Introduction to Agentic AI

What is Agentic AI?

Agentic systems are AI systems that can:

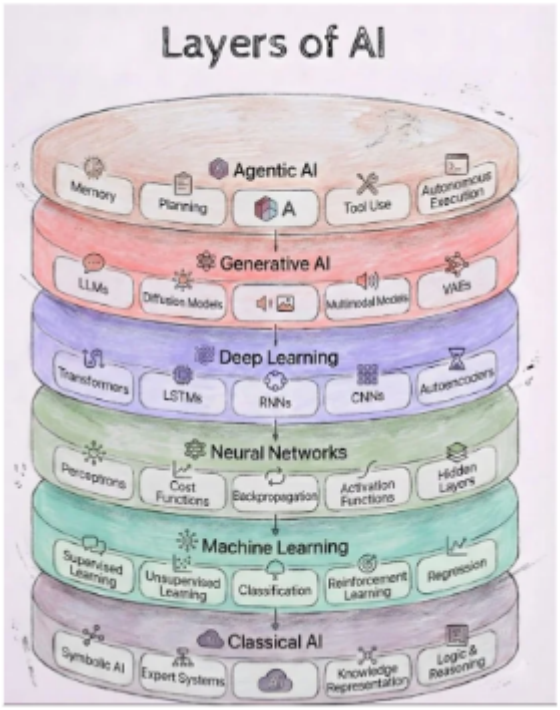
- Maintain internal state
- Plan and decompose tasks
- Use tools
- Retrieve knowledge
- Reflect and self-correct
- Operate under constraints
- Escalate to humans when needed

Controlled Autonomy

Agentic systems should never be unconstrained actors. They should operate with:

- Policy rules
- Approval gates
- Logging
- Role permissions
- Safety checks

AI Layers



The diagram presents AI as a progression of interconnected layers, from traditional rule-based systems to increasingly autonomous and goal-driven AI. Here is a paragraph explaining each layer:

1. Classical AI

Classical AI represents the foundational approaches to artificial intelligence, where systems rely primarily on predefined rules, logical reasoning, and structured knowledge. Rather than learning patterns automatically from large amounts of data, these systems are explicitly programmed with rules that guide their decisions and actions. Key approaches in this layer include symbolic AI, expert systems, logic and reasoning, and knowledge representation. Classical AI is particularly useful in situations where problems are well-defined and decisions need to be transparent and explainable.

2. Machine Learning

Machine Learning introduced the ability for AI systems to learn patterns directly from data rather than relying entirely on manually defined rules. By training algorithms on historical examples, systems can make predictions, classifications, or decisions when presented with new data. This layer includes approaches such as supervised learning, unsupervised learning, classification, regression, and reinforcement learning. Machine learning forms the foundation of many modern AI applications, including recommendation systems, fraud detection, predictive analytics, and image recognition.

3. Neural Networks

Neural Networks are a specialized area of machine learning inspired by the structure of the human brain. They consist of interconnected layers of artificial neurons that learn increasingly complex patterns from data. Neural networks are particularly effective when working with unstructured data such as images, speech, audio, and text. This layer includes concepts such as perceptrons, activation functions, hidden layers, backpropagation, and deep neural networks. Their ability to automatically learn complex representations has made them a major driver of recent advances in AI.

4. Deep Learning

Deep Learning builds on neural networks by using multiple layers to process and learn complex patterns from large datasets. These deeper architectures allow AI systems to perform advanced tasks such as understanding language, recognizing objects, processing speech, and making predictions. Common deep learning architectures include Transformers, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), and autoencoders. Deep learning provides much of the underlying technology behind today's advanced AI systems.

5. Generative AI

Generative AI represents a major shift from AI systems that primarily analyze or classify information to systems capable of creating entirely new content. Using large models trained on vast amounts of data, generative AI can produce text, images, audio, video, software code, and other forms of content. This layer includes Large Language Models (LLMs), Generative Adversarial Networks (GANs), diffusion models, Variational Autoencoders (VAEs), and multimodal models. Generative AI enables more natural interaction between humans and machines and powers applications such as AI assistants, content generation tools, and creative systems.

6. Agentic AI

Agentic AI represents a more autonomous form of AI in which systems can pursue goals, make decisions, plan actions, use tools, and interact with external environments. Unlike a traditional generative AI system that simply responds to a prompt, an AI agent can break down a task into smaller steps, determine what actions are required, access relevant tools or information, and adapt its approach based on results. This layer includes key capabilities such as memory, planning, reasoning, tool use, autonomous execution, and multi-agent systems. Agentic AI moves AI closer to functioning as an active digital collaborator capable of completing complex, multi-step tasks with reduced human intervention.

AI Factory Stack

The **AI Factory Stack** illustrates the key components required to transform a general-purpose Large Language Model (LLM) into a reliable, production-ready AI system. Each layer adds a specific capability, moving from the core intelligence of the model to the infrastructure, governance, and operational mechanisms needed to deliver real-world AI applications.

1. LLM — The Machine

The Large Language Model (LLM) is the core intelligence of the AI Factory Stack. It is responsible for understanding language, generating responses, reasoning over information, and creating content. An LLM can answer questions, summarize documents, write content, and support conversations, but on its own, it is limited to the knowledge available to it and cannot reliably access private, current, or organizational information. It therefore serves as the foundational “machine” upon which the rest of the AI system is built.

2. RAG — The Raw Material Room

Retrieval-Augmented Generation (RAG) provides the LLM with access to external knowledge and organizational information. Instead of relying only on what the model learned during training, RAG retrieves relevant documents, policies, records, or other data sources at the time a question is asked. This retrieved information is then provided to the LLM to generate a more accurate and contextually relevant response. RAG effectively gives an AI system access to the knowledge it needs to perform specialized tasks.

3. Vector Database — The Storage Warehouse

The Vector Database stores information in a format that allows AI systems to search based on meaning rather than just exact keywords. Documents and other content are converted into numerical representations called embeddings, which capture their semantic meaning. When a user asks a question, the system searches the vector database for information that is conceptually related to the query. This makes it a critical component of RAG systems, enabling fast and intelligent retrieval of relevant knowledge.

4. AI Agent — The Floor Manager

The AI Agent acts as the decision-making and coordination layer of the system. While an LLM primarily generates responses, an agent can determine what steps are required to achieve a specific goal. It can plan tasks, decide which tools or information sources to use, execute actions, and evaluate the results before moving to the next step. In this sense, the AI agent acts like a floor manager, coordinating different components of the AI system to complete more complex, multi-step tasks.

5. MCP — The Power Outlet Standard

The Model Context Protocol (MCP) provides a standardized way for AI systems to connect with external tools, applications, databases, and services. Instead of creating a separate custom integration for every tool, MCP establishes a common interface through which AI agents and models can access capabilities such as APIs, databases, files, and enterprise systems. This makes AI applications more modular, interoperable, and easier to expand as new tools and services are added.

6. Guardrails — Safety Systems

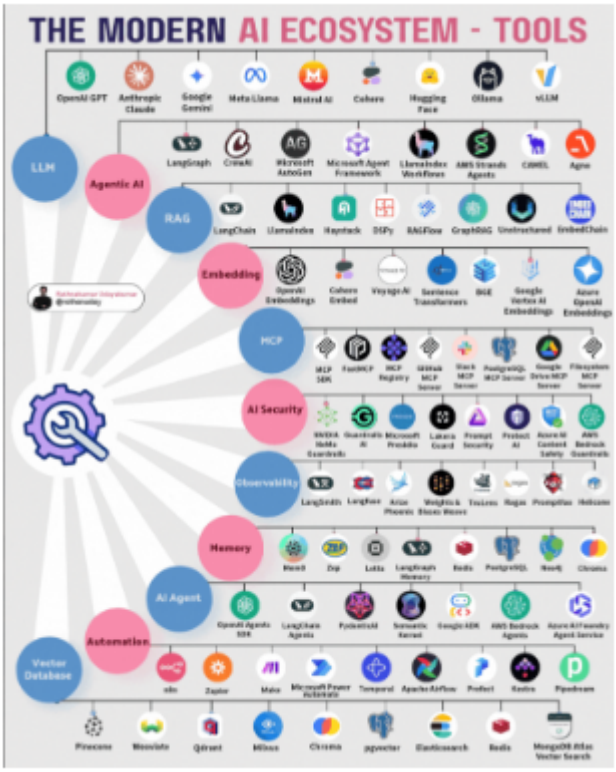
Guardrails provide the controls and protections needed to ensure that AI systems operate safely and responsibly. They can help prevent harmful or inappropriate outputs, protect sensitive information, enforce organizational policies, and restrict the actions an AI agent is allowed to perform. Guardrails are particularly important when AI systems interact with real users, confidential data, or external systems. They ensure that increased AI capability and autonomy are balanced with security, governance, and human-defined boundaries.

7. Evals — Quality Check

Evaluations, or Evals, are used to measure how well an AI system performs against defined quality standards. They test factors such as accuracy, relevance, reliability, safety, consistency, and task completion. Rather than assuming that an AI application is working correctly, Evals provide a structured way to test its performance using representative scenarios and benchmarks. They are essential for identifying weaknesses, comparing changes, and ensuring that improvements to the AI system do not introduce new problems.

Overall, the AI Factory Stack shows that a successful AI product is much more than an LLM. The LLM provides the intelligence, RAG and vector databases provide knowledge, agents provide decision-making and action, MCP enables connectivity, guardrails provide safety, and evaluations ensure quality. Together, these layers form the foundation for building AI systems that are useful, reliable, scalable, and ready for real-world deployment.

AI Tools



The **Modern AI Ecosystem – Tools** diagram presents the modern AI landscape as a collection of interconnected layers and capabilities. Rather than being built around a single model or platform, modern AI systems combine specialized tools for intelligence, data retrieval, storage, automation, safety, observability, and deployment. Each category plays a different role in turning AI models into complete, production-ready solutions.

1. LLM — The Core Intelligence

Large Language Models (LLMs) form the foundation of the modern AI ecosystem. These models are responsible for understanding and generating human language, reasoning over information, answering questions, writing content, and supporting conversations. The ecosystem includes a variety of commercial and open-source models, allowing organizations to choose models based on factors such as performance, cost, privacy, language support, and deployment requirements. However, an LLM alone is rarely enough to build a complete AI solution, which is why it connects to the other layers in the ecosystem.

2. Agentic AI — Autonomous Task Execution

Agentic AI introduces the ability for AI systems to move beyond simply generating responses and instead take actions to achieve specific goals. AI agents can break down complex tasks, make decisions, plan workflows, access tools, retrieve information, and execute multiple steps with limited human intervention. The growing number of agent frameworks and platforms reflects the increasing importance of building AI systems that can act as digital workers, assistants, and autonomous problem-solving systems.

3. RAG — Connecting AI to Knowledge

Retrieval-Augmented Generation (RAG) enables AI systems to access external and organization-specific knowledge. Instead of relying only on the information contained within an LLM, RAG retrieves relevant documents, databases, or knowledge sources when a user asks a question. This information is then provided to the model as context for generating a response. RAG is especially important for enterprise AI applications because it allows organizations to build assistants that can answer questions using current, private, and domain-specific information.

4. Embeddings — Representing Meaning

Embedding technologies convert text, images, and other forms of data into numerical representations that capture their meaning and relationships. Instead of searching for exact keywords, AI systems can use embeddings to identify information that is conceptually similar. Embeddings form a critical part of semantic search and RAG architectures, helping systems understand that different words or phrases may express similar ideas. They provide the bridge between human information and the mathematical representations used by AI systems.

5. MCP — Standardized AI Connectivity

The Model Context Protocol (MCP) represents an emerging approach to connecting AI models and agents with external tools, data sources, and services through standardized interfaces. Rather than building a completely new integration for every application, MCP aims to make it easier for AI systems to discover and use external capabilities. This creates a more modular ecosystem where models and agents can interact with databases, APIs, files, enterprise systems, and other digital tools in a consistent way.

6. AI Security — Protecting AI Systems

AI Security focuses on protecting AI models, applications, data, and users from risks and vulnerabilities. As AI systems gain access to sensitive information and external tools, security becomes increasingly important. AI security tools can help organizations manage issues such as prompt injection, unauthorized access, data leakage, malicious inputs, insecure integrations, and model misuse. This layer ensures that AI systems are not only capable but also protected against threats.

7. Observability — Understanding AI Performance

Observability tools allow developers and organizations to monitor what happens inside an AI application. They can track prompts, model responses, tool calls, retrieval quality, latency, token usage, errors, and overall system behavior. This visibility is essential when deploying complex systems involving LLMs, RAG pipelines, and AI agents. Observability helps teams identify problems, understand system behavior, optimize performance, and continuously improve their AI applications.

8. Memory — Giving AI Context Over Time

Memory enables AI systems, particularly agents and conversational applications, to retain useful information across interactions. Instead of treating every request as completely independent, a system with memory can maintain context about previous conversations, user preferences, tasks, or actions. Memory can be short-term, supporting the current conversation, or long-term, allowing the system to retain relevant knowledge over time. This capability is essential for creating more personalized, consistent, and context-aware AI experiences.

9. AI Agents — Coordinating Tools and Actions

The AI Agent layer focuses on frameworks and platforms used to build systems that can reason, plan, use tools, and execute workflows. Agents sit at the center of many modern AI architectures because they coordinate multiple capabilities, including LLMs, RAG systems, APIs, memory, and external applications. An agent can determine what needs to be done, select the appropriate tools, execute actions, evaluate the results, and continue until a task is completed.

10. Automation — Integrating AI into Workflows

Automation platforms connect AI capabilities with business processes and digital workflows. They allow organizations to trigger actions based on events, connect multiple applications, move information between systems, and automate repetitive tasks. When combined with AI agents, automation tools can enable systems to not only understand and generate information but also take meaningful action across an organization's technology environment.

11. Data Infrastructure — The Foundation for AI

Modern AI systems depend on strong data infrastructure. This includes databases, data warehouses, APIs, storage systems, and platforms that make information available to AI applications. Data infrastructure ensures that AI systems can securely access, process, retrieve, and manage the information they need. Without reliable data systems, even the most advanced AI

models have limited ability to provide useful and accurate results.

12. AI Evaluation — Measuring Quality and Reliability

AI evaluation tools are used to measure how effectively an AI system performs. They assess factors such as accuracy, relevance, faithfulness, safety, consistency, latency, and task completion. Evaluations are particularly important because AI systems can produce outputs that appear convincing but may be incorrect or unreliable. By systematically testing AI applications, organizations can identify weaknesses, compare different models or approaches, and improve system quality over time.

13. AI Governance — Managing Responsible AI

AI Governance provides the policies, controls, and accountability mechanisms needed to manage AI responsibly. It addresses issues such as compliance, privacy, transparency, risk management, model usage, data governance, and human oversight. As AI becomes more deeply integrated into organizational processes, governance ensures that its use aligns with legal requirements, ethical principles, and organizational objectives.

14. AI Safety — Setting Boundaries and Guardrails

AI Safety focuses on ensuring that AI systems behave within acceptable and defined boundaries. Safety tools and frameworks can help prevent harmful outputs, restrict unauthorized actions, protect sensitive data, and enforce organizational policies. This becomes particularly important for autonomous agents that can access external systems and perform actions. AI safety ensures that greater autonomy does not result in uncontrolled or unpredictable behavior.

Overall Perspective

The diagram shows that **modern AI is not just about choosing the best LLM**. Building a real-world AI solution requires an ecosystem of technologies working together. LLMs provide intelligence; RAG and embeddings provide access to knowledge; data infrastructure stores and manages information; agents and automation enable action; MCP connects systems; memory provides continuity; observability and evaluation measure performance; and security, governance, and safety ensure responsible operation.

Together, these layers form the **modern AI ecosystem**, transforming a standalone AI model into a complete, scalable, secure, and production-ready AI system.