

# From PoC to Production: Shipping Enterprise AI Agents with Amazon Bedrock AgentCore

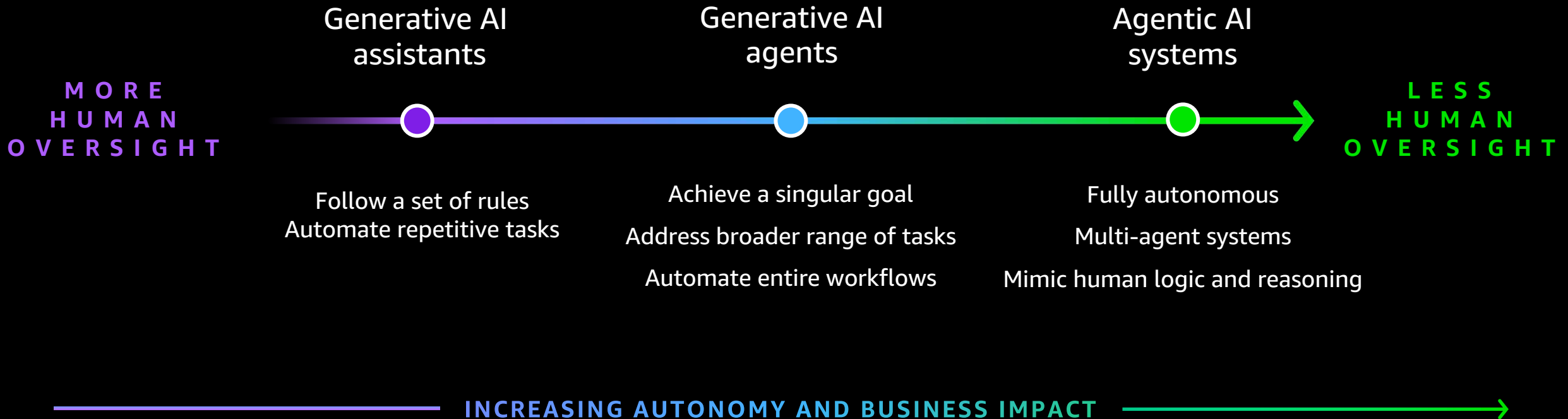
Samuel Baruffi

Principal Solutions Architect

Amazon Web Services (AWS)



# The evolution into Agentic AI



# Enterprises are doubling down on agents

33%

of enterprise software apps will include Agentic AI by 2028, up from less than 1% in 2024.

Gartner, "Top strategic Technology Trends for 2025," October 2024

15%

of day-to-day work decisions will be made autonomously through Agentic AI by 2028.

Gartner, "Top Strategic Technology Trends: Agentic AI—the Evolution of Experience" February 2025

# Agentic use cases across industries

## AUTOMOTIVE

- AI-powered development and validation
- Intelligent contact center
- Supply chain optimization

## ENERGY & UTILITIES

- Energy trading
- Renewable energy planning
- Upstream production optimization

## FINANCIAL SERVICES

- Insurance underwriting optimization
- Intelligent customer experience
- Investment research assistant

## GAMES

- AI for game production
- Code assistance for game development
- In-game AI

## MANUFACTURING

- Research intelligence and discovery
- Maintenance management
- Supply chain optimization

## TELECOM

- Customer experience
- Enterprise productivity
- Network automation

## MEDIA & ENTERTAINMENT

- Multi-agent collaboration for video understanding
- Media planning and campaign optimization
- Multi-agent collaboration for advertising sales data

## RETAIL & CONSUMER GOODS

- Agentic shopping assistants
- Supply chain reallocation
- Pricing agents

## HEALTHCARE & LIFE SCIENCES

- Clinical workflow automation
- Prior authorization automation
- Patient and member engagement



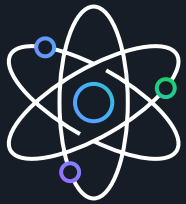
## OUR VISION

To provide the best place to build the world's most useful AI agents, empowering organizations to deploy reliable and secure agents at scale

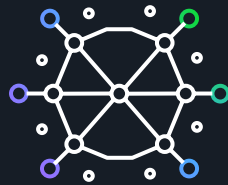


# How AWS gets you there

## Focus on the fundamentals



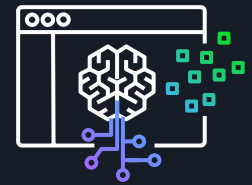
Build  
and leverage  
state-of-the-  
art science



Provide  
best-in-class  
infrastructure  
for building and  
running agents



Deliver the best  
specialized agents



Provide intuitive  
experiences so  
anyone can build &  
use powerful agents

# The broadest choice for building and deploying agents

## Agent Powered Applications



For Builders



For Business Users

## Tools for Building AI Agents



Marketplace



Bedrock Agents



Strands



Nova Act SDK

## Agentic AI Foundation



Amazon Bedrock AgentCore

# AWS AI Stack

## APPLICATIONS

Kiro

Amazon Q

AWS Transform

Amazon Connect

AWS Marketplace

## AI & AGENT DEVELOPMENT SOFTWARE & SERVICES

### SDKS FOR AGENTS

Vertically Integrated

Nova Act

Flexible/OSS

Strands Agents

### AMAZON BEDROCK

#### Models

Amazon Nova

3P Models

#### Capabilities

Optimization

Guardrails

Customization

#### AgentCore

Runtime

Gateway

Memory

1P Tools

Identity

Observability

Knowledge Bases

## INFRASTRUCTURE

### AMAZON SAGEMAKER AI

Model building

Model training

Fine-tuning

Deployment

MLOps

Governance

### AI COMPUTE

AWS Trainium

AWS Inferentia

GPUs

INTERFACES & PROTOCOLS  
(MCP/A2A)

DATA



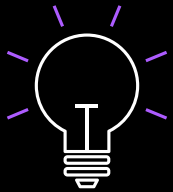


# Amazon Bedrock AgentCore



# The prototype to production “chasm”

Excitement  
and potential



POC

Challenges on the path to production



Performance



Scalability

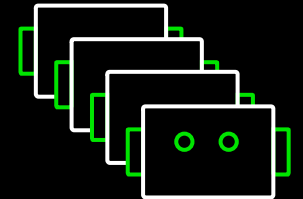


Security



Governance

Meaningful  
business value



AI production  
agents





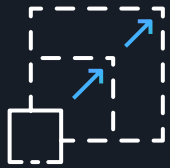
# Amazon Bedrock AgentCore Services





# AgentCore Runtime

## Scale from real-time to multi-hour workloads



- Scale from real-time to multi-hour workloads with low latency and industry-leading extended runtime (up to 8 hours)
- Supports payloads across modalities

## Accelerate time to market



- Deploy any AI agent using common open-source frameworks
- Deploy from POC to production in a few lines of code

## Secure workload with enterprise-grade isolation

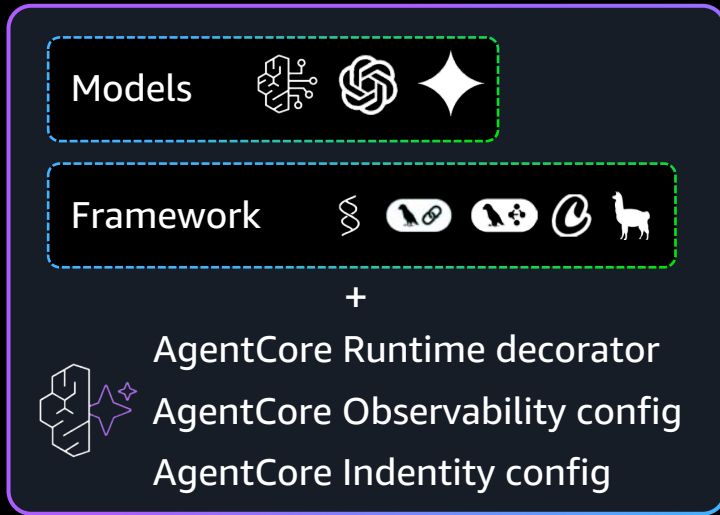


- True session isolation to protect your data
- Integrates with existing identity providers



# AgentCore Runtime

Agent or tool code

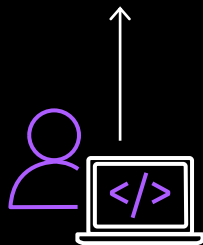
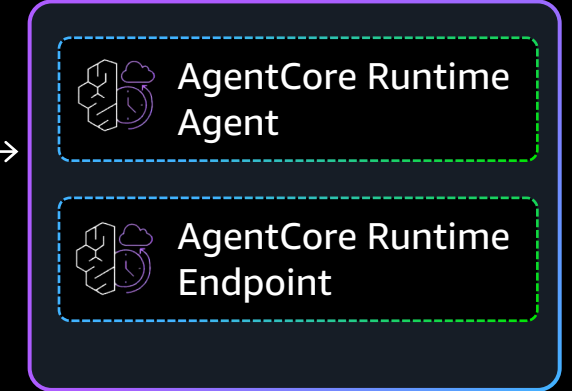


configure

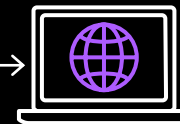


launch

Amazon ECR Repository



User



Application

invoke



# AgentCore Identity

## Secure, delegated access for AI agents



- Enables AI agents to securely access AWS resources and third-party tools such as GitHub, Google, Salesforce and Slack
- Robust access controls with just-enough access and secure permissions delegation

## Build streamlined AI agent experiences



- Minimizes consent fatigue with a secure token vault
- Streamlines authentication flows

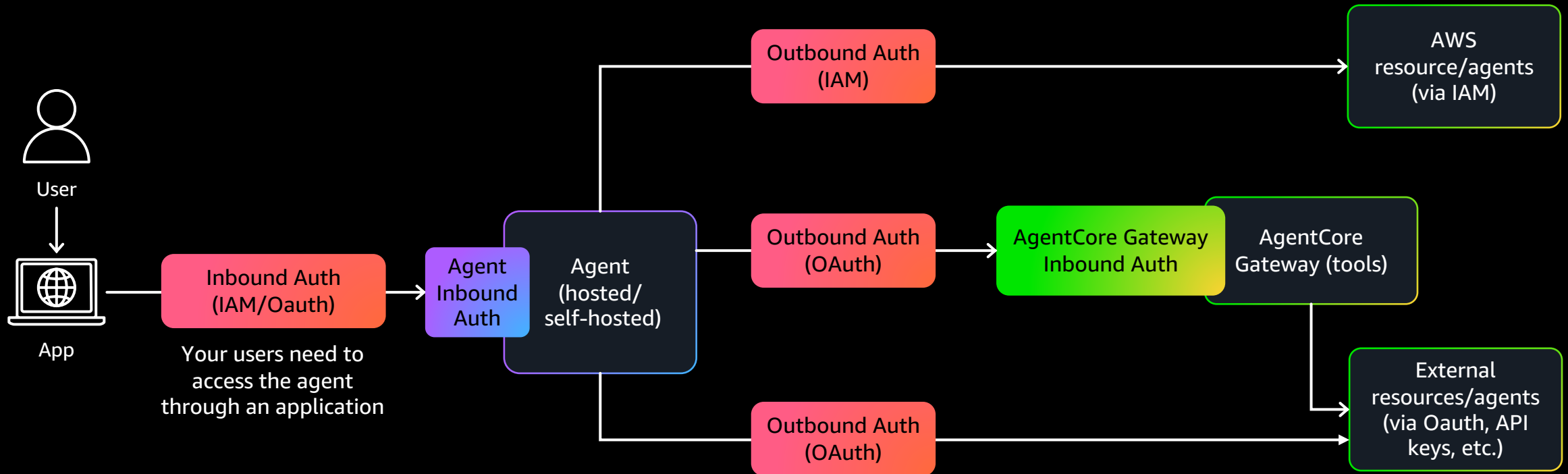
## Accelerated AI agent development



- Preserves existing identity systems such as Okta, Azure AD, or Amazon Cognito
- Lowers custom development efforts without need for migrating users or rebuilding authentication flows



# AgentCore Identity





# AgentCore Gateway

## Simplified tool development & integration



- Turn APIs, Lambda functions, and existing services into MCP-compatible tools
- Access thousands of tools through a standardized interface

## Secure and unified access



- Discover and use tools through a single, secure endpoint
- Combine multiple tools sources into one unified interface

## Intelligent tool discovery

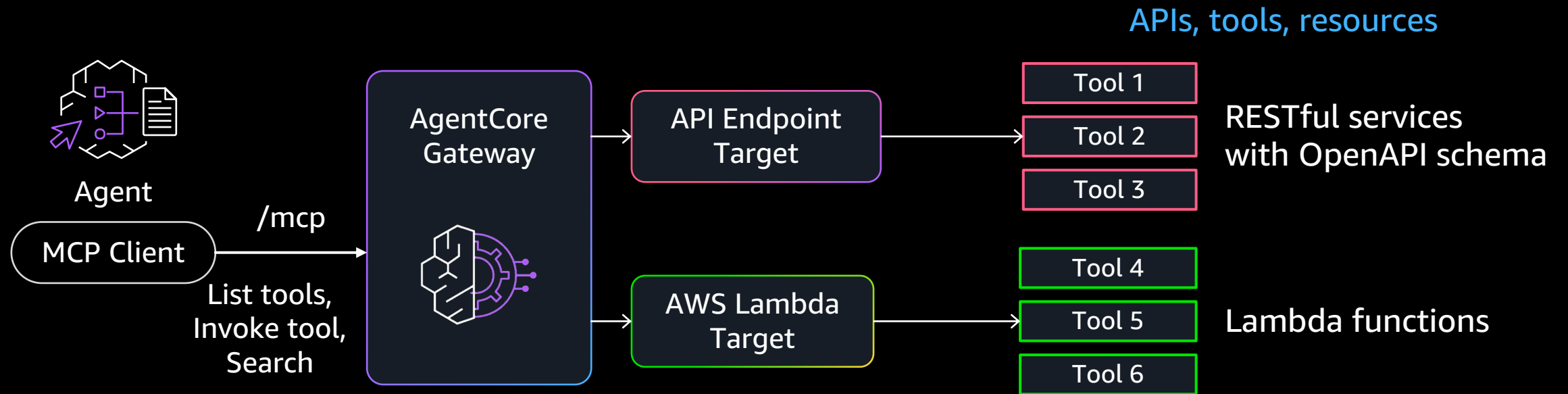


- Enable agents to find the right tools with context aware discovery
- Curated tool collections with granular permissions



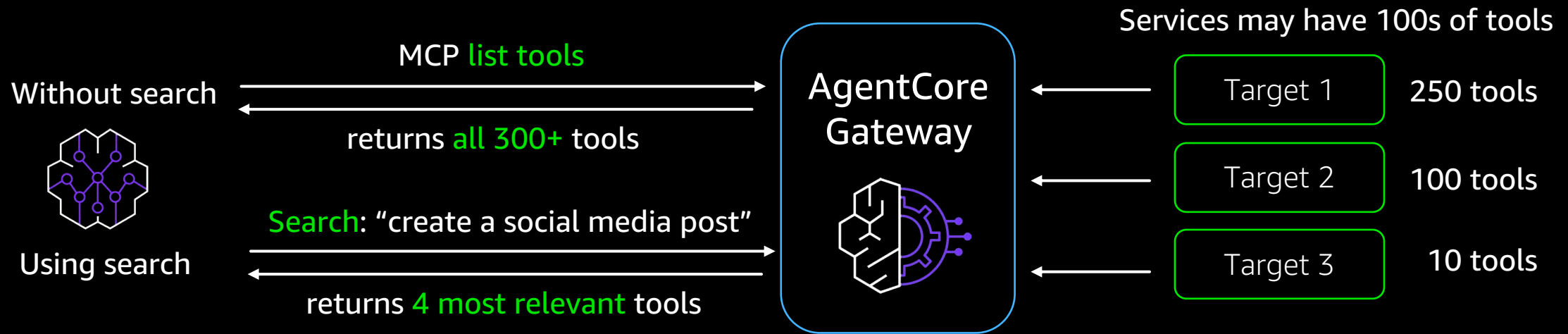


# AgentCore Gateway





# AgentCore Gateway semantic search



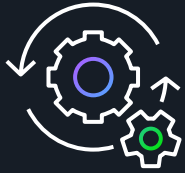
## Benefits

- AgentCore Gateway automatically indexes tools, and gives serverless semantic search
- Reduces context passed to the agent's LLM, improving accuracy, speed, and cost
- Lets agent focus on tools relevant for a given task



# AgentCore Memory

## Simplify memory management



- Abstracts memory infrastructure
- Scales automatically with serverless architecture
- Automatically stores and manages agent context across sessions

## Enterprise-grade



- Complete data privacy with dedicated storage for each customer
- Enterprise security with encryption and regional data storage options

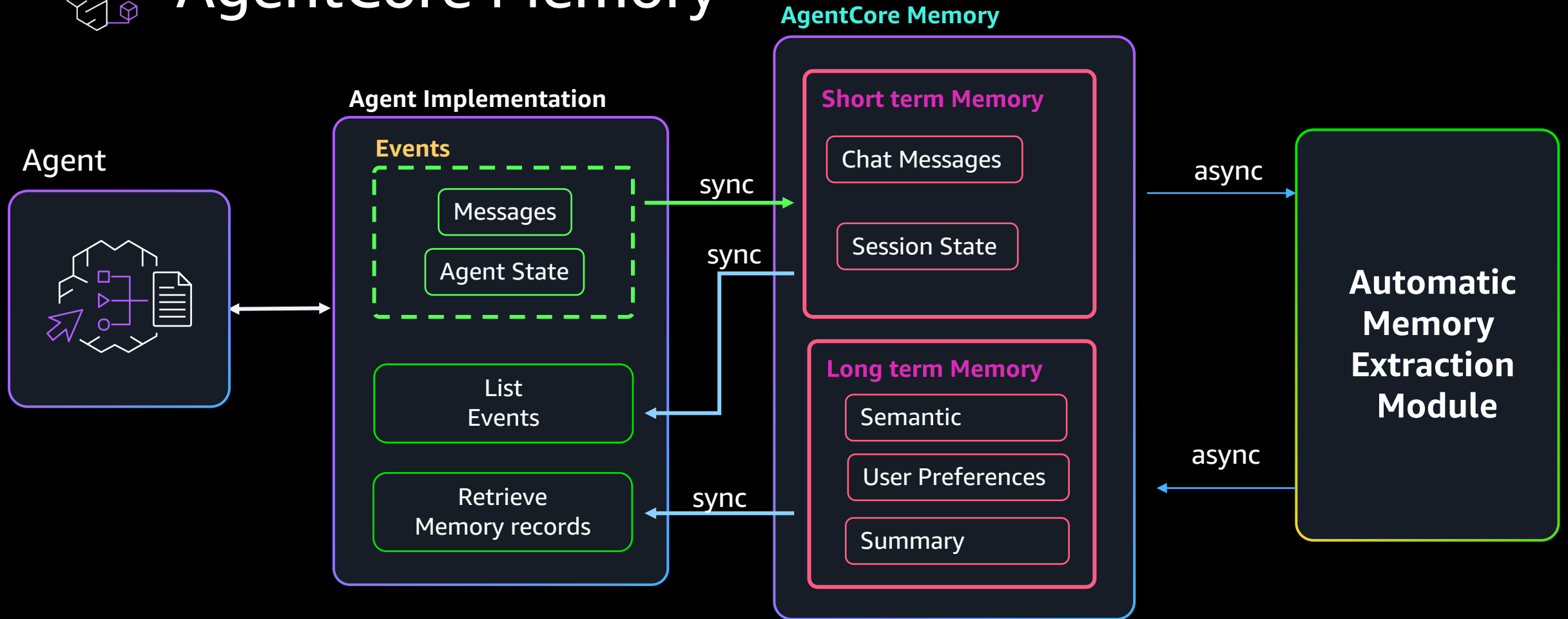
## Deep customization



- Define memory patterns based on your use case
- Configure extraction rules
- Choose models and customize prompts for memory extraction



# AgentCore Memory





# AgentCore Browser

## Serverless browser infrastructure



- Low latency browser sessions
- Auto-scales from 0 to hundreds of concurrent sessions

## Enterprise-grade security



- Session isolated compute with VM-level isolation per user
- Secure credential handling

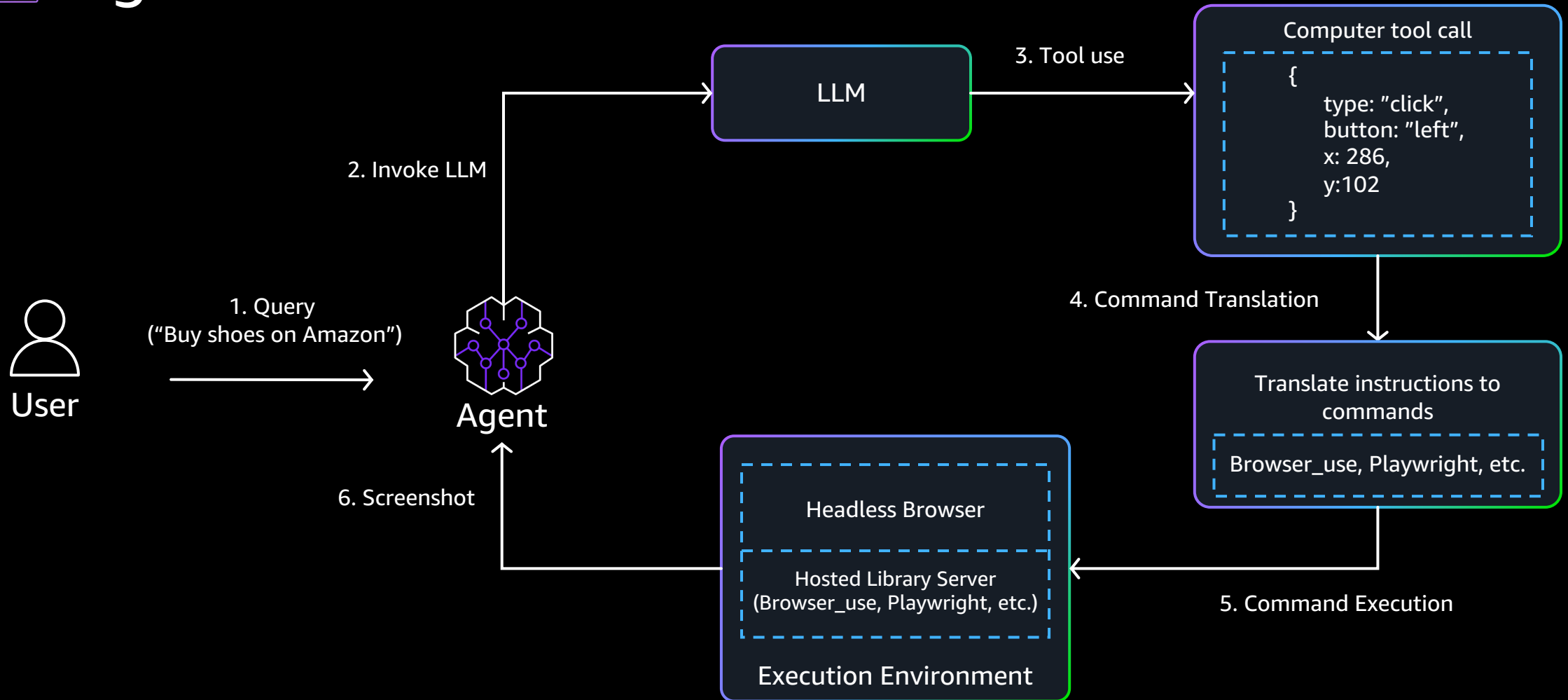
## Enterprise observability



- Live streaming URLs for real-time monitoring
- Session replays for debugging
- Extensive logging of all browser commands to CloudTrail



# AgentCore Browser





# AgentCore Code Interpreter

## Execute code securely



- Execute complex workflows and data analysis in isolated sandbox environments
- Access internal data sources securely without exposing sensitive data

## Large-scale data processing



- Process gigabyte-scale datasets efficiently using Amazon S3 integration, without API limitations

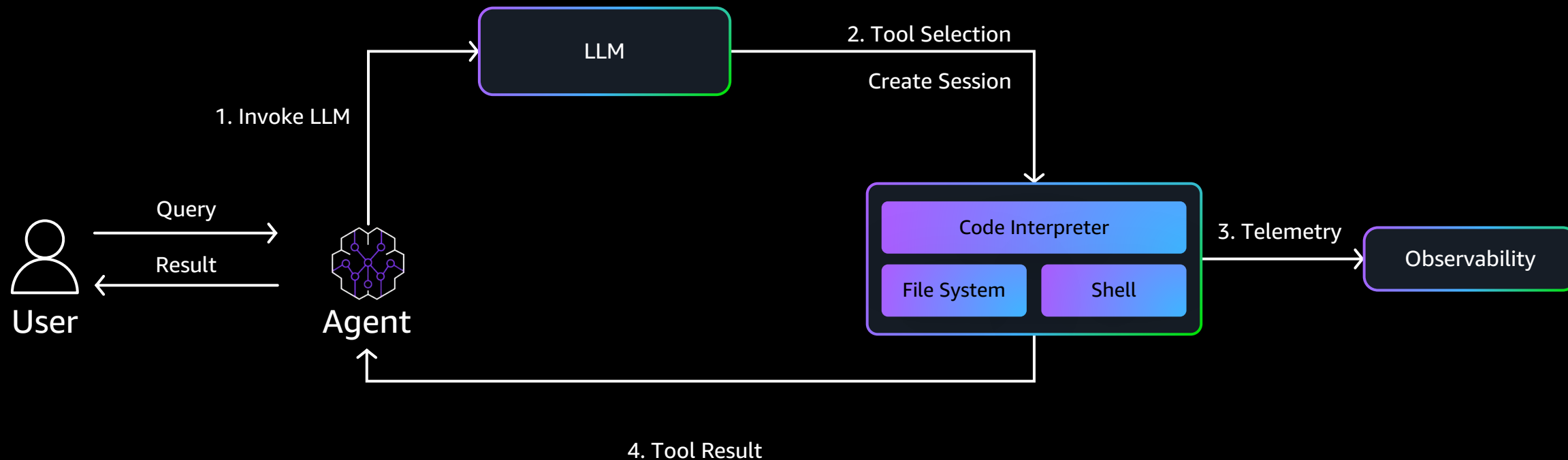
## Ease of use



- Quick start with pre-built execution runtimes for JavaScript, TypeScript, and Python with common libraries pre-installed



# AgentCore Code Interpreter

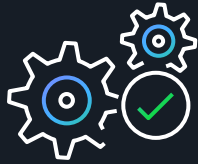






# AgentCore Observability

## Maintain quality and trust



- Comprehensive end-to-end visibility into agent behavior
- Accelerated debugging and quality audits
- Quickly detect issues and assess performance trends

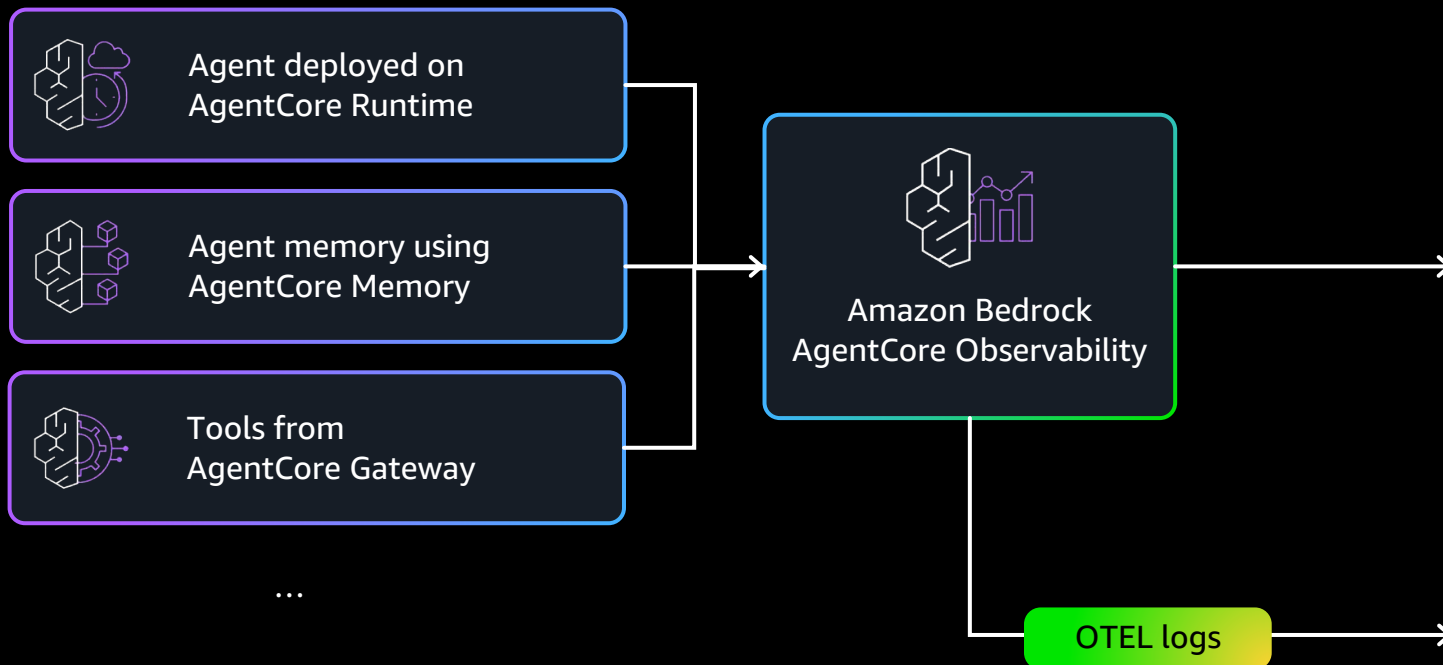
## Integrate with 3P observability tools



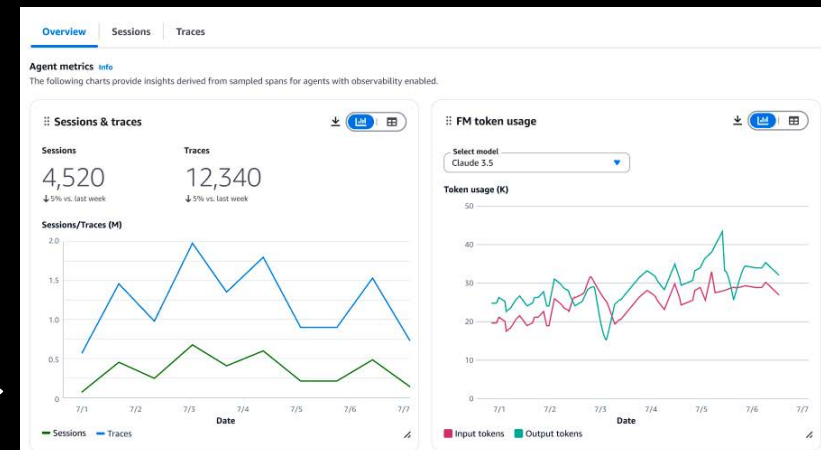
- Integration with a wide range of monitoring and observability tools, including CloudWatch
- Flexibility to leverage your existing observability stack



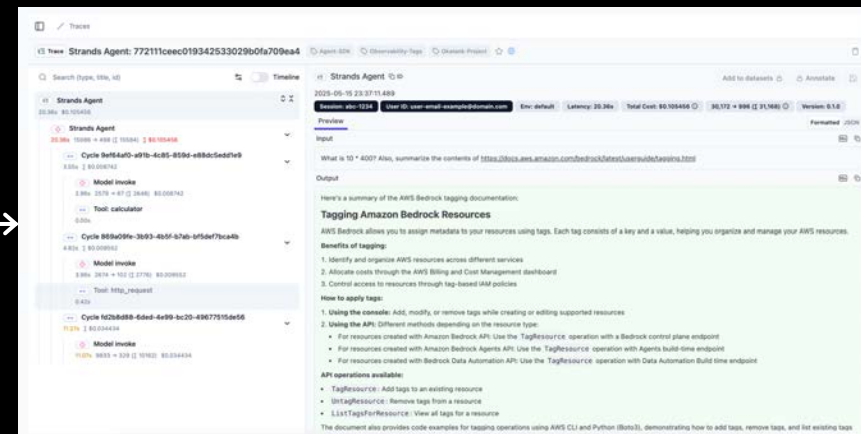
# AgentCore Observability



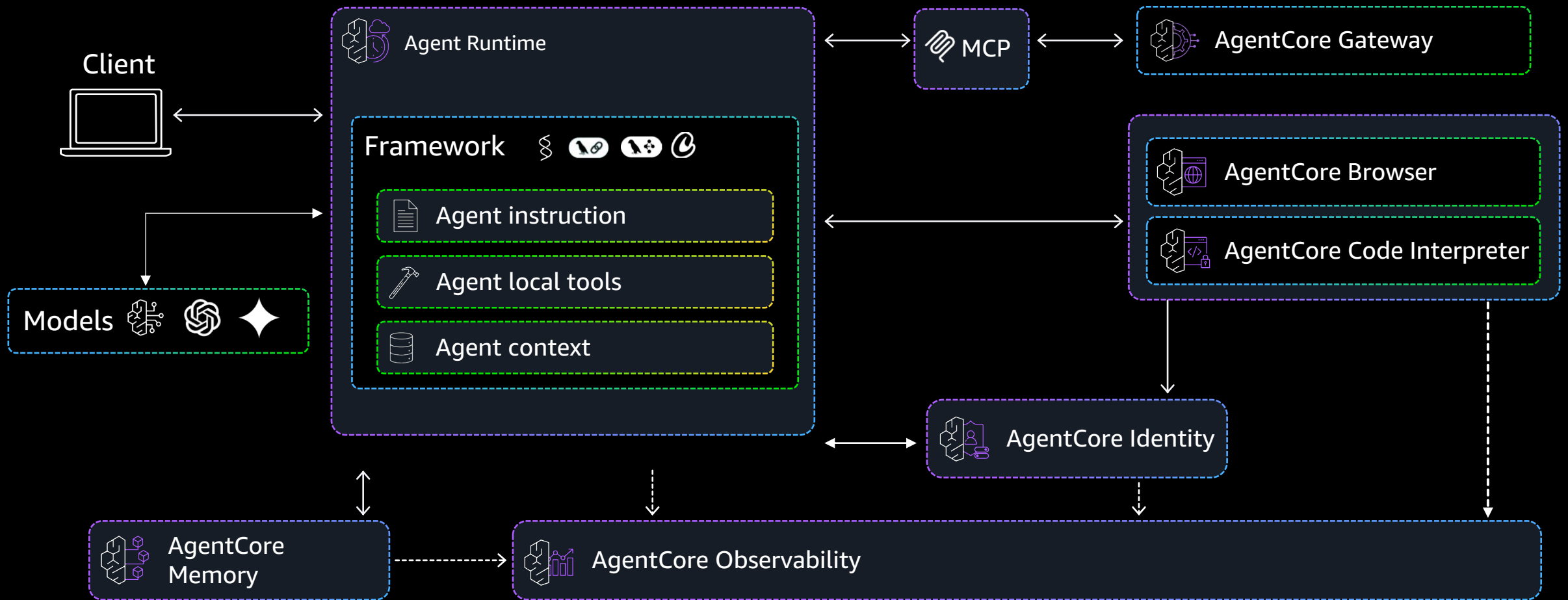
## AgentCore Observability dashboards



## Third-party observability dashboards



# AgentCore capabilities enabling agents at scale



# Amazon Bedrock



Open source

## Protocols:

MCP, A2A (coming soon)

## Frameworks:

CrewAI, LangGraph,  
Strands Agents



Amazon Bedrock



AgentCore



Runtime



Identity



Gateway



Code interpreter



Memory



Browser tool



Observability

Model Capabilities



Model access



Cost and performance  
optimization



Customization



Guardrails

Knowledge Bases

# Start building with **Amazon Bedrock AgentCore**



AgentCore  
Overview



AgentCore  
Samples



AgentCore  
Documentation



AgentCore  
blog post

# Thank you!

Samuel Baruffi

@samuelbaruffi



# Industry-specific use cases



# Top **advertising** and **marketing** agentic AI use cases



PROBLEM	Creative agencies spend months developing campaign concepts, which involve understanding challenges and audiences, conducting research, analyzing previous work, and brainstorming	Media planning tasks are complex and time-consuming due to the evolving media landscape, need for in-depth audience analysis, and lack of automated solutions for campaign insights	Effective customer segmentation requires analyzing vast data to create personalized experiences, with key challenges including data reliability and real-time insights processing
USE CASE	Advertising campaign creative concept development	Agent-enabled media planning and automation	Audience segmentation and hyper personalization
OUTCOME	<p>Retrieve insights from brand books, products descriptions, previous campaigns, audience research, brand sentiment</p> <p>Ideate concepts and develop creative copy and imagery autonomously</p> <p>Accelerate campaign optimization and fine-tuning through coordinated AI agents</p>	<p>Streamline data collection and analysis, develop insights for targeting, and decide budget allocation autonomously</p> <p>Negotiate and execute ad deals by matching advertisers with suitable ad placements</p> <p>Automatically analyze performance, attribute conversions, and optimize media mix</p>	<p>Autonomous customer segment development/fine-tuning based on real-time insights</p> <p>Coordinate personalized copy and imagery development and activation across customer experience channels like web, mobile, text/SMS, email, social, etc. and automate real-time insight retrieval</p>





# Top **aerospace** agentic AI use cases

PROBLEM	Complex geospatial data analysis requires specialized expertise, limiting accessibility	Space manufacturing processes lack integrated visibility across operations.	Satellite fleet management demands continuous monitoring and complex decision-making
USE CASE	Natural language-powered geospatial intelligence	AI-enabled space manufacturing optimization	Satellite fleet operations require constant monitoring
OUTCOME	<ul style="list-style-type: none"><li>Democratize access to sophisticated geospatial insights</li><li>Reduce analysis time through automated data processing</li><li>Enable cross-functional collaboration through simplified data access</li></ul>	<ul style="list-style-type: none"><li>Enable real-time visibility and proactive optimization of manufacturing operations</li><li>Minimize resource waste through AI-driven optimization</li><li>Increase first-time quality rates</li></ul>	<ul style="list-style-type: none"><li>Maximize orbital resource efficiency and service reliability</li><li>Enhance space situational awareness and collision avoidance</li><li>Accelerate anomaly detection and resolution</li></ul>



# Top **automotive** agentic AI use cases

PROBLEM	Automakers face challenges with growing software complexity, faster innovation cycles, and stringent quality requirements	Manual workflows, siloed data, and limited personalization make it difficult to deliver the fast, consistent, and meaningful customer support	Complex global supply chains create persistent inventory imbalances and forecast errors
USE CASE	AI-powered development and validation	Intelligent contact center	Supply chain optimization
OUTCOME	<p>Develop more quickly with a virtual engineering workbench (VEW) integrated with a gen AI-powered coding assistant</p> <p>AI agents help automate routine tasks, identify defects earlier, and support continuous validation</p>	<p>Automatically route inquiries, understand intent through natural language, and tailor responses based on real-time customer data</p> <p>Virtual agents resolve routine issues and schedule service while escalating complex cases to real agents</p>	<p>Enable autonomous inventory adjustments based on real-time conditions</p> <p>Coordinate decisions across demand, supply, and logistics agents</p> <p>Self-heal supply chain disruptions through coordinated responses across systems</p>



# Top **education** agentic AI use cases

PROBLEM	Manual processing of diverse educational documents across multiple formats and languages creates inefficient, costly, and error-prone communication between institutions and stakeholders	Staff shortages at higher education institutions are hampering their ability to effectively deliver student support services and manage varying inquiry volumes	Teachers and advisors spend significant time manually gathering and analyzing student information from various sources, limiting their time to spend more time on teaching and advising
USE CASE	Accelerated transcript processing	Streamlined education help desk	Supply chain optimization
OUTCOME	<ul style="list-style-type: none"><li>Streamlined evaluation process for faster enrollment decision-making</li><li>Enabled new approaches for predictive analytics to support enrollment and fiscal management</li><li>Reduced manual efforts and time spent on reviewing international transcripts and course articulation</li></ul>	<ul style="list-style-type: none"><li>Deliver personalized 24/7 student support across multiple departments</li><li>Automatically route inquiries and understand intent through natural language</li><li>Virtual agents resolve routine issues while escalating complex inquiries to faculty and staff</li></ul>	<ul style="list-style-type: none"><li>Generate student insight through natural language queries</li><li>Generate student information and analysis in minutes</li><li>Proactively identify at risk students with recommendations based on individual student goals</li></ul>



# Top **energy & utilities** agentic AI use cases

PROBLEM	Upstream operators spend hours analyzing massive volumes of technical data spanning well logs, drilling reports, and 3D seismic information	Energy trading floors handle massive streams of unstructured data across market feeds, trade tickets, emails, and chat transcripts that must be reconciled and analyzed in real-time	Development teams struggle to balance variable generation forecasts, supply chain constraints and interconnection timelines, leading to costly delays and overbuilding renewable energy project
USE CASE	Upstream production optimization	Energy trading	Renewable energy planning
OUTCOME	<ul style="list-style-type: none"><li>Unifies data across wells, reservoirs and surface data to recommend optimizations.</li><li>Proactively identifies equipment maintenance</li><li>Reduces analysis time for engineers</li></ul>	<ul style="list-style-type: none"><li>Ingest data across energy markets, weather and operations to generate optimal trading strategies</li><li>Automates trade reconciliation for improved compliance</li><li>Reduces trade settlement time from days to minutes</li></ul>	<ul style="list-style-type: none"><li>Simulates weather, demand and grid constraints for optimal site selection</li><li>Analyzes wind and solar configurations to maximize energy production</li><li>Accelerates permitting process</li></ul>



# Top **financial services** agentic AI use cases

PROBLEM	Investment analysts spend significant time manually gathering and analyzing data from diverse sources, limiting their ability to generate timely insights	Static rule-based underwriting leads to slow turnaround, inconsistent decisions, and missed risks	Financial institutions struggle to scale personalized customer service while maintaining compliance and operational efficiency
USE CASE	Investment research assistant	Insurance underwriting optimization	Intelligent customer experience
OUTCOME	<ul style="list-style-type: none"><li>Generate market insights through natural language queries</li><li>Create detailed financial visualizations and analysis in minutes</li><li>Proactively identify market opportunities through multi-agent analysis</li></ul>	<ul style="list-style-type: none"><li>Transform underwriting speed through instant data-driven decisions</li><li>Ensure fair and consistent underwriting decisions with explainable AI</li><li>Maximize audit efficiency through automated risk assessments while reducing costs</li></ul>	<ul style="list-style-type: none"><li>Deliver personalized 24/7 customer support across multiple channels</li><li>Increase first-contact resolutions through AI-powered analysis and suggestions</li><li>Enable secure AI experimentation for continuous service improvement</li></ul>



# Top **games** agentic AI use cases

PROBLEM	Game codebases are some of the most complex, branching and multi-faceted software projects	Creating games that adapt to individual player preferences leads to higher overall player satisfaction	As players demand more content, game development is becoming increasingly complex and resource-intensive
USE CASE	Code assistance for game development	In-game AI	AI for game production
OUTCOME	<p>Agents perform code analysis and optimization, debugging, documentation, and testing</p> <p>Agents in the code pipeline autonomously cleaning up and debugging can increase time dev teams spend on polishing the player experience</p>	<p>AI assisted agents can help players learn and master the game</p> <p>Agents collaborate to create personalized experiences and in-game commerce</p> <p>Agents to facilitate lifelike NPC interactions</p>	<p>Agents help prototype game assets including images, narrative, dialog, 3D models, music, and code.</p> <p>Agents integrate prototyped assets and convert them into playable game assets</p> <p>Agents can enable non-developer players to create and integrate content in a game</p>



# Top **government** agentic AI use cases

PROBLEM	Complex engineering design cycles create lengthy development timelines in regulated environments	Remote autonomous systems require extensive human oversight and control	Regulated operations face inefficiencies in compliance documentation and analysis
USE CASE	AI-powered engineering design for mission-critical components	Edge-optimized autonomous systems with real-time decision making	ITAR-compliant AI platform for regulated operations
OUTCOME	<ul style="list-style-type: none"><li>Reduce development cycles while maintaining security and compliance</li><li>Increase first-time quality rates through AI-assisted design validation</li><li>Enable secure knowledge sharing across regulated organizations</li></ul>	<ul style="list-style-type: none"><li>Enable efficient autonomous operations with reduced ground control requirements</li><li>Decrease time-to-deployment for critical systems</li><li>Reduce operational risk through predictive analytics</li></ul>	<ul style="list-style-type: none"><li>Achieve faster compliance analysis and reductions in procedure documentation time</li><li>Improve audit readiness through automated compliance tracking</li><li>Enable rapid response to regulatory changes through automated updates</li></ul>



# Top **healthcare** agentic AI use cases

PROBLEM	Clinical systems often contain unstructured data in various formats, slowing down decision making and adding administrative burden to clinicians	Prior authorization is a labor-intensive, time-consuming process that delays patient care, increases administrative costs, and frustrates providers, payors, and patients	Healthcare providers and payors struggle to deliver accessible and consistent personalized care plans due to resource constraints and growing patient demands
USE CASE	Clinical workflow automation	Prior authorization automation	Patient and member engagement
OUTCOME	<p>Extract relevant information from EHRs, clinical notes, images, and medical literature</p> <p>Summarize patient histories, identify trends, and flag potential issues for providers</p> <p>Autonomously manage data integration tasks like mapping between different healthcare standards and orchestrating multi-step clinical workflows</p>	<p>Automate documentation processing to reduce errors and speed processing</p> <p>Real-time verification of submission completeness and parallel processing of multiple requirements</p> <p>Suggest alternative treatments that may not require PA</p>	<p>Answer health and insurance-related questions; provide guidance on next steps</p> <p>Personalized patient/member communication for care plans, medication reminders, and follow up appointments</p> <p>Automate intake and conduct initial symptom assessments for triage</p>





# Top **life sciences** agentic AI use cases

PROBLEM	Biomarker discovery generates massive molecular and clinical datasets, requiring enormous computational power to analyze complex biological interactions	Generating clinical trial protocols requires considering complex regulatory rules, statistical best practices, recruitment feasibility, and cost considerations	Commercial teams face intense competition, limited healthcare provider availability, tight regulations, decentralized care, and rising patient expectations
USE CASE	Biomarker discovery	Clinical trial protocol development	Commercialization
OUTCOME	<p>Prepare up-to-date, comprehensive research reviews of therapeutic areas with accurate citations</p> <p>Orchestrate complex lab-in-the-loop and directed evolution workflows</p> <p>Democratize access to complex bioinformatics tools</p>	<p>Rapidly analyze historical trial data and real-world evidence to optimize protocol design and improve trial success probability</p> <p>Simulate multiple protocol scenarios across diverse patient population to identify inefficiencies</p>	<p>Use AI-generated content tagging and monitoring to accelerate medical, legal, and regulatory (MLR) review of sales content from weeks to days</p> <p>Personalize patient interactions and recommend relevant content across multiple sales channels</p> <p>Streamline pre-call planning by consolidating multiple tools and HCP data sources</p>





# Top **manufacturing** agentic AI use cases

PROBLEM	R&D teams spend 30%+ of time on non-research activities, limiting innovation speed and missing valuable patterns across siloed knowledge	Maintenance workflows require manual coordination across multiple systems and teams, causing delays and inefficiencies	Complex global supply chains create persistent inventory imbalances and forecast errors
USE CASE	Research intelligence and discovery	Maintenance management	Supply chain optimization
OUTCOME	<ul style="list-style-type: none"><li>Generate and test novel hypotheses across technical domains</li><li>Design and simulate experiments autonomously</li><li>Accelerate discovery-to-implementation through coordinated AI agents</li></ul>	<ul style="list-style-type: none"><li>Autonomously coordinate maintenance workflows across systems and functions</li><li>Automate work order planning and resource allocation in real-time</li><li>Transform equipment reliability and workforce support</li></ul>	<ul style="list-style-type: none"><li>Enable autonomous inventory adjustments based on real-time conditions</li><li>Coordinate decisions across demand, supply, and logistics agents</li><li>Self-heal supply chain disruptions through coordinated responses across systems</li></ul>



# Top **media and entertainment** agentic AI use cases



PROBLEM	M&E companies need to manage petabytes of video content and prepare assets for different downstream distribution channels, which must be organized and tagged with appropriate metadata for cross-platform distribution	Advertisers face significant challenges to create optimized media plans that maximize ROI across multiple entertainment platforms, leading to inefficient budget allocation and sub-optimal campaign performance	New ad formats across different regions make it difficult for ad sellers to keep up with innovation, trainings, and standards, which slows down sales readiness and can impact advertisers results
USE CASE	Agent-enabled video understanding	Media planning and campaign optimization	Automated advertising sales support
OUTCOME	<p>Extract insights from videos, including technical specifications, summaries, and both chapter and shot-level details</p> <p>Create media operations agents for specific tasks, such as celebrity detection, synopses, and quality and compliance control</p> <p>Reduce costs for ad break detection, skipping credits, and semantic search</p>	<p>Streamline ad campaign management by transforming natural language inputs into actionable campaign data</p> <p>Reduce campaign planning time to minutes with automated, multi-agent collaboration</p> <p>Ensure data completeness while maintaining strict compliance standards</p>	<p>Gather, analyze, and synthesize ad products, formats, and standards from various sources using a Retrieval Augmented Generation (RAG) application with AI agents</p> <p>Efficiently answer domain-specific seller questions using a chat-based assistant built on a multimodal knowledge base</p>

# Top retail and consumer goods agentic AI use cases



PROBLEM	Companies want to increase conversions, but shopping assistants built with rule-based AI cannot adequately answer complex questions	Traditional ML models can't detect or react to real-time supply chain disruptions, causing costly inventory allocation issues	Companies can struggle to set pricing that drives demand using siloed data and outdated technology
USE CASE	Agentic shopping assistants	Supply chain reallocation	Pricing agents
OUTCOME	<ul style="list-style-type: none"><li>Boost sales by personalizing recommendations through intelligent conversation and customer history analysis</li><li>Streamline purchases using chain-of-thought reasoning to guide product selection and execute transactions</li><li>Reduce returns by providing accurate, multi-step product guidance</li></ul>	<ul style="list-style-type: none"><li>Detect real-time supply chain anomalies</li><li>Optimize inventory allocation through AI-driven cause identification</li><li>Improve supply chain resilience with automated alternative action recommendations</li></ul>	<ul style="list-style-type: none"><li>Forecast market demand to ensure prices will align with market expectations</li><li>Scrape competitor websites to understand price points of similar products</li><li>Determine an optimal price point based on demand, competition, and target margins</li></ul>



# Top **telecom** agentic AI use cases

PROBLEM	Customer experiences lack intelligent automation and hyper-personalization	Employee productivity bottlenecks across departments limit efficiency	Network operators struggle with manual processes, legacy platforms, siloed data
USE CASE	Customer experience	Enterprise efficiency	Network automation
OUTCOME	<b>Transform customer experiences and reduce churn</b> Autonomous intent understanding and resolution Personalized recommendations and next actions AI-driven customer loyalty management	<b>Accelerate cross-department efficiency and innovation</b> Automated marketing campaign strategy and content creation End-to-end software development lifecycle automation Automated finance and HR workflows	<b>Enable autonomous network operations and design</b> Minimize network downtime and service disruptions Reduce operational costs through predictive maintenance Accelerate network performance and capacity optimization