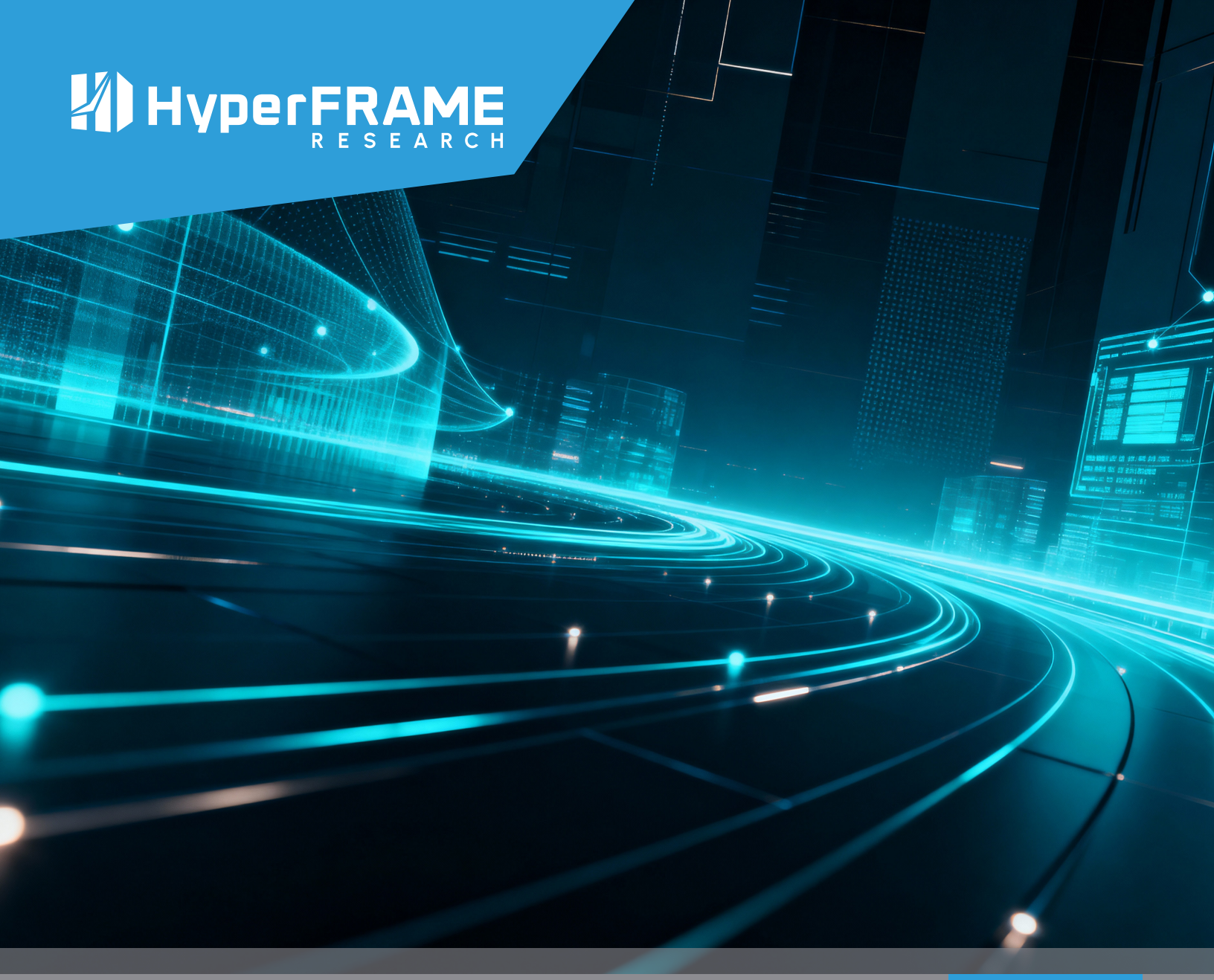
HyperFRAME RESEARCH



WHITE PAPER

Operationalizing Multi-Agent Systems with Amazon Bedrock AgentCore

How Organizations Can Move Beyond LLM Pilots and Build Secure, Scalable, and Composable Agent Workflows

Authors:

Stephanie Walter

Practice Leader - AI Stack

NOVEMBER 2025

Executive Summary

The emergence of agentic AI represents a genuine step forward in technology, but its promise has been difficult to fully realize. While AI pilots are common, most organizational deployments remain stuck in the experimental phase. As organizations attempt to scale these systems, they can encounter deep fragmentation, critical security and compliance concerns, and limited operational visibility across agent performance.

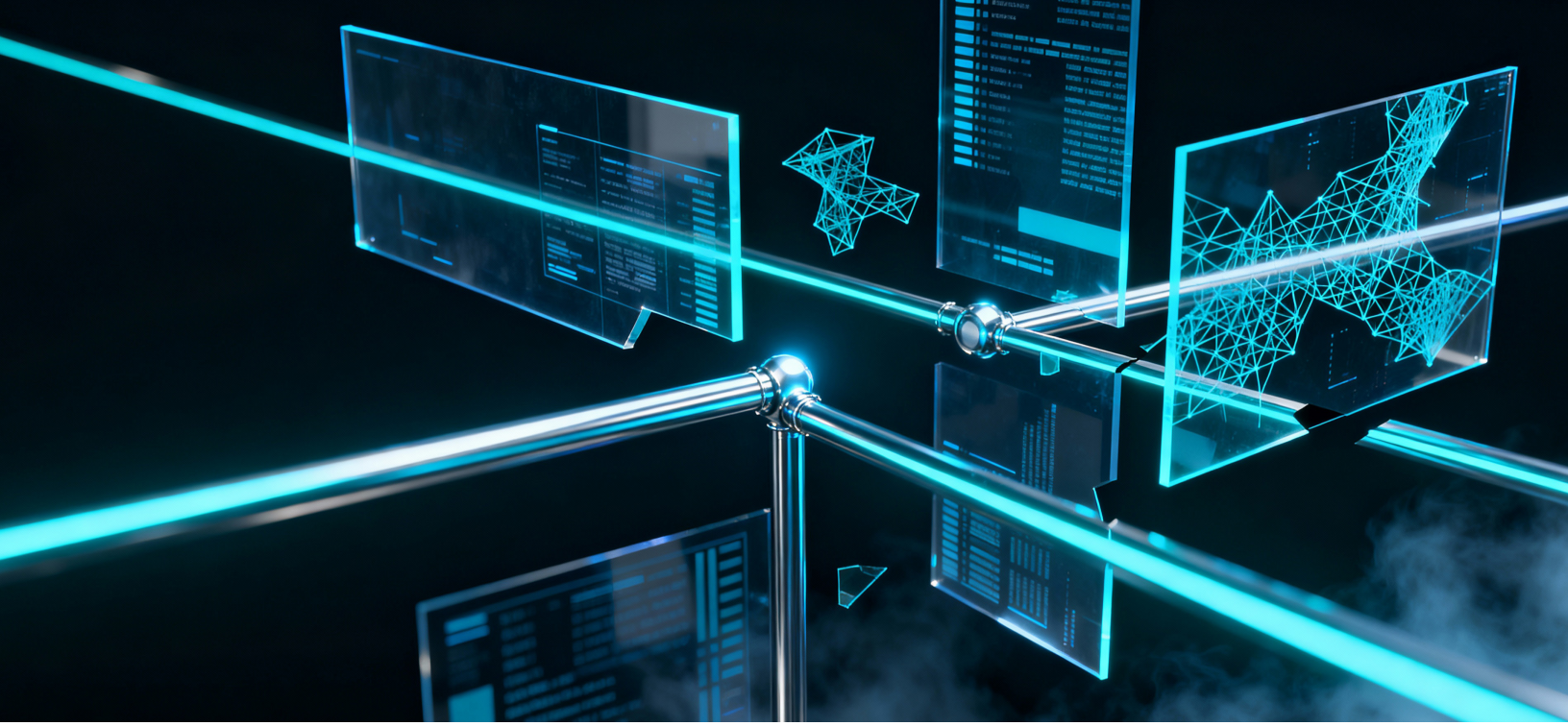
Amazon Bedrock AgentCore is architected to address this challenge head-on. It aims to deliver a unified platform designed to operationalize agents with the necessary governance, end-to-end observability, and enterprise-grade control organizations demand. This platform is crucial to deploy and operate highly capable agents at scale. By offering a comprehensive set of services, AgentCore enables the essential leap from isolated experimentation to robust, production AI within the technology stack. This white paper explores precisely how AgentCore serves as the foundation for this next phase of AI adoption.

The Rise of Agentic AI

For years, AI has focused on integrating powerful LLMs, primarily for tasks like summarization and content generation. However, the market is quickly moving to embrace agentic AI. An AI agent is a system that leverages an LLM not just for generating text, but for reasoning, planning, and taking autonomous action within an environment. The core components of this shift are LLMs combined with tool use and a defined workflow. This means that AI is no longer a passive answering machine; it's an active decision-making system.

This concept is gaining significant traction because it enables complex, multi-step business process automation that was previously impossible. Instead of writing a complicated sequence of integrations across ERP, CRM, and supply chain systems, an organization can now instruct an agent: "Check current inventory levels, evaluate open sales orders, and, if supply is below forecasted demand, automatically trigger a purchase requisition and alert the finance team for approval." The agent isn't just analyzing data, it's making conditional decisions, coordinating across applications, and executing actions based on real-time business context.

Market signals underscore this shift, shown by the proliferation of open-source orchestration frameworks like Strands Agents, LangChain, CrewAI, and LangGraph. These frameworks are excellent for initial prototypes and development efforts. The DIY approach to building production systems, however, quickly runs into walls. Hardcoding tools directly into agent logic introduces rigidity. Teams face difficulties implementing essential controls into agent workflows and integrating necessary security primitives. Furthermore, a lack of deep, end-to-end observability and agent evaluations in a self-built system prevents teams from understanding why an agent chose a specific action, which is a non-starter for regulated industries. AgentCore is designed to solve this production chasm, making AI agents meaningful for business value.



Why Most Organizations Struggle to Scale Agents

HyperFRAME Research is seeing that the largest hurdle for organizations is not building the first agent, but moving from a proof-of-concept (POC) to a reliable, full-scale production environment. The path from "excitement and potential" to "meaningful business value" is hindered by a chasm of critical challenges, including performance, security, governance, and scalability. Teams often see the following complications when deploying to production:

- **Sprawl and shadow AI.** Individual business units or development teams can launch disconnected experiments with different open-source frameworks and deployment methods. This results in sprawl, where overlapping agent functionality is deployed without central oversight. This can rapidly become "shadow AI" use, where agents access sensitive systems without IT or security department awareness, an unacceptable risk for the modern organization.
- **Security and compliance failures.** When agents are deployed without enterprise-grade security, they become potential entry points into critical systems. An agent lacking proper access controls and permissions is at the risk of performing unauthorized operations. A production agent needs distinct identity and controlled access.
- **Operational blind spots.** A core challenge is the lack of visibility into an agent's reasoning process. Without granular observability, teams cannot determine why an agent took a certain action, executed a tool, or failed a task. This absence of an audit trail or debug-ready traces cripples quality audits and slows down time-to-market.
- **Tooling chaos.** Early agent experiments often involve hardcoding tools or APIs directly into the agent's logic or prompt. This tightly couples the agent to a specific tool, limiting reusability and introducing significant maintenance overhead. True enterprise scale requires an abstraction layer that treats tools, Retrieval Augmented Generation (RAG) pipelines, and APIs as reusable assets, decoupled from the agent's core decision loop.
- **Lack of flexibility and integrations.** Many organizations discover that early agent frameworks lock them into proprietary tools or limited model choices, which restricts experimentation and innovation. True adoption requires the flexibility to integrate preferred frameworks such as CrewAI, Google ADK, LangGraph, LlamaIndex, OpenAI Agents SDK, and Strands Agents, along with access to a broad range of foundation models available on Amazon Bedrock or external providers like OpenAI and Gemini. With AgentCore, organizations can choose the services their teams need while continuing to use their preferred frameworks and models. This maintains both innovation freedom and governance.

AgentCore: A Platform Built for Agent Operations

Amazon Bedrock AgentCore represents an agentic platform comprised of fully managed services and designed to overcome the challenges of agent operationalization. It is designed to serve as the foundational agentic AI layer, providing the necessary infrastructure, security, and lifecycle management for production agents.

AgentCore is architected to support the development and deployment of various agent frameworks, including popular open-source options like Strands Agents, LangChain, LangGraph, and CrewAI. AgentCore Runtime allows agents to be deployed using common open-source frameworks, accelerating time-to-market. It also supports A2A and is model, protocol, and framework agnostic. Both real-time and multi-hour workloads can have true session isolation, providing the necessary security and scalability for multi-step tasks that last up to 8 hours.

Security with AgentCore is not a bolt-on feature, but a core architectural mandate. The identity service is designed to deliver secure, delegated access for AI agents. This is achieved by:

- Assigning distinct agent identities for secure operation at scale.
- Enabling agents to securely access third-party tools (like GitHub, Google, Slack, Salesforce, Jira, Asana, and Zendesk) and AWS resources.
- Integrating seamlessly with existing identity systems like Amazon Cognito, Okta, or Azure Entra ID to streamline authentication and lower custom development efforts.
- Propagating user identity context to agent code.

AgentCore Observability provides a unified view into the operational health of every agent across the enterprise. Built on Amazon CloudWatch, it delivers real-time visibility into performance, reliability, and resource utilization. This ensures teams can track how agents behave under load and identify issues before they affect production. Key capabilities include:

- Comprehensive operational telemetry that captures traces, session counts, latency, duration, token usage, and error rates across all AgentCore services.
- Rich metadata tagging and filtering that simplify issue investigation by linking anomalies to specific agents, workflows, or tools.
- OpenTelemetry (OTEL) compatibility, enabling seamless integration with existing monitoring ecosystems such as Dynatrace, Datadog, Arize Phoenix, LangSmith, Langfuse, and Grafana.
- Custom business and operational attributes that connect agent performance metrics directly to business outcomes, improving traceability and governance.

By unifying these signals into a single pane of glass, AgentCore Observability establishes the feedback loop necessary for responsible scaling—helping enterprises debug faster, enforce policy with confidence, and align operational insights with measurable business impact.

AgentCore Gateway serves as the integration layer for simplified tool development and secure access. It enables organizations to transform APIs and AWS Lambda functions into agent-compatible tools, supporting multiple input types including OpenAPI, Smithy, and AWS Lambda. The Gateway also connects to existing Model Context Protocol (MCP) servers and supports the Agent-to-Agent (A2A) protocol, allowing agents to interoperate and share context securely across environments.

The Gateway provides:

- A unified interface to discover and use thousands of tools through a single, secure endpoint.
- Intelligent tool discovery using semantic search, which automatically indexes tools and returns only the most relevant ones. This dramatically reduces the context passed to the LLM, improving accuracy, speed, and cost.
- IAM authorization for control in the Gateway.

AgentCore Browser is a secure, cloud-based browsing runtime that allows AI agents to interact with web applications and external sites at enterprise scale. It enables controlled access to real-time data and systems without exposing sensitive credentials or violating compliance boundaries. Key capabilities include:

- Enterprise-grade session control that preserves isolation and integrity during authenticated browsing and multi-step transactions.
- Reduced CAPTCHA interruptions and adaptive handling of dynamic web content to maintain high throughput across automated tasks.
- Centralized audit and policy enforcement that logs all agent browsing activity for compliance and traceability.
- Tight integration with AgentCore Observability, allowing teams to monitor performance, usage, and outcomes of web interactions alongside other agent metrics.

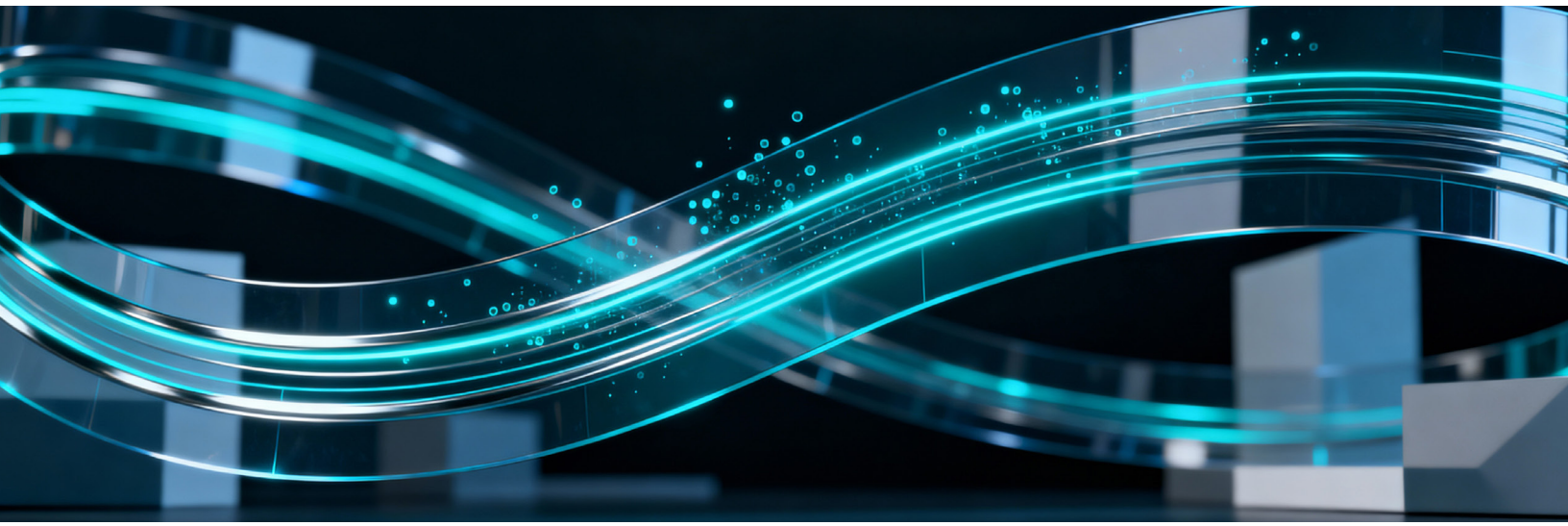
Together with the Code Interpreter, the Browser extends AgentCore's operational reach, enabling agents to reason, execute, and interact securely within dynamic digital environments.

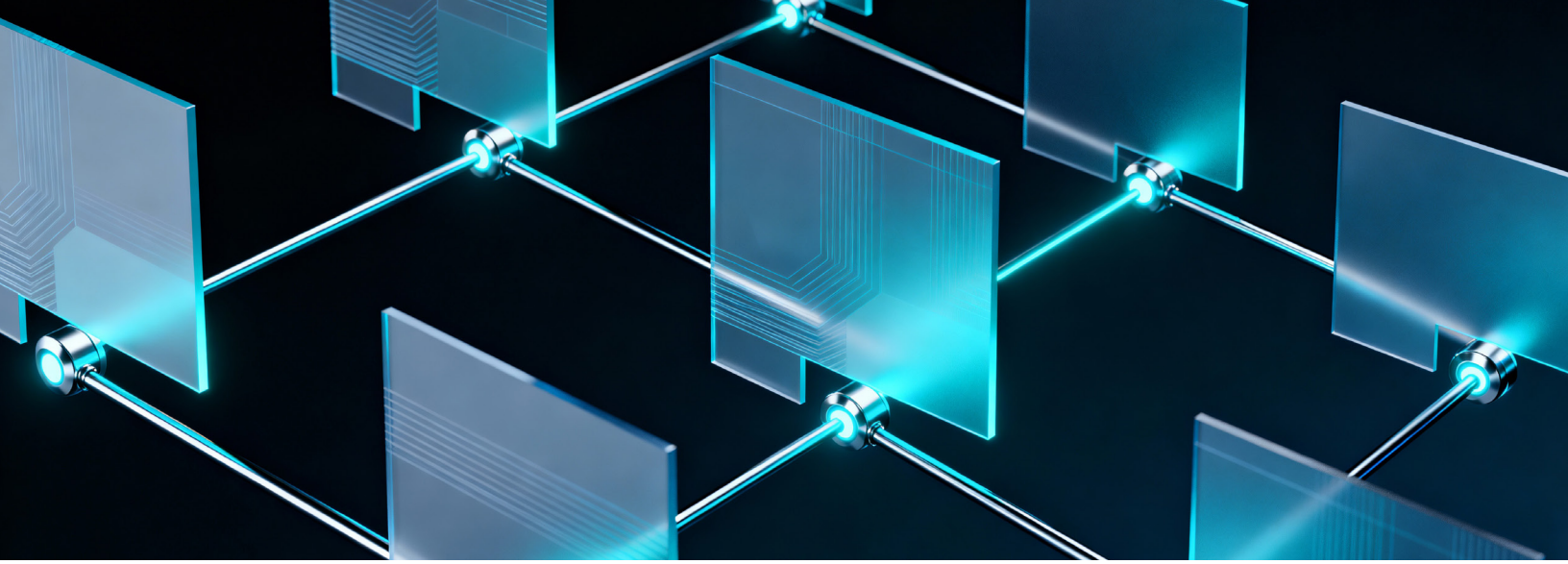
AgentCore Memory simplifies how developers build context-aware agents that can learn, recall, and adapt over time. It abstracts the complexity of managing memory infrastructure, providing automatic scaling and data isolation for enterprise use. Key capabilities include:

- Support for both short-term and long-term memory, enabling multi-turn conversations, persistent context, and shared memory stores across agents.
- Structured episode management that reduces repetitive mistakes, optimizes context usage, and improves decision-making accuracy.
- Custom extraction logic powered by developers' preferred models and prompts, giving precise control over what information is stored or recalled.
- Managed infrastructure with built-in vector embedding, reflection, and observability, freeing teams from operational overhead while ensuring data privacy and audit readiness.

By embedding secure, composable memory within the platform, AgentCore enables the development of agents that continuously improve through experience. It bridges the gap between isolated reasoning and sustained operational intelligence.

AgentCore operates within Amazon Bedrock, which serves as the AI platform for both agentic and non-agentic projects. Within this framework, AgentCore provides the agentic platform that enables the creation, orchestration, and management of intelligent agents with integrated memory, governance, and scalability.





From Pilot to Platform: AgentCore in the Stack

AgentCore is architected to be an integrating layer, not a replacement for existing core systems. It aims to fit neatly within the existing AI stack, working with:

- LLMs: It integrates with models inside Amazon Bedrock, such as Amazon Nova, and third-party models outside of Bedrock like Gemini and OpenAI, allowing for flexibility and choice. Customers value the flexibility to use models hosted on Amazon Bedrock or from external providers.
- RAG and Data Infrastructure: Agents can leverage knowledge bases for RAG tasks.
- Infrastructure: AgentCore rests on powerful AI compute resources, including AWS Trainium and AWS Inferentia.

A key reference architecture, such as a Customer Support Agent, illustrates this integration. The customer's question is routed to an agent within the AgentCore Runtime. The agent can retrieve previous interactions from AgentCore Memory and validate user access through AgentCore Identity. The agent uses AgentCore Gateway to access tools backed by AWS Lambda functions. All of this activity is tracked and sent to AgentCore Observability. This is true operationalization.

This platform approach differentiates it from raw orchestration frameworks. While Strands Agents, LangChain, or CrewAI are powerful in the build phase, AgentCore is the operationalization platform. Frameworks focus on agent logic; AgentCore provides the required enterprise-grade, fully managed services for deploying, securing, governing, and scaling that logic.

AgentCore Use Cases

Industry adoption of agentic AI platforms is accelerating but remains highly stratified by sector and solution maturity. Early experimentation with open-source orchestration frameworks has demonstrated the feasibility of AI-driven automation, yet most organizations encounter steep barriers scaling prototypes across complex, regulated environments. Organizations are now demanding agent solutions that provide not just autonomy, but also granular control, observability, and compliance. As more industries converge on the need for end-to-end production agent infrastructure, best practices now emphasize the importance of phased rollouts, cross-functional governance, and upskilling technology teams to close the emerging AI skills gap.

The power of AgentCore is best demonstrated through high-value, multi-step applications where the challenges of scale and security are paramount. The platform is purpose-built to move innovative AI solutions from controlled experiments to production systems that drive tangible business outcomes across industries. By operationalizing agentic AI, AgentCore enables systems that can reason, plan, and act autonomously across complex workflows.

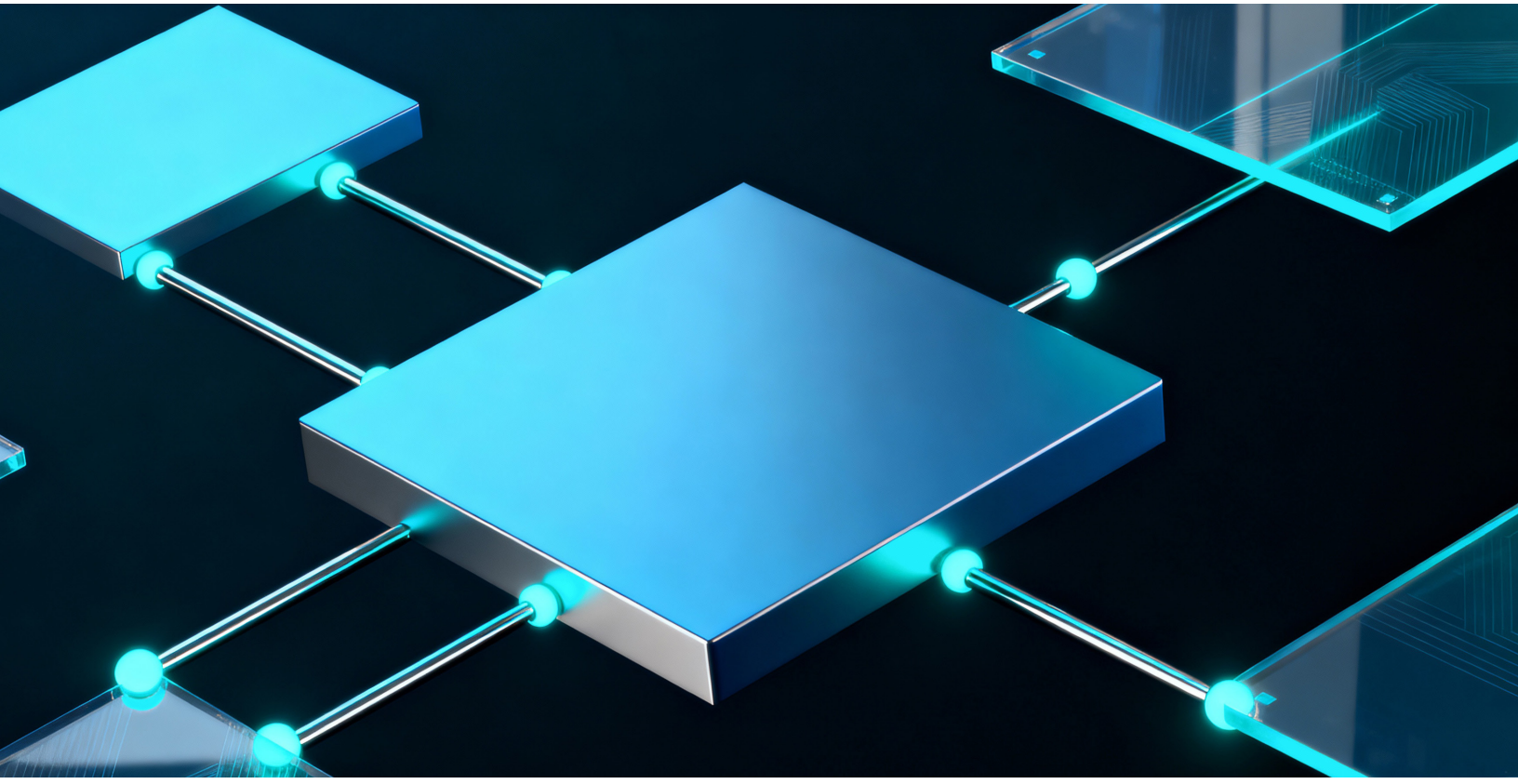
Organizations can build a robust, multi-agent framework designed to transform customer experience. This solution moves beyond basic chatbots by using multiple AI agents that work collaboratively to automatically choose the right tools and handle complex inquiries in everyday language. The system utilizes agents to understand the customer's intent, handle the necessary technical lookups and transactions behind the scenes, and retrieve past session history from AgentCore Memory. Access is securely managed through AgentCore

Identity. Business outcomes include substantial acceleration in issue resolution and a significant reduction in the need for human intervention. The agent is engineered to maintain high data integrity and minimize undesirable outputs, a critical measure of trust for any customer-facing application.

Organizations of all sizes often grapple with vast documentation, millions of lines of proprietary code, and system knowledge scattered across decades of innovation. AgentCore is used to deploy agents that connect this fragmented knowledge. By using AgentCore's framework flexibility, which supports integrations with common orchestration tools like LangGraph, companies build specialized agents. These agents analyze documents, search extensive repositories, and provide deep, unified system insights. AgentCore Gateway serves as a unified, semantic search-enabled interface to connect the agents to all relevant tools and data sources. Organizations can accelerate research and development through the rapid deployment of specialized AI agents. The solution unifies data across previously isolated information systems, creating a flexible and scalable infrastructure for thousands of engineers.

In regulated industries such as healthcare, the manual processing of complex data, like medical assessments and clinical records, creates significant delays and administrative overhead. With AgentCore Runtime, organizations can deploy secure, scalable microVM environments that support multi-agent collaboration for high-stakes workloads. In the Cohere Health case study, for example, the company used AgentCore's agentic infrastructure to accelerate clinical data analysis and automate medical necessity determinations by enabling agents to communicate, reason, and validate against large volumes of patient data. This approach reduced clinical review times, improved the accuracy of prior authorization decisions, and helped providers meet turnaround time requirements while delivering better patient outcomes.

Advertising and marketing teams struggle to create and manage thousands of personalized campaigns daily while maintaining quality and scale. Companies leverage AgentCore to build an AI-powered campaign orchestration system. They utilize AgentCore Runtime to effortlessly deploy and scale specialized marketing agents and rely on AgentCore Observability to provide clear visibility into agent performance across thousands of daily campaigns. This unified platform enables automated campaign design, audience targeting, and real-time optimization across multiple channels. This results in measurable reductions in campaign setup time and increased capacity for personalization and campaign creation. The automation saves teams a significant number of hours per week on manual campaign management tasks.



Looking Ahead

The shift toward agentic AI is more than a new feature; it is a new abstraction layer for applications. It allows developers to offload complex reasoning, planning, and multi-step execution to the AI. This is why composability, governance, and scale are non-negotiable requirements for the future of AI.

If organizations attempt to build agent infrastructure themselves, they will continue to suffer from the limitations of POC-level solutions: poor performance, high security risk, and zero auditability. AgentCore is architected to be the foundation for agent operations, moving the focus from the complexity of building the agent runtime and identity management to the business logic and value delivery. It provides the control plane in the form of the Identity, Gateway, Runtime, and Observability services. These services enable organizations to deploy, manage, and audit production-ready AI agents at scale. The future of business process automation hinges on the ability to deploy reliable, governed, and secure AI agents, and a dedicated operational platform is the clear path forward.

Our analysis suggests that the industry is rapidly consolidating around the non-negotiable elements of AI. Customers are explicitly asking for global region expansion, end-to-end observability, and readiness in a general sense. We are seeing a powerful blending of the open-source community with platform stability, characterized by the adoption of open protocols like MCP and A2A. This trend allows developers to leverage diverse frameworks, such as Strands Agents, CrewAI, and LangGraph, without sacrificing the enterprise-grade requirements for security and scalability. The overall trajectory points toward solutions that prioritize the developer experience, offer faster time to value, and maintain strict security, signaling a maturing market ready to move beyond isolated pilots

ABOUT HYPERFRAME RESEARCH:

HyperFRAME Research delivers in-depth research and insights across the global technology landscape, spanning everything from hyperscale public cloud to the mainframe and everything in between. We offer strategic advisory services, custom research reports, tailored consulting engagements, digital events, go to market planning, message testing, and lead generation programs.

Our industry analysts specialize in rigorous qualitative and quantitative assessments of technology solutions, business challenges, market forces, and end user demands across industry sectors. HyperFRAME Research collaborates closely with your Analyst Relations, Product, and Marketing teams to build and amplify your thought leadership, positioning your expertise to enhance brand and product recognition. Through content that engages readers, viewers, and listeners alike, we ensure your voice resonates across channels.

CONTACT HYPERFRAME RESEARCH:

Steven Dickens
CEO & Principal Analyst | HyperFRAME Research

Email Address:

steven.dickens@hyperframeresearch.com

Telephone Number:

+1 845 505 1678

X: @StevenDickens3

LinkedIn: Steven Dickens

BlueSky: Steven Dickens

CONTRIBUTORS

Stephanie Walter
Practice Leader - AI Stack

Steven Dickens
CEO & Principal Analyst

INQUIRIES

Contact us if you would like to discuss this report and HyperFRAME Research will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "HyperFRAME Research." Non-press and non-analysts must receive prior written permission by HyperFRAME Research for any citations.

LICENSING

DISCLOSURES

HyperFRAME Research provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.


HyperFRAME
RESEARCH