and pattern recognition can enable image, audio, and sensor analytics.

#### Business Impact

- Neuromorphic computing architectures have the potential to deliver extreme performance for use-cases pattern recognition and signal analysis at very low power and can be trained using smaller datasets than other AI models, with the potential of in situ training.
- Neuromorphic computing designs can be implemented using low-power devices, which brings the potential to drive the reach of AI techniques out to the edge of the network, thereby accelerating key tasks such as image and sound recognition.

#### Drivers

- Today's DNN algorithms and large language models (LLMs) require the use of high-performance processing devices and vast amounts of data to train these systems, limiting scope of deployment.
- Different design approaches are being taken to implement neuromorphic computing designs — large-scale devices for use in data centers, and smaller-scale devices for edge computing and endpoint designs. Both these paths utilize SNNs to implement asynchronous designs that have the benefit of being extremely low power when compared with current DNN-based designs.
- Semiconductor vendors are developing chips that utilize SNNs to implement AI-based solutions.

## Obstacles

- **Accessibility:** GPUs are more accessible and easier to program than neuromorphic computing. However, this could change when neuromorphic computing and the supporting ecosystems mature.
- **Knowledge gaps:** Programming neuromorphic computing will require new programming models, tools and training methodologies.
- **Scalability:** The complexity of interconnection challenges the ability of semiconductor manufacturers to create viable neuromorphic devices.
- **Integration:** Significant advances in architecture and implementation are required to compete with other AI architectures. Rapid developments in DNN and LLM architectures may slow advances in neuromorphic computing, but there are likely to be major leaps forward in the next decade.

## User Recommendations

- Prepare for future utilization as neuromorphic architectures have the potential to become viable over the next five years.
- Create a roadmap plan by identifying key applications that could benefit from neuromorphic computing.
- Partner with key industry leaders in neuromorphic computing to develop proof-of-concept projects.

- Identify new skill sets that need to be nurtured for successful development of neuromorphic initiatives, and establish a set of business outcomes or expected value to set management's long-term expectations.

## Sample Vendors

AnotherBrain; Applied Brain Research; BrainChip; GrAI Matter Labs; Intel; Natural Intelligence; SynSense

## Gartner Recommended Reading

[Emerging Technologies: Tech Innovators in Neuromorphic Computing](#)

[Emerging Technologies: Top Use Cases for Neuromorphic Computing](#)

[Forecast: AI Semiconductors, Worldwide, 2022-2028, 1Q24 Update](#)

[Emerging Tech Impact Radar: Artificial Intelligence](#)

## Smart Robots

Analysis By: Dwight Klappich

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

### Definition:

A smart robot is an AI-powered, often-mobile machine designed to autonomously execute one or more physical tasks. These tasks may rely on, or generate, machine learning, which can be incorporated into future activities or support unprecedented conditions. Smart robots can be classified into different types based on the tasks or use cases, such as personal, logistics and industrial.

## Why This Is Important

Smart robotics is an AI use case, while robotics in general does not imply AI. Smart (physical) robots had less adoption compared with industrial robot counterparts but received great hype in the marketplace. The placement of smart robots has moved forward several positions this year. This is due to an increased interest and investment in smart robots in the last 12 months, as companies look to further improve logistic operations, support automation and augment humans in various jobs.

## Business Impact

Smart robots will make their initial business impact across a wide spectrum of asset-, product- and service-centric industries. Their ability to reduce physical risk to humans while also doing work with greater reliability, lower costs and higher productivity is common across these industries. Smart robots are already being deployed among humans to work in logistics, warehousing and safety applications.

## Drivers

- The market is becoming more dynamic with technical developments over the last two years, enabling a host of new use cases that have changed how smart robots are perceived and how they can deliver value.
- Robots have become more affordable as the cost of components such as processors, cameras and sensors has decreased over time.
- The physical building blocks of smart robots (motors, actuators, chassis and wheels) have incrementally improved over time. Similarly, areas such as Internet of Things (IoT) integration, edge AI and conversational capabilities have seen fundamental breakthroughs. These change the paradigm for robot deployments.
- Vendor specialization has increased, leading to solutions that have higher business value, since an all-purpose/multipurpose device is either not possible or is less valuable.
- Interest in smart robots has increased across various industries. Smart robots are used in performing diverse tasks across medical/healthcare, manufacturing, last-mile delivery, inspection of industrial objects or equipment, agriculture, workplace and so forth.
- Smart robots remain an emerging technology but the hype and expectations will continue to build over the next few years, as providers expand their offerings and explore new technologies. Adding capabilities like reinforcement learning will help drive a continuous learning loop for robots and swarm management.



## Obstacles

- Companies are still struggling to identify valuable business use cases and assess ROI for robots, especially outside of manufacturing and transportation.
- Complexity and versatility of tasks require complex decision making. Current smart robots excel at repetitive and predictable tasks and they are adaptable to various tasks.
- The continuous evolution and lack of commonality of pricing models and buying options create uncertainty for organizations. Companies struggle to compare and normalize all the various buying options they encounter such as monthly leasing, hourly charges, robot as a service or buying the robot outright.

## User Recommendations

- Evaluate smart robots as both substitutes and complements to the human workforce in manufacturing, distribution, logistics, retail, healthcare or defense.
- Begin pilots designed to assess product capability and quantify benefits, especially as ROI is possible even with small-scale deployments.
- Prepare yourself to adopt and evolve your processes and robotics strategy as you gain more experience in this field.
- Examine current business processes and redesign these as necessary to support the deployment of smart robots.
- Consider different purchase models for smart robots such as robot as a service or hybrid capital expenditure/operating expenditure models.
- Dissolve the reluctance from staff by developing training resources to introduce robots alongside humans as an assistant.
- Ensure sufficient cloud computing resources to support high-speed and low-latency connectivity in the next two years.
- Evaluate multiple global and regional providers due to fragmentation within the robot landscape.

## Sample Vendors

Ava Robotics; Geekplus; GreyOrange; HAHN Group (Rethink Robotics); iRobot; Locus Robotics; SoftBank Robotics; Symbotix; temi; UBTECH Robotics

## Gartner Recommended Reading

[Emerging Technologies: Top Use Cases for Smart Robots to Lead the Way in Human Augmentation](#)

[Emerging Technologies: Top Use Cases Where Robots Interact Directly With Humans](#)

[Emerging Technologies: Smart Robot Adoption Generates Diverse Business Value](#)

## Cloud AI Services

Analysis By: Van Baker, Bern Elliot

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

### Definition:

Cloud AI services provide AI model-building tools, APIs for prebuilt services and associated middleware that enable the building/training, deployment and consumption of machine learning (ML) and generative AI models running on prebuilt infrastructure as cloud services. These services include pretrained vision, language and other generative AI services, and automated ML and fine-tuning to create new models and customize prebuilt models.

### Why This Is Important

The use of cloud AI services continues to increase. Vendors have introduced additional services including large language model (LLM) APIs and solutions with fully integrated MLOps pipelines. The addition of low-code tools has added to ease of use. Applications regularly use AI cloud services in language, vision, and tabular data and code generation to automate and accelerate business processes. Developers are aware of these offerings, and are increasingly using both prebuilt and customized ML models in applications.

## Business Impact

Cloud AI services impact applications running the business allowing developers to enhance application functions. Generative AI adds a new category to these solutions allowing for the fine-tuning of LLMs to tailor performance. Data-driven decisioning mandates inclusion of ML models to add application functionality. Some AI technologies are maturing, but generative AI includes less mature capabilities. Cloud AI services enhance applications with models that score, forecast and generate content enabling data-driven business operations.

## Drivers

- **Opportunities to capitalize on new insights:** The wealth of data from both internal and third-party sources delivers insights such as the incorporation of predictive ML models that enable data-driven decision intelligence in applications.
- **Demand for conversational interactions:** The emergence of generative AI and large language models facilitates conversationally enabled applications where users can use LLMs with data sources to gain insights.
- **The need to meet business key performance indicators (KPIs):** There is a mandate for businesses to automate processes to improve accuracy, improve responsiveness and reduce costs by deploying both AI and ML models.
- **Reduced barriers to entry:** The ability to use pretrained generative AI models and fine-tune them has reduced the need for large quantities of data to train models. Access for developers and citizen data scientists to AI and ML services due to the availability of API-callable LLMs will further expand the use of AI by development teams.
- **Automated ML as an enabler for custom development:** Use of automated ML to customize packaged services to address specific needs of the business is much more accessible and doesn't require data scientists.
- **A wide range of cloud AI services:** Cloud AI services from a range of specialized providers in the market, including orchestration layers to streamline deployment of solutions, are available.
- **Emerging AI model marketplaces:** New marketplaces should help developers adopt these techniques through cloud AI services.

## Obstacles

- **Lack of understanding** by developers and citizen data scientists about how to adapt these services to specific use cases.
- **Grounding generative AI models** is a challenge, requiring well-crafted retrieval-augmented generation (RAG) solutions that often include vector embeddings and other capabilities to implement. Many CAIDS providers are offering these capabilities as part of their generative AI offering.
- **Pricing models** for cloud AI services that are usage-based presents a risk for businesses as the costs associated with use of these services can accrue rapidly. There is a need for comprehensive cost modeling tools to address this issue.
- **Increased need for packaged solutions** that utilize multiple services for developers and citizen data scientists.
- **Limited availability of ModelOps** tools that enable integration of AI and ML models into applications.
- **Lack of skills** such as prompt engineering and fine-tuning for developers to effectively implement these services in a responsible manner.

## User Recommendations

- Choose customizable cloud AI services over bespoke models to address a range of use cases and for quicker deployment and scalability.
- Improve chances of success of your AI strategy by experimenting with AI techniques, including the use of generative AI models such as LLMs and multimodal models and other cloud services. Ensure that generative AI models are loosely coupled as the technology continues to evolve rapidly.
- Use cloud AI services to build less complex models, giving the benefit of more productive AI while freeing up your data science assets for higher-priority projects.
- Empower non-data-scientists with features such as automated algorithm selection, dataset preparation and feature engineering for project elements. Leverage existing expertise on operating cloud services to assist technical professional teams.
- Utilize pretrained generative AI models to allow for rapid prototyping and deployment of LLM-enabled solutions.
- Develop cost modeling tools that allow the enterprise to effectively predict both usage and management costs as AI models are broadly deployed in applications across the business.

## Sample Vendors

Alibaba Cloud; Amazon Web Services; Baidu; Google; H2O.ai; IBM; Microsoft; NVIDIA; Oracle; Tencent

## Gartner Recommended Reading

[Critical Capabilities for Cloud AI Developer Services](#)

[Magic Quadrant for Cloud AI Developer Services](#)

## Climbing the Slope

### Autonomous Vehicles

Analysis By: Jonathan Davenport

Benefit Rating: Transformational

Market Penetration: Less than 1% of target audience

Maturity: Emerging

#### Definition:

Autonomous vehicles use various onboard sensing and localization technologies, such as lidar, radar, cameras, global navigation satellite system (GNSS) and map data, in combination with AI-based decision making, to drive without human supervision or intervention. Autonomous vehicle technology is being applied to passenger vehicles, buses and trucks, as well as for specific use cases such as mining and agricultural tractors.

#### Why This Is Important

Autonomous vehicles have the potential to change transportation economics, cutting operational costs and increasing vehicle utilization. In urban areas, inexpensive fares and high-quality service may reduce the need for private car ownership. Road safety will increase, as AI systems will never be distracted, drive drunk or speed. Autonomous features on privately owned vehicles will enable productivity and recreational activities to be undertaken, while the vehicle handles the driving operations.

#### Business Impact

Autonomous vehicles open the potential to disrupt traditional automotive business models by providing a software-based driver that is sold as part of a service that will generate high margin revenue. Self-driving systems will stimulate demand for onboard computation to run AI software, radically increasing the vehicle's overall bill of materials. After the office and home, vehicles will become a space where digital content can be created and consumed. Over time, fleet operators will likely retrain and redeploy their human commercial drivers to other, higher-value-adding roles within the company.

#### Drivers

- The formalization of regulations and standards for autonomous vehicles will aid implementation. Automated lane-keeping system (ALKS) technology has been approved by the United Nations Economic Commission for Europe (UNECE). This is the first binding international regulation for SAE Level 3 vehicle automation, with a maximum operational speed of 37 mph. With the new regulatory landscape, automakers worldwide are beginning to announce Level 3 solutions.
- Mercedes-Benz was the first automotive manufacturer to secure internationally valid system approval and has launched in Germany. In the U.S., its Level 3 solution has secured approval in Nevada and California. BMW has announced its Personal Pilot L3 function that controls the car's speed, distance to the vehicle ahead and lane positioning for them autonomously, which is now available on new 7 Series vehicles. In China, Changan, Great Wall Motor and Xpeng have announced Level 3 systems.
- The autonomous vehicle market is expected to evolve gradually from ADAS systems to higher levels of autonomy on passenger vehicles, rather than seeing a robotaxi-based revolution. This will require flexible vehicle operational design domains (ODDs).
- Self-driving trucks are a compelling business case. Driver pay is one of the largest operating costs for fleets associated with a commercial truck, plus goods can be transported much faster to their destination because breaks are no longer necessary. The Aurora Driver product is now at a "feature complete" stage, with a plan to launch a "middle-mile" driverless truck service at the end of 2024.
- With the ability of GenAI to generate synthetic data, the training of AI algorithms in simulation can be accelerated.
- For off-road use cases, autonomous vehicles can assist, replace or redeploy human workers to improve the accuracy with which work is done, lower operational costs and improve worker safety.

## Obstacles

- Due to the complexity of designing an autonomous vehicle, the cost to bring a commercial model to market has been greater than companies had envisioned, requiring significant investments.
- When autonomous vehicles are commercially deployed, the vehicle developers, not the human occupants, will be liable for the vehicles' autonomous operations. Specific insurance solutions are needed to cover the vehicle should it be involved in an accident.

- Challenges increasingly include regulatory, legal and societal considerations, such as permits for operation and the effects of human interactions.
- Automaker plans are being set back. For example, Hyundai's Genesis G90 and the Kia EV9 vehicles were expected to be equipped with a Level 3 Highway Driving Pilot (HDP) function. The delay was caused by the variety of real-world driving scenarios that the system would need to support.
- Despite continued improvements in Level 4 perception algorithms and broader self-driving systems used for mobility use cases (such as robotaxis), driverless operations have not scaled to different cities quickly. Cruise's accident in 2023 resulted in a strategy change that saw them lay off nearly 25% of its workforce.

## User Recommendations

- **Governments must:**
  - Craft national legislation to ensure that autonomous vehicles can safely coexist with a traditional vehicle fleet as well as a framework for their approval and registration.
  - Work closely with autonomous vehicle developers to ensure that first responders can safely respond to road traffic and other emergencies and self-driving vehicles don't obstruct or hinder activities.
- **Autonomous mobility operators** should support consumer confidence in autonomous vehicle technology by remaining focused on safety and an accident-free road environment.
- **Traditional fleet operators** looking to adopt autonomous technology into their fleets should minimize the disruptive impact on driving jobs (bus, taxi and truck drivers) by developing policies and programs to train and migrate these employees to other roles.
- **Automotive manufacturers** should instigate a plan for how higher levels of autonomy can be deployed to vehicles being designed and manufactured to future-proof vehicle purchases and enable future functions-as-a-service revenue streams.

## Sample Vendors

Aurora; AutoX; Baidu; Cruise; Mobileye; NVIDIA; Oxa; Pony.ai; Waymo; Zoox



## Gartner Recommended Reading

[Emerging Tech Impact Radar: Autonomous Vehicles](#)

[Lessons From Mining: 4 Autonomous Thing Benefit Zones for Manufacturers](#)

[Emerging Tech: Synthetic Data Will Drive Future Autonomous Vehicles](#)

[Emerging Tech: Top Semiconductor Technology Trends in Autonomous Vehicles, 2023](#)

[Market Trend: Connected and Autonomous Vehicle Data Enhances Software Life Cycle Management Transformation](#)

## Knowledge Graphs

Analysis By: Afraz Jaffri

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

### Definition:

Knowledge graphs are machine-readable representations of the physical and digital worlds. They include entities (people, companies and digital assets) and their relationships, which adhere to a graph data model — a network of nodes (vertices) and links (edges/arcs).

### Why This Is Important

Knowledge graphs capture information about the world in a visually intuitive format yet are still able to represent complex relationships. Knowledge graphs act as the backbone of a number of products, including search, smart assistants and recommendation engines. Knowledge graphs support collaboration and sharing, exploration and discovery, and the extraction of insights through analysis. Generative AI models can be combined with knowledge graphs to provide context for more accurate outputs in a technique becoming known as GraphRAG or G-RAG.

### Business Impact

Knowledge graphs can drive business impact in a variety of different settings, including:

- Digital workplace (such as collaboration, sharing and search)
- Automation (such as ingestion of data from content to robotic process automation)
- Machine learning (such as augmenting training data)
- Investigative analysis (such as law enforcement, cybersecurity and risk management)
- Digital commerce (such as product information management and recommendations)
- Data management (such as metadata management, data cataloging and data fabric)

## Drivers

- The need to complement AI and machine learning methods that detect only patterns in data (such as the current generation of foundation models) with the explicit knowledge, rules and semantics provided by knowledge graphs.
- The desire to make better use of unstructured data held in documents, correspondence, images and videos, using standardized metadata that can be related and managed and provide the foundation for AI-ready data.
- The increased usage of knowledge graphs with large language models (LLMs) to provide enhanced contextual understanding when answering questions on large quantities of enterprise data.
- The increasing awareness of the use of knowledge graphs in consumer products and services, such as smart devices and voice assistants, chatbots, search engines, recommendation engines and route planning.
- The emerging landscape of Web3 applications and the need for data access across trust networks, leading to the creation of decentralized knowledge graphs to build immutable and queryable data structures.
- The need to manage the increasing number of data silos where data is often duplicated, and where meaning, usage and consumption patterns are not well-defined.
- The use of graph algorithms and machine learning to identify influencers, customer segments, fraudulent activity and critical bottlenecks in complex networks.

## Obstacles

- Awareness of knowledge graph use cases is increasing, but business value and relevance are difficult to capture in the early implementation stages.
- Moving knowledge graph models from prototype to production requires engineering and system integration expertise. Methods to maintain knowledge graphs as they scale – to ensure reliable performance, handle duplication and preserve data quality – remain immature.
- The graph DBMS market is fragmented along three properties: type of data model (Resource Description Framework or property), implementation architecture (native or multimodal) and optimal workload (operational or analytical). This fragmentation continues to cause confusion and hesitation among adopters.
- Organizations want to enable the ingestion, validation and sharing of ontologies and data relating to entities (such as geography, people and events). However, making internal data interoperable with external knowledge graphs is a challenge.
- In-house expertise, especially among subject matter experts, is lacking, and identifying third-party providers is difficult. Often, expertise resides with vendors of graph technologies. Skills in scalability and optimization are also hard to acquire.

## User Recommendations

- **Create a working group of knowledge graph practitioners and sponsors** by assessing the skills of data and analytics (D&A) leaders, practitioners and business domain experts. Factors like use case requirements, data characteristics, scalability expectations, query flexibilities and domain knowledge of knowledge graphs should be addressed.
- **Run a pilot to identify use cases that need custom-made knowledge graphs.** The pilot should deliver not only tangible value for the business, but also learning and development for D&A staff.
- **Create a minimum viable subset that can capture the information of a business domain to decrease time to value.** Assess the data, both structured and unstructured, needed to feed a knowledge graph, and follow Agile development principles.
- **Utilize vendor and service provider expertise** to validate use cases, educate stakeholders and provide an initial knowledge graph implementation.
- **Include knowledge graphs within the scope of D&A governance and management.** To avoid perpetuating data silos, investigate and establish ways for multiple knowledge graphs to interoperate and extend toward a data fabric.

## Sample Vendors

Cambridge Semantics; Diffbot; eccenca; Fluree; Neo4j; Ontotext; Stardog; TigerGraph; TopQuadrant

## Gartner Recommended Reading

[How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications](#)

[3 Ways to Enhance AI With Graph Analytics and Machine Learning](#)

[How Large Language Models and Knowledge Graphs Can Transform Enterprise Search](#)

## Intelligent Applications

Analysis By: Justin Tung, Stephen Emmott, Alys Woodward

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

**Maturity:** Early mainstream

**Definition:**

Intelligent applications utilize learned adaptation to respond autonomously to people and machines. While applications can behave intelligently, intelligent applications are inherently smart/proactive. Rule-based approaches based on conditional logic are giving way to math-based training to elicit an appropriate response across a range of circumstances, including those that are new or unique. This enables the augmentation and automation of work across a broad variety of scenarios and use cases.

**Why This Is Important**

AI is the current competitive play for enterprise applications, with many technology providers now enabling AI in their products via inbuilt, added, proxied or custom capabilities. Recent developments in AI are continuing to enable applications to work autonomously across a wider range of scenarios with elevated quality and productivity. Integrated intelligence and AI can also support decision-making processes alongside transactional processes.

**Business Impact**

Three benefits to organizations buying or augmenting intelligent applications are:

- **Automation** — They increase automated and dynamic decision making, reducing the cost and unreliability of human intervention, and improving effectiveness of business processes.
- **Augmentation** — They increase the speed and quality of dynamic decision making based on context and risk, whether automated or via improved decision support.
- **Contextualization** — Applications can adapt to the context of the user or process, creating personalized experiences.

## Drivers

- **The hype wave for generative AI and large language models (LLMs), and the prevalence of conversational user interfaces (UIs)** as a way to interact with them, has inspired innovation and surfaced valuable ways to add AI functionality within preexisting applications. Features such as recommendations, insights and personalization are now more easily accessible via natural language prompts. Looking ahead, wider incorporation of chat-based interfaces will blur the line between interface and intelligence in an easily composable manner.
- **AI capabilities and features are increasingly integrated** into ERP, CRM, digital workplace, supply chain and knowledge management software within enterprise application suites. Embedded generative AI (as detailed above with LLMs) and traditional AI capabilities, such as predictive analytics, help to derive more insights from data in such applications. The 2023 Gartner AI in the Enterprise Survey shows that the top way to fulfill GenAI use cases is to use GenAI embedded into existing (purchased) applications (see [Survey Shows How GenAI Puts Organizational AI Maturity to the Test](#) for more information).
- **Organizations are demanding more functionalities from applications**, whether built or bought, expecting them to enhance current processes for both transactions and decision making with recommendations, insights and additional information. This in turn allows vendors to deliver higher value and drives higher prices.
- **The trend toward composable business architectures** has highlighted the possibilities for delivering advanced and flexible capabilities to support, augment and automate decisions, which have traditionally required an underlying data fabric and packaged capabilities to build. However, the increased adoption of LLMs has the potential to be used as a composable interface layer, kick-starting the ability to deliver on the composable architecture.

## Obstacles

- **Lack of data** — Intelligent applications require access to data from a range of systems, meaning application vendors need to think about data management technology and processes outside their own solutions.
- **Adding AI adds complexity to operations** — Models have to be trained and maintained, and users must understand what data is being used. Contextualizing insights requires business metadata.
- **Overuse of AI in marketing** — Vendors sometimes neglect the focus on business impact, which can generate a cynical response in business buyers, particularly when AI has not delivered value in the past.
- **Trust in system-generated insights** — It takes time for business users to see the benefit and trust such insights and some explainability is key.

## User Recommendations

- Challenge your packaged software providers to outline in their product roadmaps and/or ecosystems how they are incorporating AI to add business value in the form of a range of AI technologies.
- Evaluate the architecture of your providers by considering that the best-in-class intelligent applications are built from the ground up to be constantly collecting data from other systems, with a solid data layer in the form of a data fabric.
- Prioritize investments in specialized and domain-specific intelligent applications delivered as point solutions, which help solve problem areas such as customer engagement and service, talent acquisition, collaboration, and engagement.
- Bring AI components into your composable enterprise to innovate faster and safer, to reduce costs by building reusability, and to lay the foundation for business-IT partnerships. Remain aware of what makes AI different, particularly how to refresh ML models to avert implementation and usage challenges.

## Sample Vendors

Alkymi; ClayHR; Creatio; Eightfold AI; JAGGAER; OpenText; Prevedere; Salesforce; Sievo; SugarCRM

## Gartner Recommended Reading

[Top Tech Provider Trend for 2023: Intelligent Applications](#)

Survey Shows How Generative AI Puts Organizational AI Maturity to the Test

Maximize Competitiveness in Banking With Behavioral and Data Science



## Entering the Plateau

### Computer Vision

Analysis By: Nick Ingelbrecht

**Benefit Rating:** Transformational

**Market Penetration:** 20% to 50% of target audience

**Maturity:** Early mainstream

#### Definition:

Computer vision is a set of technologies that involve capturing, processing and analyzing real-world images and videos to extract meaningful, contextual information from the physical world.

#### Why This Is Important

Computer vision (CV) comprises a transformational collection of technologies, including AI/generative AI, advanced sensors and analytics that are essential to sensing and understanding the environment. Computer vision technology is driving innovation across many industries and use cases and is creating unprecedented business applications and opportunities.

#### Business Impact

CV technologies are used across all industries, and address a broad and growing range of business applications. These include physical security, healthcare, retail, automotive, robotics, manufacturing, supply chain/logistics, banking and finance, agriculture, government, and media and entertainment. Computer vision operates in the visible and nonvisible spectrum, including infrared, hyperspectral imaging, lidar, radar and ultraviolet.

#### Drivers

CV adoption is driven by an increasing demand for automation to reduce costs and improve monitoring and response capabilities. Other drivers include:

- **Improvements** in the availability and application of machine learning (ML) methods, tools and services, hardware processing efficiencies, and data generation and augmentation techniques.

- **Multimodal training** of CV with large language models (LLMs) facilitates natural language contextual search of unstructured image data and correlation of data on video streams at scale.
- **New architectures, models and algorithm enhancements** steadily improve the price/performance of CV applications. Combinations of convolutional neural networks (CNNs) and vision transformers are delivering leading levels of performance.
- **Advances in image and video generation**, such as Google Lumiere and OpenAI Sora, have broken new ground in the sophistication and realism of text to video generation.
- **The proliferation of cameras and other sensors is generating exponential increases in image data**, creating a critical and growing demand for methods to automate analysis and manage and extract value from that data. [Dynamic vision systems](#) are now being integrated into smartphones and lower-cost lidar products are opening new innovation areas.
- **3D capture, modeling and editing** of real-world objects and environments have been enabled by new techniques such as [Proximity Attention Point Rendering \(PAPR\)](#) and applications of [3D Generative AI](#).
- **Edge-enabled frameworks**, developer ecosystems, model compression and chip advancements.
- **New business models and applications** range from smartphone cameras and filters, through to global video content production and distribution, life-saving medical image diagnostics, autonomous vehicles, video surveillance for security, robotics, and manufacturing automation.
- **Sensor fusion**, multispectral and hyperspectral imaging expand the range of applications.
- **Improved reliability**, price, performance and functionality generate compelling business value.
- **Open-world recognition** using GenAI can identify and classify known objects, as well as handle unknown/unseen classes of objects and activities in novel contexts, without the need to train a model on specific examples.

## Obstacles

- High-end systems are expensive, and building business cases with adequate ROI is challenging.
- The CV market lacks independent standardization and performance benchmarks/KPIs, advanced solutions are far from being commoditized, and reliability remains an obstacle for mission-critical applications like autonomous vehicles.
- Integration is problematic due to a lack of open interfaces.
- Enterprises struggle to activate CV models in business processes and face data security as well as organizational challenges and user resistance to visual monitoring.
- Scaling solutions is challenging due to the hardware costs and high levels of customization and service support.
- Adequate training data may be hard or expensive to acquire, especially in areas where available open-source CV datasets are declining.
- Proprietary algorithms and patent pools deter innovation.
- Ethical, privacy and regulatory issues have emerged, including the use of deepfakes for embezzlement, misleading advertising and blackmail, as well as the capture of facial and other biometric data and the impact of new CV technologies on copyright and authenticity.

## User Recommendations

- Assess change management impacts of CV projects on the organization and its people.
- Focus initially on a few small projects, using fail-fast approaches, and scale the most promising systems into production using cross-disciplinary teams.
- Test production systems early in the real-world environment because lighting, color, object disposition and movement can break CV solutions that worked well in the development cycle.
- Build internal CV competencies and processes for exploiting image and video assets.
- Exploit third-party CV tooling and services to accelerate data preparation and reduce costs.

- Evaluate legal, regulatory, commercial and reputational risks associated with CV projects at the outset.
- Reduce the barrier to CV adoption by addressing two of the main challenges — lack of training data, and costly and constrained hardware — by investing in synthetic and augmented data solutions and model compression to improve model performance and expand the range of more valuable use cases.

## Sample Vendors

Adobe; Amazon Web Services; Baidu; Clarifai; Dragonfruit AI; Landing AI; Matroid; Microsoft; Prophesee; Tencent

## Gartner Recommended Reading

[Emerging Tech Impact Radar: Computer Vision](#)

[Emerging Tech: Revenue Opportunity Projection of Computer Vision](#)

[Innovation Guide for Generative AI in Computer Vision](#)

[Emerging Tech: Revenue Opportunity Projection of Computer Vision: Growth Markets](#)

[Emerging Technologies: Emergence Cycle for Computer Vision](#)

## Appendixes

See the previous Hype Cycle: [Hype Cycle for Artificial Intelligence, 2023](#)

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Table 2: Hype Cycle Phases

(Enlarged table in Appendix)

Phase	Definition
<i>Innovation Trigger</i>	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
<i>Peak of Inflated Expectations</i>	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.
<i>Trough of Disillusionment</i>	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
<i>Slope of Enlightenment</i>	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.
<i>Plateau of Productivity</i>	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 it as it enters this phase.
<i>Years to Mainstream Adoption</i>	The time required for the innovation to reach the Plateau of Productivity.

Source: Gartner (June 2024)

Table 3: Benefit Ratings

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

Source: Gartner (June 2024)

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 main stream	Robust technology Not much evolution in vendors or technology	Several dominant vendors
Legacy	Not appropriate for new developments Cost of migration constraints replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner (June 2024)

Document Revision History

- [Hype Cycle for Artificial Intelligence, 2023 - 19 July 2023](#)
- [Hype Cycle for Artificial Intelligence, 2022 - 8 July 2022](#)
- [Hype Cycle for Artificial Intelligence, 2021 - 29 July 2021](#)
- [Hype Cycle for Artificial Intelligence, 2020 - 27 July 2020](#)
- [Hype Cycle for Artificial Intelligence, 2019 - 25 July 2019](#)
- [Hype Cycle for Artificial Intelligence, 2018 - 24 July 2018](#)
- [Hype Cycle for Artificial Intelligence, 2017 - 24 July 2017](#)
- [Hype Cycle for Smart Machines, 2016 - 21 July 2016](#)
- [Hype Cycle for Smart Machines, 2015 - 24 July 2015](#)
- [Hype Cycle for Smart Machines, 2014 - 18 July 2014](#)

Recommended by the Authors

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

[Understanding Gartner's Hype Cycles](#)

[Tool: Gartner's Hype Cycle Builder](#)

[AI Design Patterns for Large Language Models](#)

[10 Best Practices for Scaling Generative AI Across the Enterprise](#)

[When Not to Use Generative AI](#)

[How to Build Knowledge Graphs That Enable AI-Driven Enterprise Applications](#)

---

© 2024 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates. This publication may not be reproduced or distributed in any form without Gartner's prior written permission. It consists of the opinions of Gartner's research organization, which should not be construed as statements of fact. While the information contained in this publication has been obtained from sources believed to be reliable, Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although Gartner research may address legal and financial issues, Gartner does not provide legal or investment advice and its research should not be construed or used as such. Your access and use of this publication are governed by [Gartner's Usage Policy](#). Gartner prides itself on its reputation for independence and objectivity. Its research is produced independently by its research organization without input or influence from any third party. For further information, see "[Guiding Principles on Independence and Objectivity](#)." Gartner research may not be used as input into or for the training or development of generative artificial intelligence, machine learning, algorithms, software, or related technologies.



Table 1: Priority Matrix for Artificial Intelligence, 2024

Benefit	Years to Mainstream Adoption			
↓	Less Than 2 Years ↓	2 - 5 Years ↓	5 - 10 Years ↓	More Than 10 Years ↓
Transformational	Composite AI Computer Vision	Decision Intelligence Embodied AI Foundation Models Generative AI Intelligent Applications Responsible AI	Autonomic Systems Autonomous Vehicles First-Principles AI Neuromorphic Computing	Artificial General Intelligence
High	Edge AI	AI Engineering AI-Ready Data AI TRiSM Causal AI Cloud AI Services Knowledge Graphs Prompt Engineering Sovereign AI Synthetic Data	AI Simulation ModelOps Multiagent Systems Neuro-Symbolic AI Smart Robots	
Moderate				
Low				Quantum AI

Source: Gartner (June 2024)

Table 2: Hype Cycle Phases

Phase	Definition
<i>Innovation Trigger</i>	A breakthrough, public demonstration, product launch or other event generates significant media and industry interest.
<i>Peak of Inflated Expectations</i>	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.
<i>Trough of Disillusionment</i>	Because the innovation does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
<i>Slope of Enlightenment</i>	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.
<i>Plateau of Productivity</i>	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 it as it enters this phase.
<i>Years to Mainstream Adoption</i>	The time required for the innovation to reach the Plateau of Productivity.

Source: Gartner (June 2024)

Table 3: Benefit Ratings

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

Source: Gartner (June 2024)

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 constraints replacement	Maintenance revenue focus
Obsolete	Rarely used	Used/resale market only

Source: Gartner (June 2024)