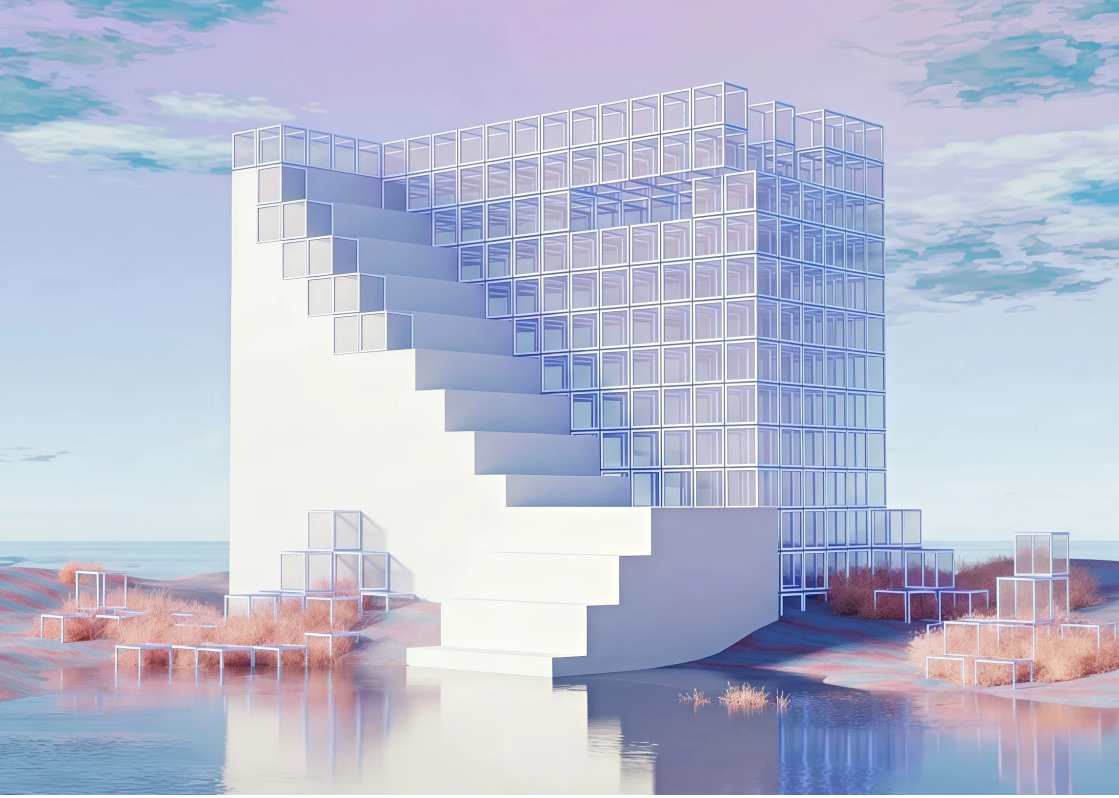


The agentic enterprise: Building an ecosystem of continuous evolution and reliable impact.

/thoughtworks

Design. Engineering. AI.

aws PARTNER AI Services Competency

- Agentic AI Consulting Services
- Generative AI Consulting Services

The collapse of traditional transformation.

Most transformation strategies start with defining an ideal state: a target that shapes every action on the organization's journey. But in the age of AI, this model is obsolete.

AI is progressing rapidly, with the length of tasks large language models (LLMs) can complete doubling every seven months.[1] But these capabilities are uneven, creating a "jagged frontier" where it's difficult to predict exactly how AI models improve with each new model iteration compared to humans performing a similar task.[2]

Agentic AI: From task automation to outcome orchestration.

While early AI implementations have mostly focused on predefined task automation using workflows, agentic AI is focused on outcomes. It's a more independent and proactive approach, needing less human intervention to complete complex tasks.

1 https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/
2 https://www.hbs.edu/faculty/Pages/item.aspx?num=64700

Instead of pre-defining a complex process for AI to follow, it's about defining a goal and a set of guardrails and allowing the AI to find the best path to success. The specific steps to reach the goal are non-deterministic, but they can deliver more complex and powerful outcomes.

Most current AI deployments are narrow. Individual people are using AI tools to complete their daily tasks faster, or with less manual effort. But that only offers marginal productivity gains. The major potential value of AI lies in higher levels of autonomy and intelligence, when organizations are able to apply its capabilities at broader, more impactful levels. Think teams, processes and functions supported by AI.

But there's a gap between that one model, one output position and an organization-wide deployment, and businesses won't be able to bridge it following the traditional digital transformation playbook. By the time organizations follow the traditional route to identify the challenge, design and build a solution and manage the deployment, it's increasingly likely that AI will have evolved and their original target state needs to be revisited. Teams risk investing time and budget in a project that will deliver minimal value, or could be unfit for purpose.

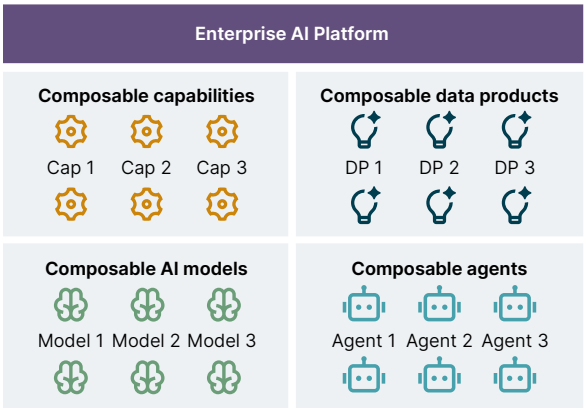
With no fixed future, organizations need a strategy that accommodates an evolving set of tools. Teams must move from "building to a blueprint" to "building for adaptation," with the understanding that the industry is frequently redrawing the execution and delegation lines between humans and AI.

The new business architecture.

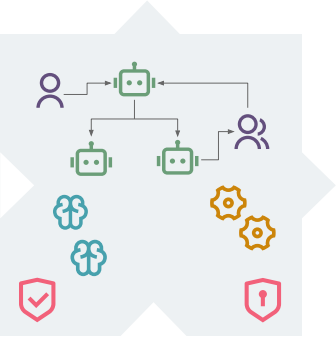
The question is, if real AI transformation is a moving target, how can organizations achieve their goals?

By breaking down functional silos and building flexibility into the way processes are designed and built, organizations can stay adaptable as they explore ways to support and augment their workforce with AI's evolving capabilities. We call it the agentic enterprise.

This composable strategy transforms how core systems, data products, AI models and agents can be combined into modular components that work independently. They can be assembled and reassembled into capabilities that evolve the processes performed by humans and AI, and that can be exposed through different experiences — text, voice, video, AR — where that human-AI collaboration takes place.



A new enterprise architecture paradigm to breakdown an enterprise into independent and composable capabilities, data products, AI models and agents



They can be assembled to build enterprise agentic capabilities, built using agile and AI-assisted methods

Building an effective agent.

In an agentic enterprise, human experts set the strategy and parameters, and handle any ambiguous, high-complexity exceptions, while agents are empowered to make decisions and learn from their environment and activities. Together, they create a process orchestration layer where they collaborate for widespread productivity gains that far exceed the incremental improvements of traditional task automation.

To be successful, each agent should be assembled from five components:

Models	Data	Tools	Orchestration	Governance
For planning, reasoning, evaluating and learning	For memory, context, facts and knowledge	For acting and reacting to the real world	For collaboration with humans and other agents	For safety, bias, ethics, security, cost control and quality

Models

AI models bring intelligence to agents. Whether you use a fine-tuned domain-specific AI model or the wealth of general-purpose models available, they're the core of an effective agent.

Platforms such as Amazon Bedrock provide access to these foundational models, which give agents the ability to reason, plan, evaluate different options, make data-led decisions and learn and improve over time.

Amazon Bedrock provides access to hundreds of foundation models from leading AI companies. That means organizations can choose the best model for their specific use case, based on performance and cost. And as needs evolve and new models emerge, teams can easily switch between models to future-proof their AI strategy, all within a single platform with built-in evaluation tools.

Organizations can also use Bedrock to customize models while maintaining enterprise-grade security and compliance. The platform's comprehensive capabilities allow teams to fine-tune domain-specific models or leverage general-purpose models, ensuring agents have the intelligence they need to reason, plan and make data-led decisions at scale.

Data

Data provides the knowledge and context agents need to operate within the organization. Certain tasks will require factual information rather than LLM-generated content alone, so being able to connect your agents with data products helps improve the reliability and performance of agentic workflows.

Tools

Tools connect agents to enterprise systems, enabling them to execute actions in the real world based on a dynamic understanding of their goals. Exposing your core APIs and

microservices, or modernizing your legacy systems for AI become additional levers to improve reliability, so agents can perform deterministic tasks in your systems. This improves overall reliability as you don't need to rely only on the non-deterministic output from LLMs in the agentic workflow.

Orchestration

In a complex process, agents won't operate in isolation. The orchestration capability is fundamental for sharing agent-to-agent, agent-to-human and human-to-agent workloads. It keeps humans in the loop while allowing agents to work across different functions with shared needs.

Governance

Policies and guardrails are a vital governance layer for ensuring agents behave as expected and adhere to organizational policies and compliance demands. Measures such as content filters, input validation and frequent output audits can help ensure agents are operating responsibly. These guardrails are also a useful tool for making end-users more comfortable working alongside AI agents.

Deploying agents with Amazon Bedrock AgentCore

Amazon Bedrock AgentCore helps organizations deploy agents in production environments by providing a comprehensive agentic platform that eliminates infrastructure management complexity. It offers serverless agent deployment with complete session isolation to prevent data leakage, supporting both low-latency conversations and long-running asynchronous workloads

of up to eight hours. The platform enables deployment through code upload or containers, automatically scaling to hundreds of sessions as needed.

AgentCore accelerates production deployment through enterprise-grade security features, including secure agent identity and access management that allows agents to access AWS resources and third-party services. The platform provides real-time monitoring and observability through Amazon CloudWatch integration, enabling teams to trace, debug and monitor agent performance in production. Additionally, it includes continuous evaluation capabilities that sample and score live interactions to improve agent quality over time, making it easier for organizations to maintain production-ready agents at scale with confidence.

Beyond these core services, AgentCore also includes *Identity* for secure multi-identity-provider authentication, *Policy* for defining and enforcing agent-level access controls, *Browser* for enabling agents to interact with web-based interfaces, and *Code Interpreter* for executing code within isolated environments — further extending the platform's enterprise-grade capabilities.

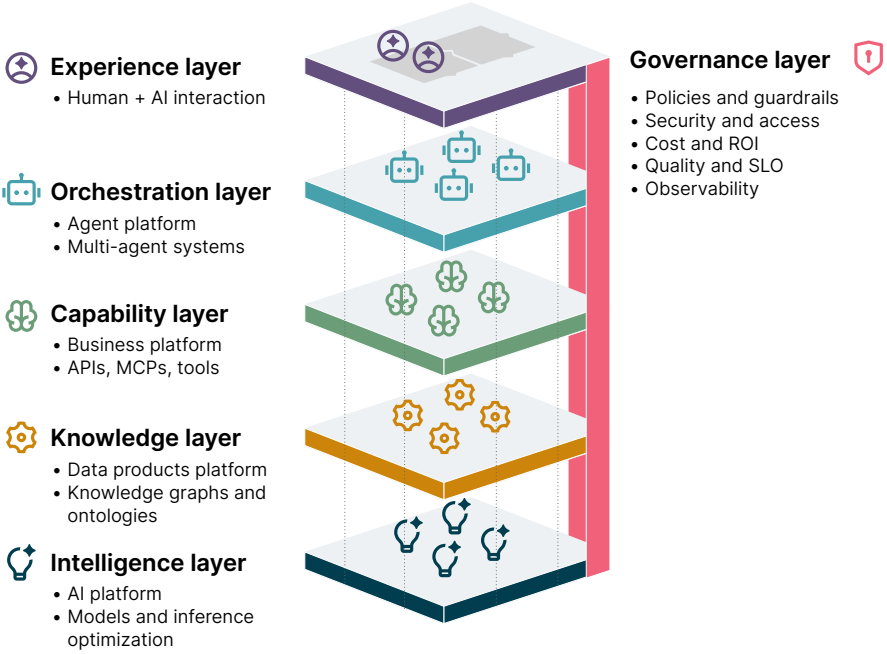
Adaptability in the agentic enterprise.

Agents are one of the core building blocks for agentic capabilities, where:

Humans and agents work together to automate and orchestrate processes that support business KPIs, using **domain-specific models** and data products to plan, reason and execute with **APIs**. At every step, **policies** and **guardrails** provide the governance controls needed to manage cost, security and quality.

These agentic capabilities can be composed in different ways to serve a huge variety of business use cases. With multiple agents, teams can automate and orchestrate across entire functions and, crucially, between functions. The efficiency and value produced will be outsized compared to individual productivity gains when agents work together crossing traditional organizational boundaries, creating connections between departments such as marketing, HR, finance and procurement.

For agentic capabilities to make a real impact, they need to be supported by layers of foundational architecture that align with the key components outlined above. Enterprises will need to manage platforms for data, business capabilities, AI/ML models, agents and explore how to expose rich human-AI experiences.



The continuous improvement loop.

As the AI frontier continues to advance, agents need to evolve to offer improved capabilities and match the pace of changing business demands. The modular nature of an agentic AI system means that agents are as simple to adapt as they are to build.

In an agentic enterprise, there are constant opportunities for improvement and, crucially, these opportunities can be triggered from any source. Business needs, user experience, technology and even data itself can be the driver for innovation and change.

A top-down, use case-focused view can solve the visible challenges of users' everyday experiences, while a bottom-up, data-led view can uncover hidden areas for agents to deliver value. Data flows can be abstract, but they represent the reality of an organization's operations; analyzing from this perspective uncovers processes that are ideal candidates for automation, such as those that are data-rich without needing human input. This also highlights the importance of building a solid data foundation to benefit from AI.

No end state. The evolving "jagged intelligence frontier" redraws the delegation boundary between human and AI constantly.

Workflows need to adapt, requiring people's roles to evolve to supervise and delegate at higher, more strategic levels. Process follows tech and data.

Enhancing agents in live production environments with AgentCore.

AWS facilitates continuous improvement of agents through Amazon Bedrock AgentCore's monitoring and evaluation capabilities, which include:

- The Evaluations service, which samples and scores live interactions using built-in and custom evaluators.
- The Observability service, which provides dashboards powered by Amazon CloudWatch for tracing, debugging and monitoring agent behavior, as well as issue detection and analysis through OpenTelemetry integration.
- The Memory service, which maintains context across interactions and builds knowledge that improves performance over time.
- The Gateway service, which enables semantic search for tool discovery, helping agents access more capabilities and data sources.
- The serverless Runtime service, which supports rapid deployment of improvements without infrastructure management, handling workloads from low-latency conversations to eight-hour asynchronous tasks.

The reliability gap: Why most AI transformations fail.

Despite the growing power and potential of AI in the enterprise, most organizations have yet to see value from their pilot projects. Successfully deploying enterprise-grade AI is difficult, with progress often slowed by challenges such as poor integration within the enterprise, governance and compliance blockers, a lack of capabilities needed to deploy at scale and low trust, especially for generative AI tools.[3]

This challenge stems from being too focused on what's possible and not focused enough on deploying AI that really works. If leaders want consistent adoption, which is vital for realizing value at scale, AI systems must be reliable, focused on real pain points and tightly integrated with enterprise systems.

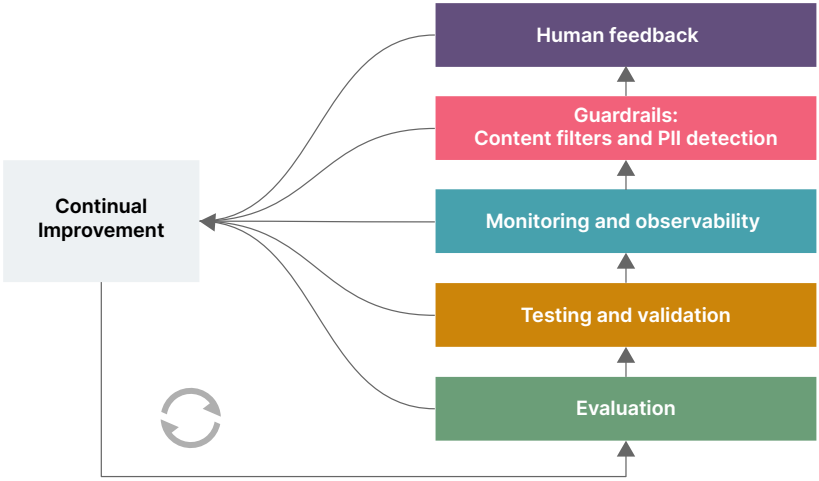
People need to see that agents offer a consistent, meaningful improvement to their workflows. Impressive capabilities aren't enough to convince teams to trust agents with their data and decisions when the value proposition is unclear.

A framework for AI reliability.

At Thoughtworks, we've created a framework of rigorous analytical and engineering techniques for building reliable AI

3 https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/

systems. This iterative process ensures AI is fit for purpose and ready to deploy at scale.



The framework is a five-stage, continuous process that uses feedback loops to drive performance and reliability improvements, both during the build phase and after the AI system is deployed and used in production.

1. For **evaluation**, we develop an understanding of how the AI system should perform as a whole, in our specific use case. This helps assess performance beyond the public benchmarks, while also allowing us to understand how different combinations of new AI models and our prompts behave when recombined.
2. In **testing and validation**, we compare the system behavior against expectations, to ensure it is delivering the right outputs or achieving the desired goals.

3 **Observability** is key for effective monitoring; we log every interaction associated with a user, agent or system to assess end-to-end performance, to ensure every step and action can be explained, and to reconstruct traces and troubleshoot when occasional problems occur.

4 **Guardrails** provide strict operating parameters for AI tools and agents to protect people, data and security. It's also where governance policies are inserted, to ensure the appropriate checks and controls are in place for a fast, reliable user experience while adhering to enterprise compliance.

5 Keeping **humans in the loop** allows us to gather feedback from real users, monitor interactions and annotate data to ensure the AI system learns and improves over time.

This framework is derived from more than 30 years of software engineering excellence, combined with ongoing Thoughtworks AI research and experimentation in reliability areas such as evaluations, explainability and robustness.

Our approach complements the AWS framework for responsible AI, which is built on eight dimensions:

Fairness	Explainability	Privacy and security	Safety
Controllability	Veracity and robustness	Governance	Transparency

To ensure reliability in its own AI models and services, AWS conducts independent assessments of service fairness, explainability, robustness and privacy and security. Tools like Amazon Bedrock Guardrails block up to 88% more harmful

content and filter over 75% of hallucinated responses, while Model Evaluation helps assess and select the best foundation models for each use case.

Additionally, AWS publishes AI Service Cards that provide transparency through information on intended use cases, limitations, responsible AI design choices and performance optimization best practices for its AI services and models.

The path to scaling agentic AI.

Decades of scoping, designing, deploying and evolving technology implementations have led us to create an effective, structured approach to scaling AI that helps organizations become agentic enterprises.

Plan and pilot

We work alongside your teams to assess organization-wide AI readiness and plot an adaptable strategy for agentic AI.

By identifying the right use cases for AI agents, we help your teams prioritize deployment and focus on a high-impact pilot.

We also assess and establish the foundational platform and architecture to enable scaled impact.

Establish and expand

We design and build AI agents, defining the necessary guardrails, evaluations, model selection, tools, connectors, processes and data foundations.

Working in thin slices, we can then extend agentic AI use cases to support more functions and departments.

With more platform enterprise capabilities established, the time-to-value for each new use case reduces and you get to value faster.

Scale and evolve

With agents in use across multiple functions and delivering value, we create a flexible long-term roadmap for continued operations and evolution of existing systems.

We continuously optimize these systems across the entire AI enterprise stack, beyond the application layer, using performance engineering to improve cost, throughput and quality.

Start with a thin slice.

Without focus, ambitious projects derail, innovation stalls and colleagues lose faith in the transformation.

Thin slices are vertically integrated working packages across the entire agentic enterprise stack. They're narrow, but complete deployments that deliver value from the very beginning to act as a proof of concept while driving the platform capabilities forward to accelerate further thin slices.

A successful thin slice usually tackles a high-visibility, high-impact use case, such as creating a customer support agent that connects to multiple data sources to bring the full picture of a customer, and uses appropriate tools to triage and automatically resolve issues.

The agentic enterprise in action.

Accelerating research and development at Bayer.

Thoughtworks initially created Bayer's Preclinical Information Center, or PRINCE, as a cloud-based data integration platform on AWS.

We then set about expanding its capabilities, first with an LLM-powered conversational assistant and now with a multi-agent AI system. PRINCE's agentic components can complete complex tasks such as planning and writing, as well as evaluate findings and provide precise citations.

Bayer's scientists can now focus on designing and trialing new treatments with fast, easy access to insights from more than 18,000 research reports.

Learn more

Supporting appeals for denied medical insurance claims with Nayya.

Every year, US medical insurers deny 850 million claims, but very few people have the tools and support they need to appeal these decisions.

Thoughtworks collaborated with Nayya to create the Nayya Claims Advocate, an agentic AI application that automatically creates an appeal letter. AWS Bedrock, including its Knowledge

Bases, provides a rigorously secure and highly capable environment for the agent to assess the circumstances of a claim and build an optimized response.

Learn more

Build an architecture where humans and AI can evolve together.

Agentic AI is gaining traction, transforming organizations and developing new capabilities all the time. Becoming an agentic enterprise can help leaders augment their workforce's expertise, break down silos between functions and deliver more value — all with the agility to adapt to whatever comes next.

To maximize the impact of agentic AI, organizations need to start by building a solid foundation, establishing the use cases, data sources, platforms and governance that shape and guide agents.

Together with Thoughtworks and AWS, you can reimagine business processes and the way your teams collaborate with technology and each other, supported by industry-leading expertise, decades of experience and the most advanced and flexible AI tools on the market.

What's next?

- Discover your AI readiness with our personalized assessment, and see whether your organization has what it takes to maximize the value of AI.
  Get your report
- Reach out to discuss your agentic enterprise transformation with our expert team, and explore how Thoughtworks and AWS can accelerate your journey.
  Let's talk about agentic AI

Authors

Author

Danilo Sato

Global VP of AI
Thoughtworks

Contributor

Baran Karlidag

Senior Manager, Specialist Partner Solution Architects, EMEA
AWS

We are a global technology consultancy that delivers extraordinary impact by blending design, engineering and AI expertise.

For over 30 years, our culture of innovation and technological excellence has helped clients strengthen their enterprise systems, scale with agility and create seamless digital experiences.

We're dedicated to solving our clients' most critical challenges, combining AI and human ingenuity to turn their ambitious ideas into reality.

thoughtworks.com

/thoughtworks

Design. Engineering. AI.