Gartner®

Generative AI Will Impact the Design and Control of Sociotechnical Systems

Published 13 October 2023 - ID G00796161 - 38 min read

By Analyst(s): Anthony Mullen, Erick Brethenoux, Pieter den Hamer, Bettina Tratz-Ryan

Initiatives: Artificial Intelligence

Generative AI is reshaping our approach to the design and safe control of the sociotechnical systems underpinning our society. IT leaders must consider the deep impacts of GenAI on systems development and "how work gets done" to avoid dangerous and potentially costly impacts to their organizations.

Overview

Impacts

- **Impact on jobs, processes and tasks.** Generative AI (GenAI) will radically change how we design and develop systems, deeply impacting enterprise jobs by shifting the emphasis on human work from lower-level tasks in development and design to higher-level framing of "work to be done."

- **Impact on technical architecture.** Generative AI will create a Cambrian explosion of AI agents, software, process and artifacts, necessitating a new architectural approach that will accelerate the shift from single-agent to multiagent computing.

- **Impact on trust, risk and security.** GenAI — by virtue of creating more code and integration between systems — will be a catalyst for emergent properties and feedback loops in AI systems. These will pose major risks to human value alignment and will require new approaches to control and oversight.

Recommendations

- **Generate multiyear plans to redesign tasks, processes and — eventually — jobs within your organization.** Immediately create a broad-reach training programme with HR for all employees aligned to your current enterprise AI use case portfolio. Provide employees with rich user interfaces (UIs) to help them shape and understand the behavior of AI, and empower business technologists and domain experts to create reusable and explainable semantic models of "how work should be done."

- **Clarify roles, functions and technical architecture by shifting to an agent-environment model approach rooted in AI simulation technology.** As the building of AI models and agents flows out to the rest of the business — boosted by GenAI's ability to discover and create capabilities — central functions like data science, software engineering and operations must look at how they build reusable environments (econometric models, social models, logistics models) that support the use and distributed development of agents across the organization.

- **Implement trust, risk and security approaches focused on three core actions to meet the challenge of dangerous feedback and emergent behaviors engendered by GenAI:**

    - Operate at not just the AI model level, but at the application, agent and AI systems levels.

    - Prioritize transparency, explainability and richer communicability with AI systems to ensure alignment with human goals and values.

    - Engage at government, governing body and grass roots levels to agree on legislation and shared approaches to bounding the risks of AI.

Strategic Planning Assumptions

- By 2025, the top three enterprise security applications will include the use of simulation techniques.
- By 2025, 60% of all low-code/no-code usage will involve multimodal interfaces powered by AI.
- By 2025, the daily tasks required by a software engineer will be 50% different than they are today.
- By 2026, investment in semantic technologies to steer, control and support dialogue with AI systems will see a 500% increase from today.
- By 2026, 20% of top data science teams will have rebranded as cognitive science or science consultancies, increasing diversity in staff skills by 800%.
- By 2027, generative AI models powered by large language models (LLMs) will be the biggest customer of technology marketplaces.
- By 2027, business unit staff will generate 50% of the technical components needed to develop AI systems — up from less than 5% today.
- By 2027, generative AI will have generated one trillion processes for enterprises.

Introduction

Use this index to navigate the document:

Introduction:

Impacts and Recommendations:

Notes:

Foreword

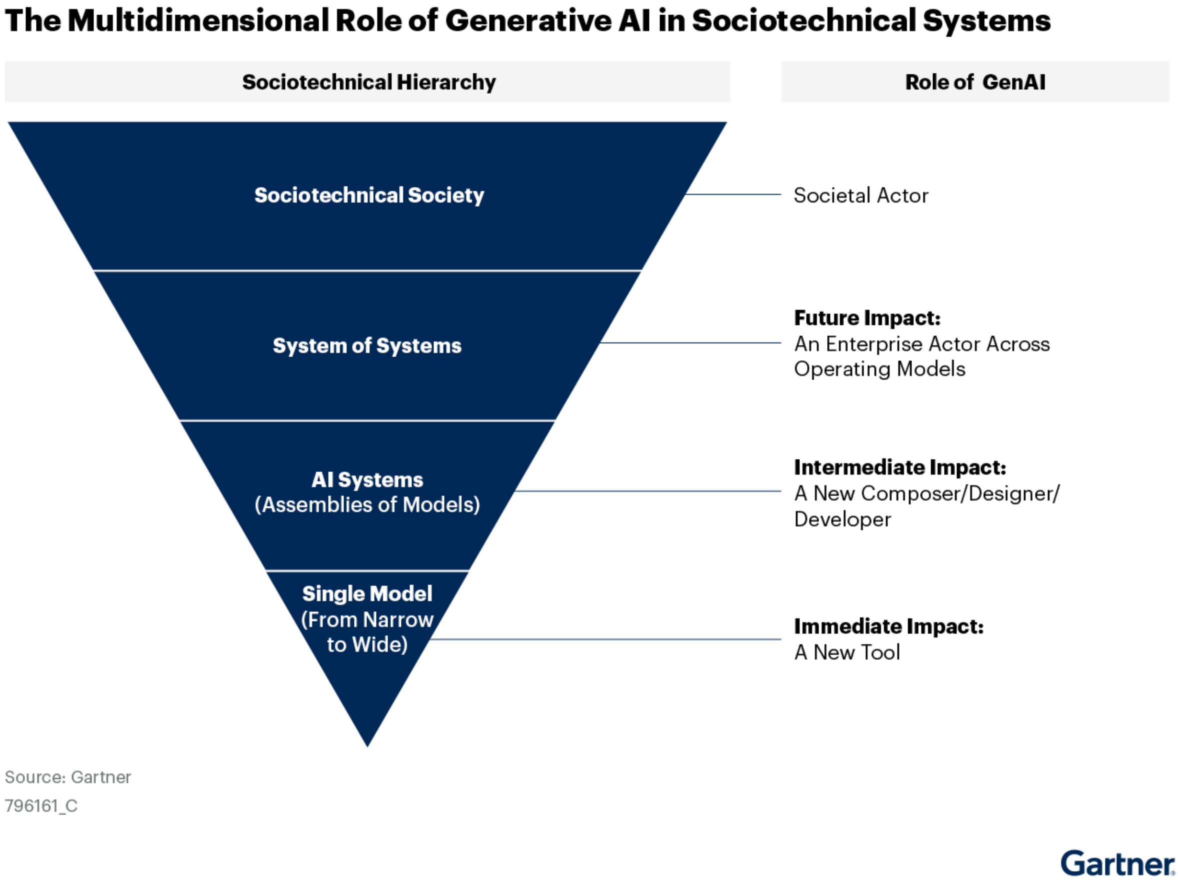
This research does not focus on society, per se, but rather on the systems, mechanisms and engineering that underpin the intersection of people and technology — sometimes called "sociotechnical systems." As AI is woven into sociotechnical systems, we create behaviors and relationships between people, systems and technology that have useful but risky mechanics such as autonomous decision making, synthetic content, feedback loops and networking effects. Generative AI accelerates this across applications, enterprise technical landscapes, ecosystems and society.

These systems and the implications of GenAI on them are of great interest to system designers — CIOs, enterprise architects, application leaders, data science teams, software engineers and UI designers — all of whom operate at the important intersection of AI and human communication. For a nontechnical look at the impact on society, see 9 Social and Cultural Implications of Generative AI.

The Multidimensional Impact of GenAI on Sociotechnical Systems

It is clear that AI technologies are joining humans and social constructs on the sociotechnical stage in a way that all other technologies have never previously done. AI is more like the carpenter than the hammer — it is not simply inert like a hammer. It has fitness functions and goals, just like humans do. Within AI, generative AI has a multidimensional role acting as both an individual agent and as a system — and even systems of systems — shaping the development of sociotechnical systems that underpin society (see Figure 1).

Figure 1: The Multidimensional Role of Generative AI in Sociotechnical Systems


The Multidimensional Role of Generative AI in Sociotechnical Systems
Sociotechnical Hierarchy
Role of GenAI
Sociotechnical Society
System of Systems
AI Systems (Assemblies of Models)
Single Model (From Narrow to Wide)
Societal Actor
Future Impact: An Enterprise Actor Across Operating Models
Intermediate Impact: A New Composer/Designer/ Developer
Immediate Impact: A New Tool
Gartner

Source: Gartner
796161_C

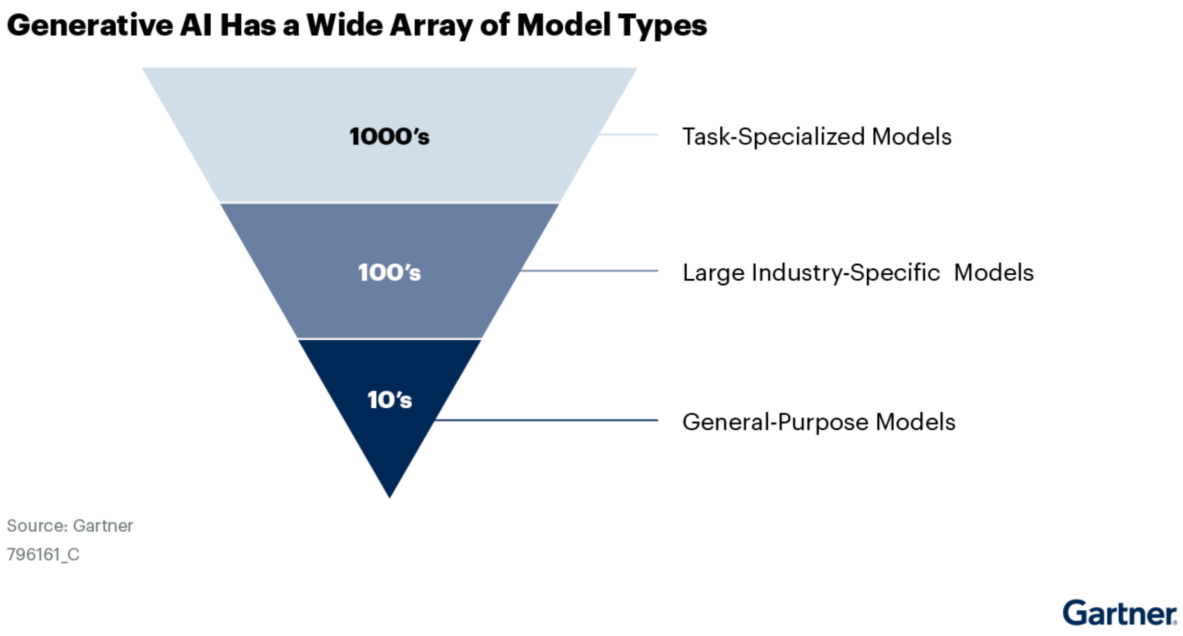
> AI, self-assembly, collaboration and a reduction in physical limitations of speed and scale are coming together to create an emergent machine society. We need to find a balance between machine and human value systems.

Until a few years ago, self-assembly of AI components was difficult. Sure, we could build AI models in isolation, but it took a lot of procedural code, skill and hours to connect them together with much of it not reusable. Generative AI has profoundly changed that. Large language models (LLMs), coupled with a conversational interface like ChatGPT, have proven to be a Rosetta Stone of sorts in connecting systems in a loosely coupled and flexible way. What was once complex and the preserve of data scientist teams is now open to anyone with the will and a little technical knowledge.

Generative AI Is Fast Accelerating Autonomous AI

Enterprises should expect the GenAI model marketplace to rapidly accelerate over the next 12 months when large general-purpose models (like the GPT series) are joined by industry-specific models and even small targeted models (see Figure 2).

Figure 2: Generative AI Has a Wide Array of Model Types


Generative AI Has a Wide Array of Model Types
1000's
Task-Specialized Models
100's
Large Industry-Specific Models
10's
General-Purpose Models
Gartner

Source: Gartner
796161_C

Further, organizations should not think of GenAI models as a singular technology, but rather as a mechanism to leverage a pool of resources at varying levels of generalism and specialism. These resources, which could be AI models or code, will be combined into innumerable combinations supported by GenAI, which in turn will support the creation of workflow to connect components together.

Generative AI creates a step change in the acceleration of systems assembly, as well as the diversity of system components. In 2023, we saw:

- **Autonomous generation:** Auto-GPT, released onto the market back in March, allows an AI agent to be spun up from a simple text prompt by a user. It then engages web and applications capabilities as it deems fit in order to fulfill the hopefully sensible high-level task with minimal human involvement, possibly running indefinitely.
- **Agent marketplace acceleration**: Hugging Face, one of the largest AI model marketplaces, has made APIs and interfaces to its hundreds of thousands of models available via simple nontechnical prompting. Before GenAI, piecing together different AI models required rare data science and AI expertise. Today, with systems like the Hugging Face Transformer Agent, there is a powerful democratizing shift from single agent to multiagent design from which consumers and business users alike can benefit.

The Three Waves of Impact From GenAI

Sociotechnical systems designers should envisage GenAI as having three waves of evolution, with each radically changing the design of jobs, systems architecture and security.

Most people are familiar with generative AI at an “artifact” level, where GenAI produces a “thing.” These things/artifacts range from prose, images, videos and code to 3D models and even protein folding sequences. The impacts of automating artifact creation on human work and play are multidimensional, spanning technical, political, economic, social, cultural, ethical, regulatory and social dimensions of human life — see more in 9 Social and Cultural Implications of Generative AI.

As profound and deep reaching as that is, it is only the first wave of disruption that GenAI will engender. We see three waves of GenAI overlapping but maturing at different rates, and these will directly impact the design and operation of sociotechnical systems. Today, we see the market in the midst of the first wave while experimenting with the other two:

- **Wave 1: Artifact creation and prompt engineering.** This first wave creates content, personalizes content, and democratizes the development of applications and interfaces to AI systems and models. Along with risks of hallucinations and hard-to-control behavior, GenAI has many legal challenges around intellectual property that will take years to resolve (see Prepare for AI Regulation by Addressing 4 Critical Areas). Even with these risks we see mass experimentation and rapid adoption.

- **Wave 2: Generative processes and multiagent systems.** GenAI, such as LLMs, can already break large tasks specified in language into subtasks and components. It can select third-party agents and software tools to use to fulfill these subgoals and tasks. In today's marketplace, this composable approach to development using multiple AI capabilities is mostly realized as LLM plugins, but the approach will evolve to be more agent based and multimodal (see Quick Answer: How Will the Generative AI Plug-In Market Evolve?). Over the next five years, the use of a pool of agents dynamically assembled to perform a task will become a much more common development method and will begin to displace existing procedural systems. Over time, we'll see a shift to real-time generation of process and workflow.

- **Wave 3: Adaptive networked systems.** These flexible process flow capabilities, combined with multiagent systems, will create the third wave of adaptive networked systems (see Top Strategic Technology Trends for 2023: Adaptive AI). Multiagent systems accelerate adaptability in sociotechnical systems by distributing tasks among autonomous agents that collaboratively learn and respond to dynamic environments. Through inter-agent communication, shared grounding and shared learning, they collectively adapt to unforeseen challenges and evolving scenarios.

Autonomous generation of process and workflow, along with a diverse set of marketplace agents to fulfill delegated tasks, invariably means the development of more systems and at greater speed. An ecosystem of models, content and data developing and evolving at high speed invariably creates a complex ecosystem that must be examined at both the individual model level and the systems level.

Explore the following impacts of GenAI in more detail below:

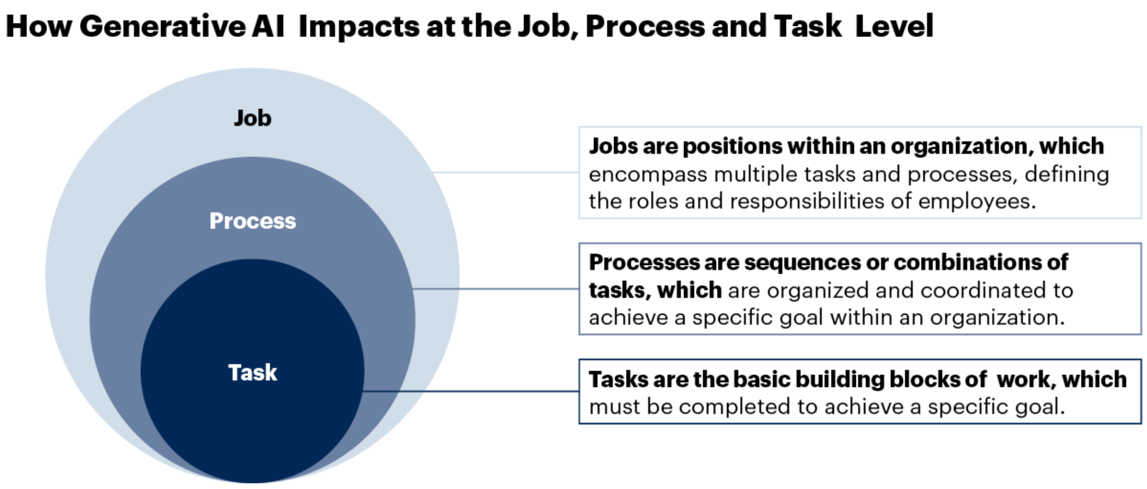
- Impact on job design
- Impact on architecture
- Impact on trust, risk and security

Impacts and Recommendations

Impact 1: Generative AI Will Radically Change How We Design and Develop Systems, Deeply Impacting Enterprise Jobs

GenAI has the capability to completely change how "jobs get done" in systems development. See Fig 3 on how GenAI impacts at the job, process and task level.

Figure 3: How Generative AI Impacts at the Job, Process and Task Level


How Generative AI Impacts at the Job, Process and Task Level
Job
Process
Task

Jobs are positions within an organization, which encompass multiple tasks and processes, defining the roles and responsibilities of employees.

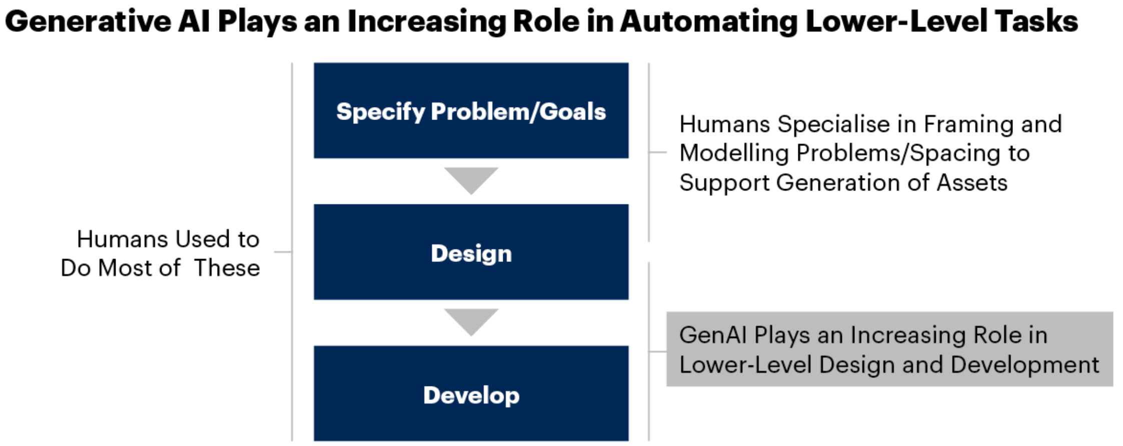
Processes are sequences or combinations of tasks, which are organized and coordinated to achieve a specific goal within an organization.

Tasks are the basic building blocks of work, which must be completed to achieve a specific goal.

Source: Gartner
796161_C

Because GenAI can augment or replace human effort in task and process work, this will create a shift in the emphasis of "who does what" when it comes to systems development. GenAI will play an increasingly large role in lower-level design and development. It will shift humans to specialize in framing and modeling their goals and business needs, and steer the behavior of underlying systems by being clear about the way in which a task should be performed (Figure 4).

Figure 4: Generative AI Plays an Increasing Role in Automating Lower-Level Tasks

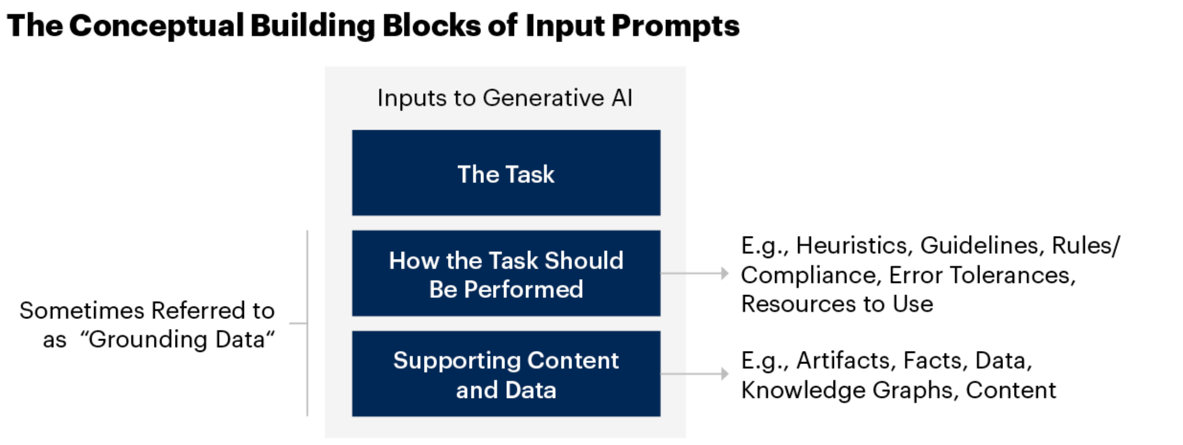

Generative AI Plays an Increasing Role in Automating Lower-Level Tasks
Specify Problem/Goals
Design
Develop
Humans Used to Do Most of These
Humans Specialise in Framing and Modelling Problems/Spacing to Support Generation of Assets
GenAI Plays an Increasing Role in Lower-Level Design and Development

Source: Gartner
796161_C

From a systems design and development perspective, enterprises face new choices on how processes and tasks get done — by humans or AI, or a mix of both. The broadly applicable nature of tools like ChatGPT impacts a wide gamut of workers, from analytics teams and software engineers to contact center staff and other business unit teams. Job roles will inevitably change as tasks and processes shift (see Plan for Generative AI's Impact on Jobs).

With humans specifying problems and goals rather than executing lower order tasks, a greater emphasis will be placed on how they engage with AI systems and frame their goals (Figure 5). Increasingly, workers will not be judged on how they develop work, but rather on how well they can frame problems and provide context.

Figure 5: The Conceptual Building Blocks of Input Prompts



The Conceptual Building Blocks of Input Prompts

Inputs to Generative AI

The Task

How the Task Should Be Performed

Supporting Content and Data

Sometimes Referred to as "Grounding Data"

E.g., Heuristics, Guidelines, Rules/ Compliance, Error Tolerances, Resources to Use

E.g., Artifacts, Facts, Data, Knowledge Graphs, Content

Source: Gartner
796161_C

Gartner

Recommendations

- **Create a broad reach training programme with HR for all employees aligned to your enterprise AI use case portfolio.** Educate employees on how you are using the technology today and planning to use it tomorrow, which will stimulate demand for AI across the organization. Be sure to examine the impacts that GenAI has across task, process and job automation before investment — a desire to automate too quickly will have negative social impacts in the company and may remove a cycle of workers that are hard to regain. Develop continuous learning programmes for existing employees to adapt to the changing technological landscape (see Quick Answer: How Should CXOs Structure AI Operating Models?).

- **Connect employees with rich UIs to support human-in-the-loop dialogues between humans and AI systems.** This will not just simply give them a result, but one that bears scrutiny under improved critical thinking and questioning. GenAI requires users to be clear, concise and bounded with what they ask of it, so writing and critical thinking workshops that support framing context or writing unambiguously will also provide systemic benefit (see Quick Answer: How Should CXOs Structure AI Operating Models?). Business unit leaders should invest in AI talent and upskilling, but ensure that their staff can both quiz and enrich AI systems.

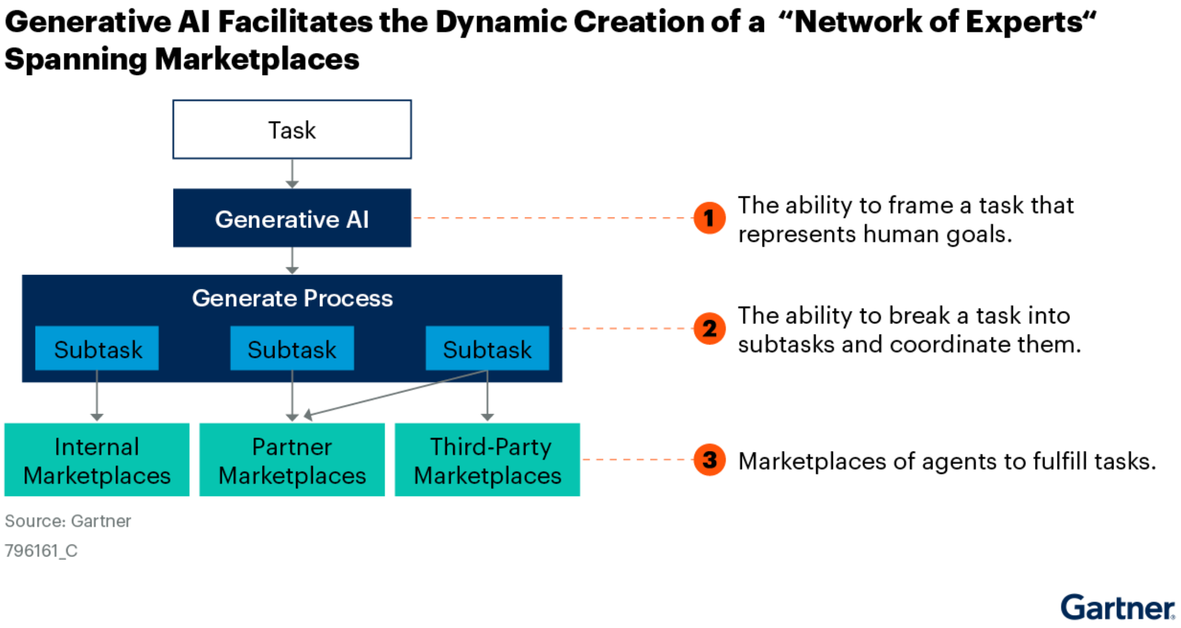
- **Empower business technologists and domain staff to create reusable and explainable semantic models of "how work should be done."** As the emphasis over the next five years shifts from developing code, models and integrating processes to specifying how work should be done and delegated to a mix of internal and external AI agents and humans, it will become important to consistently ground requests sent to an evolving set of GenAI systems. To date, the best tools that both allow for human understandable models of how concepts and rules can be developed and used for computation are semantic technologies (see Demystifying Semantic Layers for Self-Service Analytics). These allow employees, customers and citizens to communicate, disambiguate and use heuristic models of the world to engage with systems. Without their use we cannot connect explainable human heuristics (rules, norms) to the safe development of AI systems under our control. Software engineering teams should counter reductions in code generation demand with skills in design and information engineering.

- **Plan for a major shift in technical staff skills due to the shift from development to sociotechnical design.** As GenAI builds momentum in taking over low-level and specialized tasks spanning coding, process, model and composite AI development, you should create a medium term roadmap for your engineering group. Ensure that technical staff responsible for architecture, applications, AI and analytics can safely use simulation-based approaches and develop multiagent systems, which we foresee as being a major architectural piece of the enterprise stack (see Predicts 2023: Simulation Combined With Advanced AI Techniques Will Drive Future AI Investments). Data science teams should evolve into more diverse cognitive science teams to design safe, productive systems that benefit society.

Impact 2: Generative AI Will Create a Cambrian Explosion of AI Agents, Software and Artifacts, Necessitating a New Architectural Approach

While GenAI has come to the attention of most for generating words and images, its capabilities reach much further. It can interpret high-level requests and then flexibly seek to deliver solutions employing the use of first-, second- and third-party services — creating processes, code and integration as needed. As a result, GenAI will assist many more people in building many more models, processes and code, resulting in more AI within the enterprise landscape (see Figure 6).

Figure 6: Generative AI Facilitates the Dynamic Creation of a "Network of Experts" Spanning Marketplaces


Generative AI Facilitates the Dynamic Creation of a "Network of Experts" Spanning Marketplaces
Task
Generative AI
Generate Process
Subtask
Subtask
Subtask
Internal Marketplaces
Partner Marketplaces
Third-Party Marketplaces
1 The ability to frame a task that represents human goals.
2 The ability to break a task into subtasks and coordinate them.
3 Marketplaces of agents to fulfill tasks.
Gartner

Source: Gartner
796161_C

For Wave 2 of generative AI (generative processes and multiagent systems) to begin to impact, we need three things in place:

1. **The ability to frame a task that represents human goals.** Wave 1 of GenAI is addressing this area today with prompt engineering, artifact creation and retrieval-augmented generation (RAG) using content, and knowledge bases. This approach will mature to leverage neurosymbolic techniques to connect the enterprise semantic layer with underlying AI models.

2. **The ability to break a task into subtasks and coordinate them.** This is present in the industry today for GenAI, sometimes called AI orchestration (see Innovation Guide for Generative AI Technologies), but the idea is evolving rapidly. Orchestration of AI is shifting from using "procedural code" (to manage the interplay of models) to using AI techniques like "multiagent systems" (to coordinate tasks). These systems combine procedural elements like code and engineering processes with multiple AI models that collaborate, compete and cooperate with each other to fulfill higher-order tasks (see Innovation Insight: AI Simulation).

3. **Marketplaces of agents to fulfill tasks**. Early marketplaces of agents are seen today in the form of OpenAI ChatGPT plugins, where hundreds of vendor plugins can be included in an LLM-led conversation to extend the capability of LLMs (see Quick Answer: How Will the Generative AI Plug-In Market Evolve?). Today, we see companies like Hugging Face place LLMs across their AI model marketplaces, where generative tools can dynamically select models and capabilities at run time from a pool of hundreds of thousands of models. Companies like Unity are creating marketplaces for gaming covering GenAI, machine learning (ML) solutions and behavioral modes like pathfinding and nonplayer character (NPC) responsiveness.

Generative AI will rapidly accelerate the development of content and workflows, and form major parts of systems over the next five years. While individual agents, code and other components may become a commodity due to having large marketplaces or (in the case of agents) the ability to generate code and process, the arrangements of them and building them into smart, safe, reusable systems will not be a commodity. The challenge, then, for systems designers shifts from the building of individual models and components to framing the work to be done and steering and creating safe self-assembly of systems components. As these AI models and capabilities flood into a digital ecosystem, they will have complex and interlinked sociotechnical implications.

Evolution in the Approach to Grounding the Behavior of AI

The data within the GenAI space is often referred to as grounding data — it can be content like documents and lists, metadata or actual real-world data. This approach to grounding generative models can range from providing a comma-delimited list of 10 risk categories through to automated vehicles adhering to the laws of physics inside a simulation. Grounding data is very diverse, as are methods to facilitate its use.

Today, we see three broad approaches to grounding model behaviors with GenAI:

1. **Method 1: Sequential grounding** — This approach provides a “targeted slice” of context into generative models, where grounding data is sent into a generative agent; for example, a company document, sets of compliance rules or customer data being sent into a generative model like an LLM. This is by far the most popular current approach, but it depends heavily on custom-made orchestration and human effort to coordinate systems.

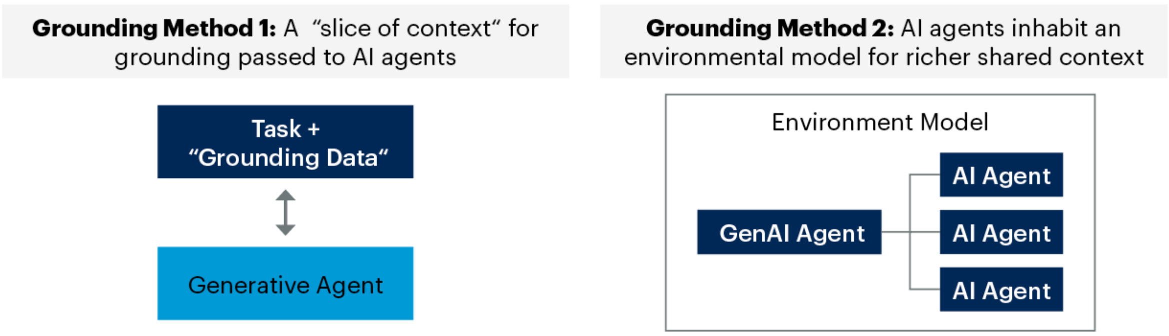
2. **Method 2: Hosted grounding** — This approach, mostly used in simulation and reinforcement learning, is where agents inhabit an environment to “ground” models, where they can both interact with the environment and enrich it — just as humans do in the real world. This approach gives weight to not just AI models but also to environmental models (made of data and metadata) — for example, pension investors represented as AI agents interacting with one another inside an econometric simulation. This approach is most common with asset-centric companies.

3. **Method 3: Hybrid grounding** — Where complex AI systems spanning central and edge AI use a mixture of hosted grounding (a central approach) and sequential grounding by local AI systems. This is very rare today.

Figure 7 outlines Methods 1 and 2.

Figure 7: Grounding Data as Argument to Generative AI vs. Hosted Grounding of Generative AI

Grounding Data as Argument to Generative AI vs. Hosted Grounding of Generative AI


Grounding Method 1: A “slice of context“ for grounding passed to AI agents
Task + “Grounding Data“
Generative Agent
Grounding Method 2: AI agents inhabit an environmental model for richer shared context
Environment Model
GenAI Agent
AI Agent
AI Agent
AI Agent

Source: Gartner
796161_C

There are major benefits to steering your architectural approach to Methods 2 and 3:

- Build your industry "world model" and associated grounding data incrementally and reuse it more easily.
- Observe and control (amplify/dampen) the behavior of AI models in aggregate.
- Easily swap agents in and out of your environment (rather than redesign workflows).
- Provide a richer context window to models for better performance.

The Agent and Environment Approach via Simulation

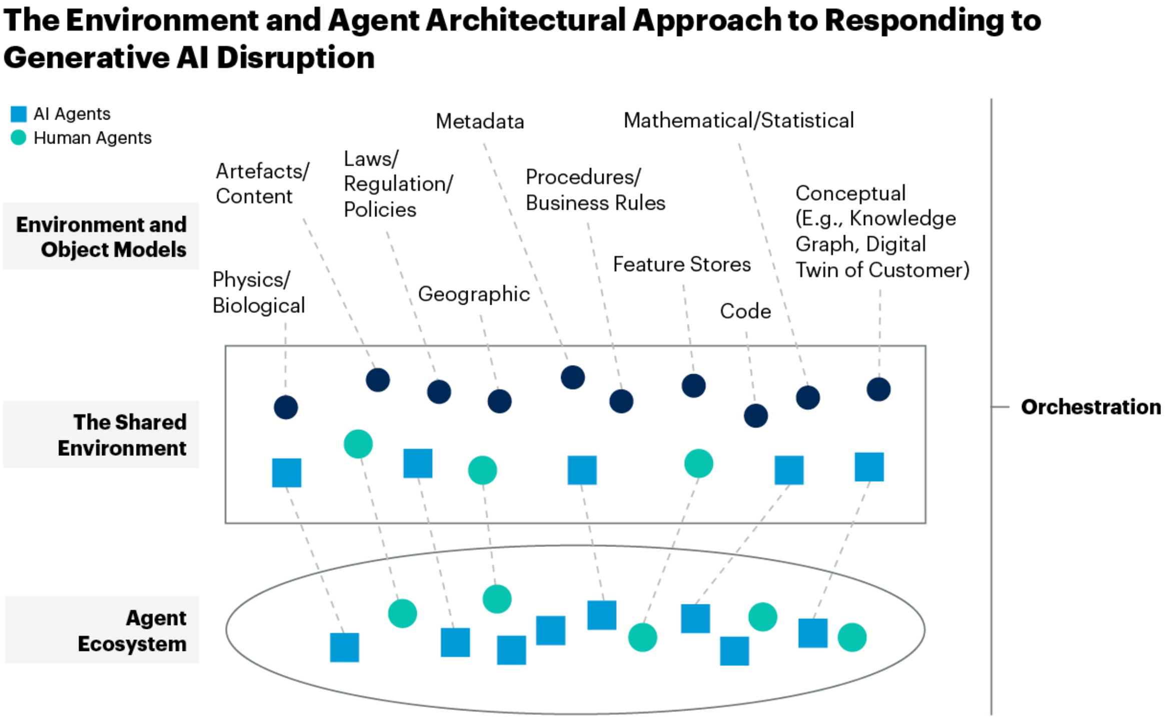
Simulation tools challenge existing data science and machine learning (DSML) platforms by not just enabling the development of individual agents and models, but also the environments to support them. DSML platforms are intrinsically multipersona platforms used by data scientists, software and hardware engineers, and domain experts. The major building blocks for a simulation-based design approach are as follows:

- **Environment models.** An environment model represents the surrounding context or the "world" in which agents operate — both AI and procedural agents. It encompasses the various states, conditions, entities and rules that dictate how agents will interact and how they and the environment itself will evolve over time. These models of the world are diverse and cover physical, econometric, social, conceptual and procedural/process features. Environment modeling is a crucial component when shifting to a multiagent system-based approach. It represents the capability of agents to understand and predict the environment in which they operate.
- **The shared environment.** This is the "stage" where events take place (real or generated) across time. Where environmental models interact with humans and AI models. The shared environment generates event data. The stage can be backstage in the form of development and testing platforms (including simulation platforms), as well as front-stage operations and customer-facing platforms.
- **Agent ecosystem.** An ecosystem of AI models (and humans) that can fulfill a range of tasks, from general purpose capabilities like foundation models to specific domain and task-specific functions within industries.

- **Orchestration.** Orchestration is the coordination of agents and the environment. This used to take lots of time with data science and software engineering development and integration. But increasingly, orchestration of models and code to deliver tasks and processes is shifting to generative and multiagent system approaches. While AI will play an increasing role in orchestration for enterprises, it will not be solely responsible for safe, explainable systems aligned to human interests. Orchestration needs to use composite AI methods mixing semantics, procedural code and data-driven AI to achieve a balance of control and explainability.

Simulation technologies, combined with GenAI, will accelerate agent-based computing. See Fig 8 for an illustration of a multiagent approach to sociotechnical.

Figure 8: The Environment and Agent Architectural Approach to Responding to Generative AI Disruption


The Environment and Agent Architectural Approach to Responding to Generative AI Disruption
■ AI Agents
● Human Agents
Environment and Object Models
The Shared Environment
Agent Ecosystem
Artefacts/ Content
Laws/ Regulation/ Policies
Metadata
Procedures/ Business Rules
Mathematical/Statistical
Conceptual (E.g., Knowledge Graph, Digital Twin of Customer)
Physics/ Biological
Geographic
Feature Stores
Code
Orchestration

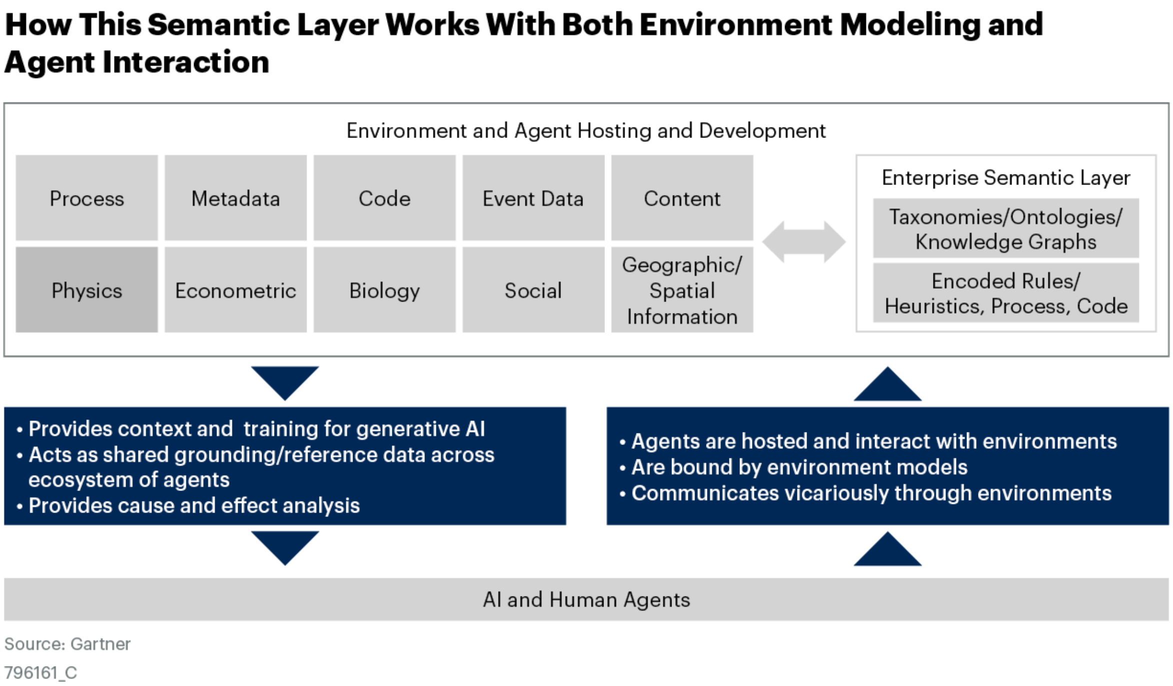
Source: Gartner
796161_C

Accurate environment modeling is essential for agents to make informed decisions, anticipate future states, and collaborate or compete effectively with other agents (see Note 2). Enterprises will spend more time modeling their environments in data, metadata and code than they will developing individual AI agents. It is therefore critical that these environment models can be collaboratively developed and reused. Here are some ways to think about environment modeling:

- **Self-serve car insurance —** A customer persona and business rules bounding the behavior of AI recommendation and journey automation algorithms.
- **Supply chain —** A GIS platform bounding the behavior of AI route planning.
- **Digital commerce —** A metaverse art gallery bounding the behavior of how users or AI engage with the store.
- **Predictive maintenance simulation —** A physics environment bounding the behavior of AI models predicting part failure of metallic components.

As our sociotechnical systems begin to contain more AI-enabled components, the need to have a strong, understandable "lingua franca" between AI and humans is critical. Language and the world of symbols are how we achieve this communication and transparency. From a technical point of view, enterprises must develop a semantic layer to achieve this explainable communication in multiagent systems and environment modeling (see Demystifying Semantic Layers for Self-Service Analytics). Figure 9 shows how this semantic layer works with both environment modeling and agent interaction.

Figure 9: How This Semantic Layer Works With Both Environment Modeling and Agent Interaction


How This Semantic Layer Works With Both Environment Modeling and Agent Interaction
Environment and Agent Hosting and Development
Process
Metadata
Code
Event Data
Content
Physics
Econometric
Biology
Social
Geographic/ Spatial Information
Enterprise Semantic Layer
Taxonomies/Ontologies/ Knowledge Graphs
Encoded Rules/ Heuristics, Process, Code
• Provides context and training for generative AI
• Acts as shared grounding/reference data across ecosystem of agents
• Provides cause and effect analysis
• Agents are hosted and interact with environments
• Are bound by environment models
• Communicates vicariously through environments
AI and Human Agents

Source: Gartner

796161_C

Recommendations

- **Redefine data and information architecture practices to take an environment modeling approach.** Construct models that capture the intricacies and nuances of your business in the real world — from products and attributes to geographical information and compliance rules. These models should be able to evolve based on new data and interactions, which often means connecting employees at scale. The richer the environment for AI agents, the better they can anticipate and respond to changes. Look to AI simulation platforms as the eventual architectural approach and build steps in your roadmap to get there.

- **Develop an agent ecosystem to support the shift toward composable AI.** The ecosystem for AI agents to fulfill work will be diverse, ranging from general purpose algorithms to domain- and task-specific agents — sourced from inside or outside your company. Turn your own enterprise capabilities into a set of agents through API services, and begin to explore how generative process and composite AI approaches so you can combine models for better business benefit.

- **Reduce the technical debt in developing multiagent systems by using semantic standards.** While there are existing standards and frameworks for multiagent systems (e.g., FIPA, OMB, JADE), they are not actively maintained and haven't yet been applied in depth to a post-GenAI ecosystem. Enterprises can use semantic AI platforms to support both development of grounding data and rules for agents, which benefit from standards like OWL and RDF (see Tool: Vendor Identification for Natural Language Technologies).

- **Clarify the roles of functions across the organization with the agent and environment model approach.** As the building of AI models flows out to the rest of the business (e.g., HR, customer service, ITSM) — propelled by GenAI's ability to discover and create capabilities — central functions like data science, software engineering and operations must begin to look at how they build reusable environments (econometric models, social models, logistics models) that support the use and distributed development of agents across the organization. Use simulation tools to ensure domain experts, software engineers and AI teams can collaborate on environment modeling and make it easy for employees to contribute to these models.

- **Shape AI architecture roadmaps to support simulation and multiagent system approaches.** Generative AI is only a part of the broader AI strategy and tooling landscape. If you haven't yet created an AI strategy, then begin by identifying key areas where AI can create value, addressing ethical considerations, ensuring data privacy and security, and fostering collaboration across departments. Those that do have an AI strategy must look at the impact of GenAI on it — the main impacts of which we have explored in this document. If you don't have an AI innovation lab or similar, set one up now or partner with vendors to get things moving.

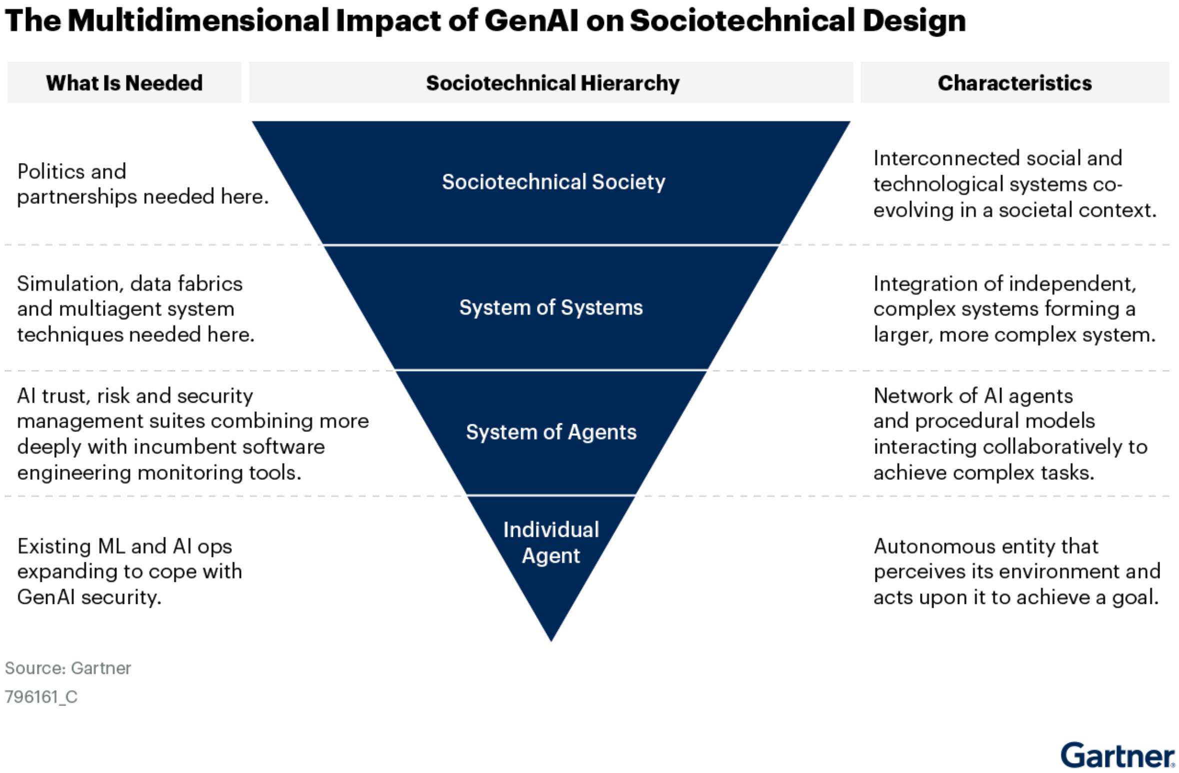
Impact 3: Generative AI Will Be a Catalyst for Emergent Properties and Feedback Loops

It is unlikely that we will be able to universally stop or singularly steer AI research in an even manner — no more than industrialization could be halted or universally controlled. This then leaves the biggest challenge to solve — the "alignment problem" in communicating with AI. That is, aligning AI with human goals and perhaps even having "shared goals" with AI — whatever that may become. The breadth of the alignment problem in AI was best crystallized in Zenon W. Pylyshyn's classic 1987 research collection "The Robots Dilemma — The Frame Problem of Artificial Intelligence" [1] and, of course, in fiction by Isaac Asimov.

Aside from the apocalyptic predictions of a few people (even experts), it could be that the true threat to humanity when it comes to AI is the ability to generate (and quickly disseminate) very damaging information that cannot be distinguished from facts. This poisons data, skews application behavior, and reduces trust in AI and human networks.

Sociotechnical designers must seek to design systems that engender trust and risk as they propagate, and then secure them locally and in aggregate. They must be able to do this even though parts of the system are opaque (especially neural networks). It is the engineering challenge of our age (see Figure 10).

Figure 10: The Multidimensional Impact of GenAI on Sociotechnical Design



The Multidimensional Impact of GenAI on Sociotechnical Design

What Is Needed	Sociotechnical Hierarchy	Characteristics
Politics and partnerships needed here.	Sociotechnical Society	Interconnected social and technological systems co-evolving in a societal context.
Simulation, data fabrics and multiagent system techniques needed here.	System of Systems	Integration of independent, complex systems forming a larger, more complex system.
AI trust, risk and security management suites combining more deeply with incumbent software engineering monitoring tools.	System of Agents	Network of AI agents and procedural models interacting collaboratively to achieve complex tasks.
Existing ML and AI ops expanding to cope with GenAI security.	Individual Agent	Autonomous entity that perceives its environment and acts upon it to achieve a goal.

Source: Gartner
796161_C

Generative AI impacts trust, risk and security systems that challenge existing single model-centric practices. While development of complex interlinked development will continue apace, existing trust, risk and security management (TRiSM) systems are not ready to handle it.

Moving Beyond a Single Model

Networked sociotechnical systems assurance requires approaches beyond managing a single model.

Today, TRiSM solutions for AI are just beginning to break out of managing and observing a single AI agent or collections (rather than integrations) of them. For enterprises owning or building these models, the largest market addressing the assurance of AI behavior (bias, drift, context collapse) today is MLOps and the emerging ModelOps space (see Use Gartner's MLOps Framework to Operationalize Machine Learning Projects). These operational tools overlap with more targeted security capabilities presented in the TRiSM space (see Market Guide for AI Trust, Risk and Security Management). Across all of these categories new solutions and tools are constantly emerging (see Innovation Guide for Generative AI in Trust, Risk and Security Management).

None of these markets are ready to support the levels of disruption we expect from GenAI over the next three years. Today, we see:

- **MLOps and ModelOps tools** that are able to manage and observe AI model behavior. These tools are being extended to include foundation models like LLMs. However, the majority of AI-related ops platforms do not have the ability to look at collective behavior and influences of AI models within an ecosystem. They also lack a unified approach to observing both AI models and applications.
- **Emerging tools in the TRiSM space** to observe and control generative model inputs and outputs, to detect content anomalies, thwart prompts and model hijacking. TriSM tools for GenAI are not yet embedded into existing enterprise platforms like CRM and ERP systems, and require effort to both appraise them and integrate them into existing applications and workflows. Further, because TRiSM tools are still in the very early days of development, they do not yet have consistent and reliable performance.

These markets aren't ready for a dramatic increase in the volume of AI-powered systems and they are nowhere near assuring trust, risk and security across deeply interconnected networks of AI, applications and humans. Today, we see social media platforms struggle to maintain a healthy network of AI and human actors at both the individual and societal level.

The Threat of Feedback Loops and Emergent Behavior

Generative AI dramatically shortens decision automation and systems development time, thereby creating dangerous characteristics to control that appear at speed like feedback loops and emergent systems behavior.

Just as language and models of the world empower humans to communicate, collaborate and to subsequently build networks and feedback loops of behavior, so too does AI empower "networks of AI" to develop. Sociotechnical systems designers must first understand feedback loops and emergent behaviors before they can design and control them to be trustworthy, low risk and secure.

GenAI accelerates the creation of:

- **Feedback loops that are based on bad (intentional and unintentional) information reinforcing undesired behaviors across networks.** At a single model level, this could be a computer vision algorithm generating bias information. At a system of systems level, it could be a poorly calibrated enterprise supply and demand ecosystem wasting money or overinvesting in a commodity. Designers must be aware of the following mechanics:
    - **Echo chambers and bias amplification.** If AI systems are continuously trained on data they generate, it will lead to amplification of biases already present in the original dataset.
    - **Quality degradation.** With each generation, the quality of the generated content may degrade either semantically or factually. This could lead to an increasing amount of nonsensical or uninformative content. Within a single model this can sometimes be referred to as model collapse, [2] but the idea extends to "system collapse" too.
    - **Homogenization.** Repeatedly training AI models on their own output could result in loss of diversity and novelty in the generated content, leading to homogenization. This might stifle creativity and limit the scope of ideas that AI can generate.
    - **Misinformation/disinformation.** False information can weight the behavior of models, or false "facts" may be used in claims or legal situations.

- **Emergent behaviors that can't be easily observed or controlled.** GenAI will rapidly accelerate the network of AI agents in our enterprise ecosystem spanning enterprise boundaries. Just as with people and social media networks, these networks of AI models and systems will exert influence on one another as goals either collaborate or compete. Emergent behaviors, both good and bad, arise from simple interactions in bulk and GenAI fuels conditions for emergent behavior. We see this emergent behavior in AI on two levels:
    - **Individual emergent behavior.** With bulk content and high-dimensional AI model features, along with a simple mechanic of predicting the next word (e.g., ChatGPT series), you have good conditions for emergent behavior. Tools like ChatGPT can perform hundreds of tasks but were only designed with a fraction of the hundreds in mind. Along with the need to manage feedback loops, there is also the risk of emergent properties from these generative-powered AI systems.
    - **Collective emergent behavior.** As discussed earlier, GenAI has the built-in mechanics to accelerate the creation of multiagent systems. These systems exhibit emergent behaviors, which means that when multiple individual agents work together, they create patterns and outcomes that are more complex and sophisticated than the sum of their individual actions. In 2019, OpenAI demonstrated intelligent emergent and unplanned behavior in a multiagent system, realized as a simulated hide and seek environment. [3] This collective behavior reaches further than just a single AI system — also operating at an enterprise, ecosystem and cultural level.

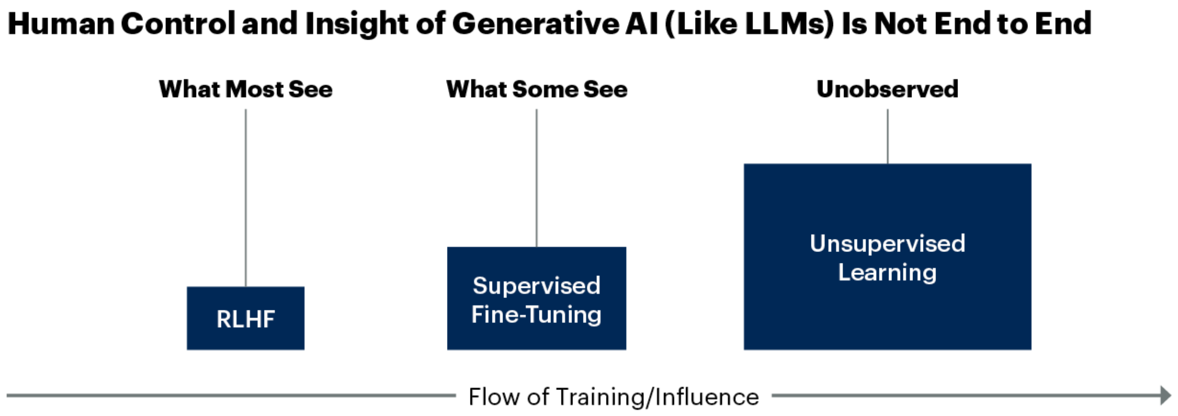
Automated and dynamic process, along with the collaboration of multiple AI agents and humans, will be accelerated by GenAI creating feedback loops and emergent behavior in sociotechnical systems. This systems behavior could have catastrophic results for individuals, groups, enterprises and wider society. While we see many initiatives to align AI systems behavior with human values and goals — such as OpenAI's alignment research [4] (using techniques like Recursive Reward Modeling [RRM], debate and amplification) — the pace of building GenAI-powered systems is outstripping our technical ability to control them.

Full Causal Explainability Will Not Be Possible in AI Systems

Full causal explainability of AI systems will not be possible, meaning neurosymbolic approaches are required to create a bridge between human and AI communication.

Human control and oversight of individual AI models is not end to end. Even for a single model like a LLM, the bulk of behavior (“neuron” processing in a deep neural network) goes unseen. This is quite unlike software code, which can be traced step by step (see Figure 11).

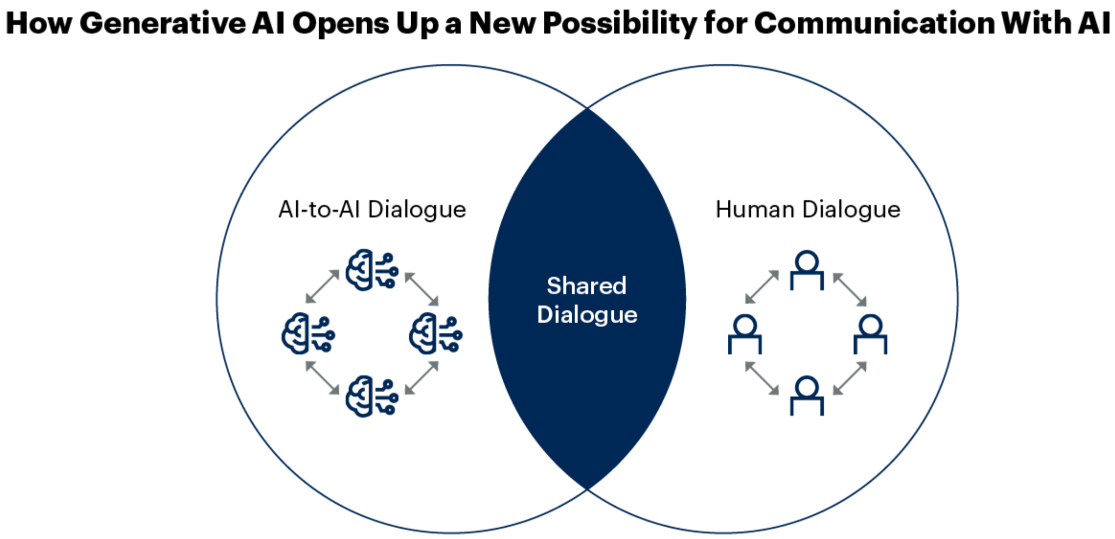
Figure 11: Human Control and Insight of Generative AI (Like LLMs) Is Not End to End


Human Control and Insight of Generative AI (Like LLMs) Is Not End to End
What Most See
What Some See
Unobserved
RLHF
Supervised Fine-Tuning
Unsupervised Learning
Flow of Training/Influence

Source: Gartner
RLHF = reinforcement learning from human feedback
796161_C

This lack of visibility into model behavior isn’t new; it applies to both AI and GenAI data-driven approaches. It is amplified further when we have AI systems in aggregate influencing one another. Simply put, if we can’t understand all the system machinations and ultimate decisions of complex emergent AI systems in a step-by-step, explainable, cause and effect manner, then we must at least be able to ask questions and to broadly shape their behavior. GenAI accelerates the imperative for humans to be able to communicate successfully with AI systems. In this regard, its novel capabilities with language will aid alignment with AI (see Figure 12).

Figure 12: How Generative AI Opens Up a New Possibility for Communication With AI


How Generative AI Opens Up a New Possibility for Communication With AI
AI-to-AI Dialogue
Shared Dialogue
Human Dialogue

Source: Gartner
796161_C

Generative AI, while challenged with hallucinations and inaccuracies, is a major boon for communication between humans and AI systems. It provides simple conversational access to knowledge bases, data and tasks via a flexible, natural language interface. That should give us hope.

Control Mechanisms Lack Maturity

Methods of control are not yet mature enough to manage AI feedback loops and emergent properties, meaning innovation is needed beyond current model-centric AI risk tooling.

At present, AI risk approaches are primarily focused on the observation and control of a single model. But that will not be enough — we need to evolve to look at the behavior of systems and systems of systems. Looking at other complex sociotechnical systems like economic models or social media, it is clear that a fine-grained level of control will not be possible. Rather, a set of mechanisms to slow down and interrupt AI systems is needed to avoid cascading viral negative effects. See Table 1 for an illustration of this control.

Table 1: Control Mechanisms Versus Level or Application Within a Sociotechnical Systems Hierarchy

Control Lever		Level Applied to
Amplify/dampen behavior Meter/throttle behavior Delay behavior (by time or resources) Apply limits and quotas on behaviors Semantic gates/rules Factcheck behavior Shape behavior with ground truth Gain consensus on behavior Human interrupt Intentional hallucination (adjust weights or reweight priors)	X	First-party model (e.g., custom price prediction AI model) Second-party model (e.g., partner recommendation algorithm) Third-party model (e.g., ChatGPT or LLM) Composite AI system level (e.g., multimodal chatbot or IoT project) Enterprise AI system of systems (e.g., shared metadata infrastructure) Industry level National level Societal level

Source: Gartner

Recommendations

- **Engage in multiple dimensions of policy and procedure making — from government to grass roots.** Given the impact of GenAI on sociotechnical systems, it is essential to engage with a wide range of stakeholders. These range from top-down government and governing bodies to more grass roots engagements with customers, staff, partners and peers to agree on shared approaches to bounding the risks of AI. Participate in dialogues and initiatives that aim to develop industry standards, best practices and regulatory frameworks for GenAI. Importantly, these should be open and applicable to many. Through common interest, shape the responsible and ethical development and deployment of GenAI technologies. Stay abreast of legislation by reading Prepare for AI Regulation by Addressing 4 Critical Areas.

- **Mitigate risk by tying decision automation and adaptive AI to behavioral modeling.** Integrating automation and adaptive AI with behavioral modeling facilitates dynamic adaptations grounded in human behavioral data, thereby bolstering predictive precision and mitigating associated operational risks. To observe emergent properties and risks in sociotechnical systems, both technical and social models of the world are required. Evolve marketing, sales and HR to begin to look at psychographics techniques for customers and employees. This will facilitate a joined-up view of systems and human behavior.

- **Develop a “simulate and test” mindset supported by multiagent and simulation techniques.** Explore AI simulation, AI optimization, process simulation and decision intelligence platforms to support controlled experiments in identifying potential systems automation problems and risks before they occur in the real world. Use data fabrics to observe the behavior of applications and data sources, and also to interpret the meaning of the behavior. Do so by connecting the data fabrics metadata and infrastructure to semantic platforms to support analysis. Share procedural and application model details with I&O, as well as trust, risk and security teams, so they may explore multiagent system implications.

- **Prioritize development of a UI and gamification techniques to tackle the alignment problem with humans and AI.** It is imperative at this early stage of AI development that customers and employees can effectively engage with AI in a clear, dialectic and rich way (see Design and Implement Human-in-the-Loop Interfaces for Control, Performance and Transparency). UI designers should use (but rightsize) techniques like anthropomorphization in UI design to promote dialogue (see Anthropomorphize AI to Make It Safe and Successful). Without communication, dialogue and explanation, we create a very weak root in AI evolution. Enterprises should accelerate their efforts to develop explainable and communicable interfaces to shape and steer the behavior of AI.

- **Mandate greater transparency of GenAI, code/process workflow and data.** Enterprises should be mindful that, while the marketplace of third-party agents will rapidly accelerate through 2025, there are as yet no trust and reputation mechanisms in this ecosystem creating risk, especially when tools like LLMs dynamically commission third-party AI models. Elsewhere, build “watermarked” modular designs to ensure better isolation of systems components to examine individual and collective emergent behavior.

Evidence

[1] Z. W. Pylyshyn, “The Robots Dilemma: The Frame Problem of Artificial Intelligence,” Praeger, 1987.

[2] The AI Feedback Loop: Researchers Warn of “Model Collapse” as AI Trains on AI-Generated Content, VentureBeat.

[3] Emergent Tool Use From Multi-Agent Autocurricula, arXiv , Cornell University.

[4] Our Approach to Alignment Research, OpenAI.

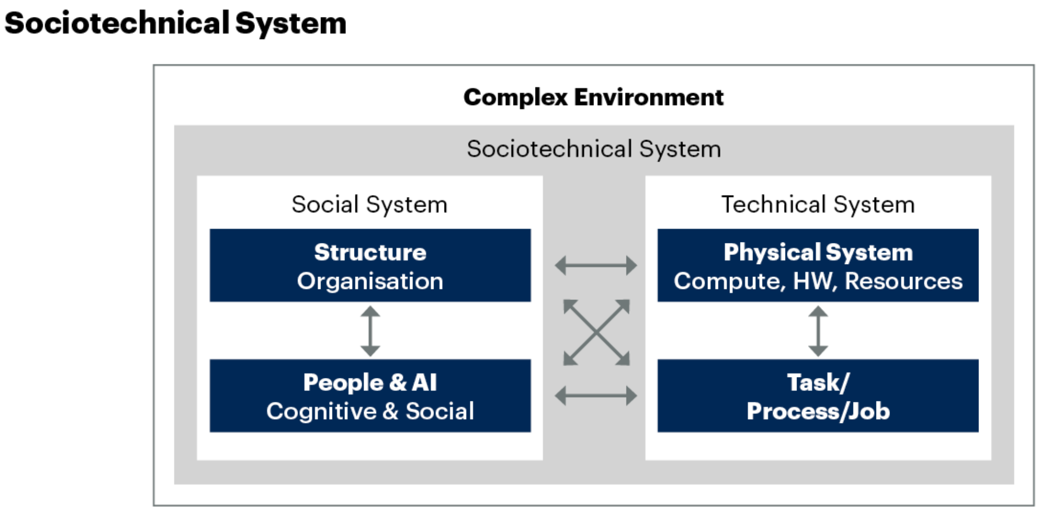
Note 1: What Is a Sociotechnical System?

A sociotechnical system (STS) is a framework that emphasizes the symbiotic relationship between:

- **Social systems:** Formal and informal relationships between individuals and organizations including communication, hierarchy, norms, behaviors, group dynamics and skills in organizations — the “human side” of the organization.
- **Technical systems:** Tools and processes, including hardware, software, and procedures and AI used in work environments — the “technical side” of the organization.

These two systems are not independent; they influence and shape each other in a complex, dynamic interaction. Sociotechnical systems thinking, developed at the UK’s Tavistock Institute, emphasizes the interplay between technology and social factors in workplaces. It was inspired by the introduction of coal mining machinery and the need to consider behavioral changes in work practices.Thus, a sociotechnical system examines how social and technical elements co-evolve and impact system success, efficiency and sustainability.

Figure 13: Sociotechnical System



Source: Gartner
796161_C

IT and business leaders should care about sociotechnical systems approaches due to their key benefits:

- **Technology adoption:** Helps anticipate and manage human responses to new technologies, minimizing resistance and enhancing adoption.
- **System design:** Enables design of work systems that balance human needs and technological efficiency, promoting performance and satisfaction.
- **Innovation:** Supports innovation by fostering a holistic approach to system design and management, boosting long-term success.

Note 2: Design Characteristics to Enable Multiagent Systems

Evolving from ad hoc AI developments to multiagent systems approaches requires a strategic overhaul of many enterprise components. Systems designers should enable the following characteristics in their enterprise architectures for multiagent systems:

- **Shared and personal models of the world:** Shared ontologies are useful for effective collaboration between agents, allowing them to have a shared understanding (or ontology) about the world. This allows them to communicate and make decisions based on common ground — understandable to humans (see Demystifying Taxonomies, Ontologies and Data Models). Individual agents should also maintain their own specific world models based on their unique experiences and roles. This can enable specialized expertise or perspectives; enterprises should strive to differentiate here.

- **Learning and adaptation:** Agents can employ reinforcement learning techniques to understand and adapt to their environment, refining their world models based on feedback from actions taken. Simulation platforms can provide this kind of environment for reinforcement learning. Further, transfer learning can be used to ensure knowledge gained in one context can be transferred to another. If an agent refines its world model in one setting, transfer learning can allow it to apply this knowledge in a different, but related, setting.

- **Handle uncertainty:** Given that agents often operate in environments with uncertainty, their world models should be able to handle and represent probabilistic information. Bayesian techniques can allow agents to update their world models in a principled manner based on new evidence.

- **Feedback loops:** Regularly test the agents' world models against actual outcomes to determine their accuracy. These goals and outcomes should be modeled in a human understandable and AI computable way. Semantic technologies can help here in providing a data modeling technique to capture terms, rules, relationships and outcomes, as well as exploration of cause and effect. Simulated environments can allow for rapid iteration and refinement — something crucial in the early days of multiagent system usage.

- **Feedback from other agents:** Agents should be able to receive and incorporate feedback from other agents to refine their local (versus enterprisewide) world models, facilitating collaborative learning.

- **Ethical and bias considerations:** Ensure that the processes and data sources used to create world models and agents are transparent and "watermarked" or tracked. Today's approach in AI of using a "model card" to represent a model's behavior is evolving. This transparency aids in understanding biases or inaccuracies that might creep in.

Recommended by the Authors

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

Quick Answer: How Will the Generative AI Plug-In Market Evolve?

9 Social and Cultural Implications of Generative AI

Top Strategic Technology Trends for 2023: Adaptive AI

Plan for Generative AI's Impact on Jobs

Quick Answer: How Should CXOs Structure AI Operating Models?

Demystifying Semantic Layers for Self-Service Analytics

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

Innovation Guide for Generative AI Technologies

Innovation Insight: AI Simulation

Table 1: Control Mechanisms Versus Level or Application Within a Sociotechnical Systems Hierarchy

Control Lever		Level Applied to
Amplify/dampen behavior Meter/throttle behavior Delay behavior (by time or resources) Apply limits and quotas on behaviors Semantic gates/rules Factcheck behavior Shape behavior with ground truth Gain consensus on behavior Human interrupt Intentional hallucination (adjust weights or reweight priors)	X	First-party model (e.g., custom price prediction AI model) Second-party model (e.g., partner recommendation algorithm) Third-party model (e.g., ChatGPT or LLM) Composite AI system level (e.g., multimodal chatbot or IoT project) Enterprise AI system of systems (e.g., shared metadata infrastructure) Industry level National level Societal level

Source: Gartner