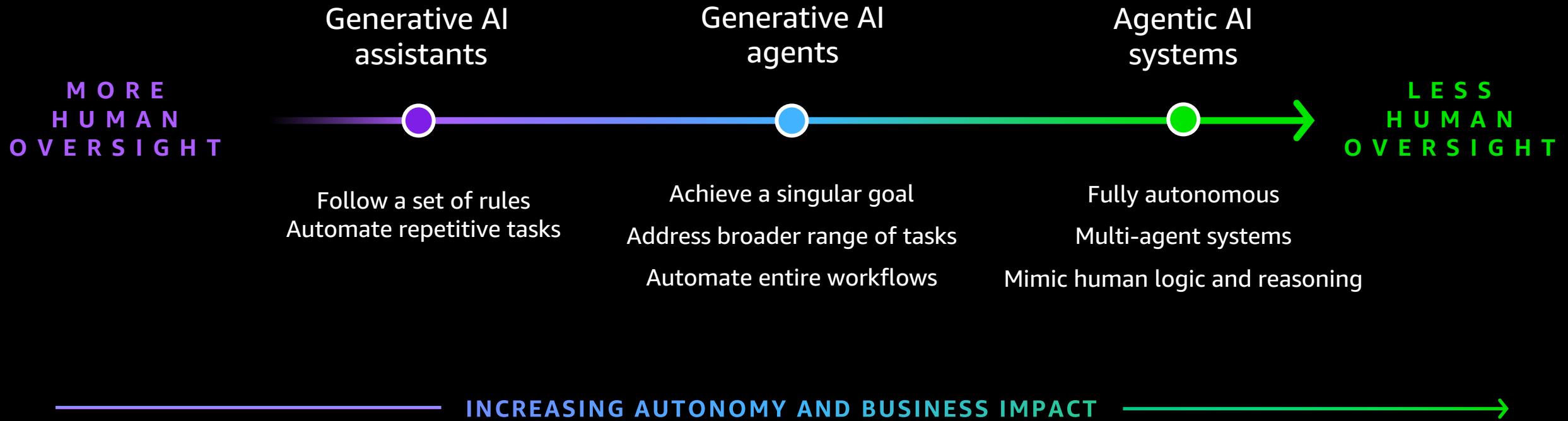


Building AI Agents on AWS



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.

15%

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

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

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



Top agentic AI use cases

PROBLEM	Staff shortages at higher education institutions are hampering their ability to effectively deliver student support services and manage varying inquiry volumes	PROBLEM	Clinical systems often contain unstructured data in various formats, slowing down decision making and adding administrative burden to clinicians	PROBLEM	Biomarker discovery generates massive molecular and clinical datasets, requiring enormous computational power to analyze complex biological interactions
USE CASE	Streamlined education help desk	USE CASE	Clinical workflow automation	USE CASE	Biomarker discovery
OUTCOME	Deliver personalized 24/7 student support across multiple departments Automatically route inquiries and understand intent through natural language Virtual agents resolve routine issues while escalating complex inquiries to faculty and staff	OUTCOME	Extract relevant information from EHRs, clinical notes, images, and medical literature Summarize patient histories, identify trends, and flag potential issues for providers Autonomously manage data integration tasks like mapping between different healthcare standards and orchestrating multi-step clinical workflows	OUTCOME	Prepare up-to-date, comprehensive research reviews of therapeutic areas with accurate citations Orchestrate complex lab-in-the-loop and directed evolution workflows Democratize access to complex bioinformatics tools



O U R V I S I O N

To provide the best place to build
and scale the most useful AI agents



AWS Agentic AI portfolio

INTERFACES & PROTOCOLS (MCP/A2A)

SECURITY & POLICIES

APPLICATIONS

Kiro

Amazon Quick Suite

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

Custom Model Building

Model building

Model training

MLOps

Deployment

Hyperpod

Data Foundation for AI

Data Processing

Governance

Storage and Databases

Vector Search

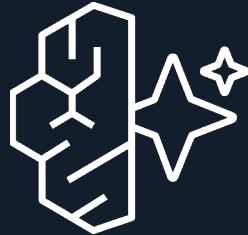
AI COMPUTE

Trainium

Inferentia

GPUs





Amazon Bedrock

The most flexible, comprehensive way to build and scale generative AI applications including agents

Amazon Bedrock

BROAD SELECTION OF FULLY MANAGED MODELS FROM LEADING AI COMPANIES

AI21labs

Highly efficient processing & grounded generation for long context lengths

JAMBA

amazon

Frontier intelligence and industry leading price performance

NOVA

ANTHROPIC

Excels at complex reasoning, code generation, and instruction following

CLAUDE

cohere

Powering efficient, multilingual AI agents with advanced search & retrieval

COMMAND
EMBED
RERANK

deepseek

Advanced reasoning models that solve complex problems step-by-step

DEEPSEEK

Luma

High-quality video generation with natural, coherent motion & ultra-realistic details

RAY

Meta

Advanced image and language reasoning

LLAMA

MISTRAL AI

Specialized expert models for agentic reasoning and multimodal tasks

MISTRAL
MIXTRAL
PIXTRAL

OpenAI

Automate tasks, enhance creativity, and solve complex problems efficiently

GPT-OSS

poolside

Software engineering AI for large enterprises

Coming soon

Qwen

Advanced reasoning with agentic intelligence

QWEN

stability.ai

Professional-grade images with creative control, deployable at scale

STABLE DIFFUSION
STABLE IMAGE

TwelveLabs

CTRL + F for video data: unlock the full potential of enterprise video assets

MARENGO
PEGASUS

WRITER

Purpose-built models for building and scaling AI agents across the enterprise

PALMYRA



Amazon Nova Foundation Models

State-of-the-art foundation models that deliver frontier intelligence and industry-leading price performance

UNDERSTANDING MODELS

Amazon Nova
Micro

Amazon Nova
Lite

Amazon Nova
Pro

Amazon Nova
Premier

CREATIVE CONTENT GENERATION MODELS

Amazon Nova
Canvas

Amazon Nova
Reel

SPEECH - TO - SPEECH MODEL

Amazon Nova
Sonic

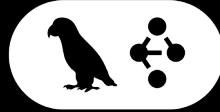


Frameworks for building agents

EMBRACING OPEN SOURCE



STRANDS
AGENTS SDK



LANGGRAPH,
LANGCHAIN

OpenAI



CREW.AI



GOOGLE ADK



LLAMAINDEX

+ many more

< Hello, World! >

Strands Agents

Strands Agents is an open source python SDK for building agents using just a few lines of code



The image shows a laptop screen with a dark background. On the screen, the GitHub repository page for "Strands Agents" is displayed. At the top, there are navigation links: README, Code of conduct, License, and Security. Below the header, the title "Strands Agents" is shown in bold. Underneath the title, it says "A model-driven approach to building AI agents in just a few lines of code." There are links for Docs, Samples, Tools, and Agent Builder. A paragraph describes the tool as a simple yet powerful SDK for building AI agents. The "Feature Overview" section lists several bullet points: Lightweight & Flexible, Model Agnostic, Advanced Capabilities, and Built-in MCP.

• Lightweight & Flexible: Simple agent loop that just works and is fully customizable

• Model Agnostic: Support for Amazon Bedrock, Anthropic, Ollama, and custom providers

• Advanced Capabilities: Multi-agent systems, autonomous agents, and streaming support

• Built-in MCP: Native support for Model Context Protocol (MCP) servers, enabling access to thousands of



```
from strands import Agent, tool
from strands_tools import calculator
from strands.models import BedrockModel
import json
```



```
from strands import Agent, tool
from strands_tools import calculator
from strands.models import BedrockModel
import json

@tool
def weather():
    """ Get the weather """
    return "sunny"
```

```
● ● ●

@tool
def weather():
    """ Get the weather """
    return "sunny"

model_id = "us.anthropic.claude-3-7-sonnet-20250219-v1:0"

model = BedrockModel(
    model_id=model_id,
)

agent = Agent(
    model=model,
    tools=[calculator, weather],
    system_prompt="You're a helpful assistant. You can perform simple math
calculations and tell the weather."
)
```



```
def strands_agent_bedrock(payload):
    """
    Invoke the agent with a payload
    """
    user_input = payload.get("prompt")
    response = agent(user_input)
    return response.message['content'][0]['text']
```



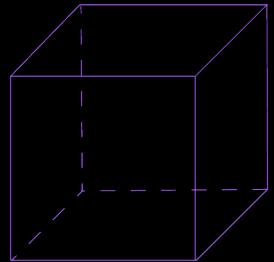
Deploy and operate agents with Amazon Bedrock AgentCore

Build, deploy, and operate highly capable agents securely at scale using any framework and model – no infrastructure management required

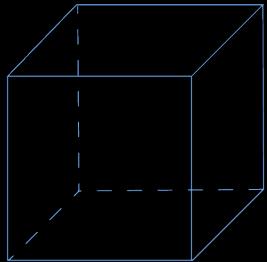
The prototype to production “chasm”



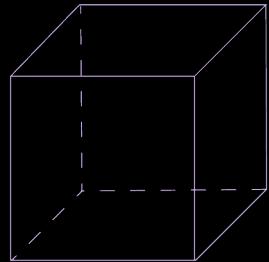
GETTING AGENTS TO PRODUCTION IS STILL TOO HARD



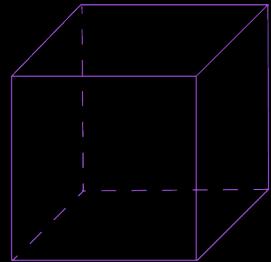
Securely execute
and scale agent
code



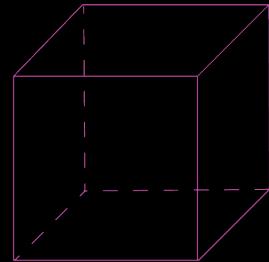
Remember past
interactions +
learning



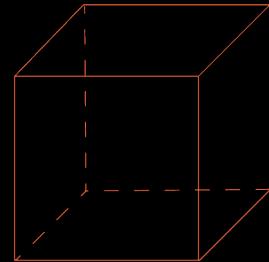
Identity and access
controls for all
agents and tools



Agentic tool use for
executing complex
workflows



Discover and connect
with custom tools
and resources

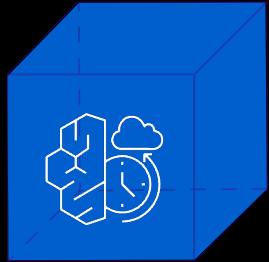


Understand and
audit every
interaction

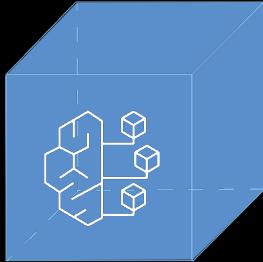
GENERALLY AVAILABLE

Amazon Bedrock AgentCore

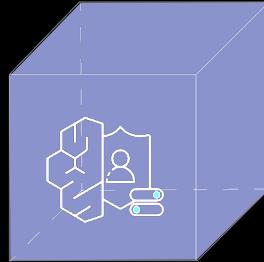
Comprehensive agentic platform: Everything you need for getting agents into production



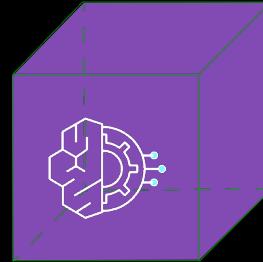
Runtime



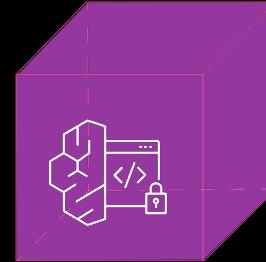
Memory



Identity



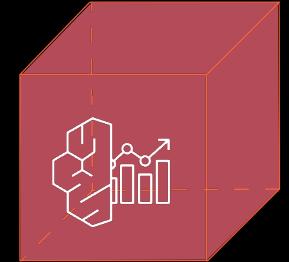
Gateway



Code
Interpreter

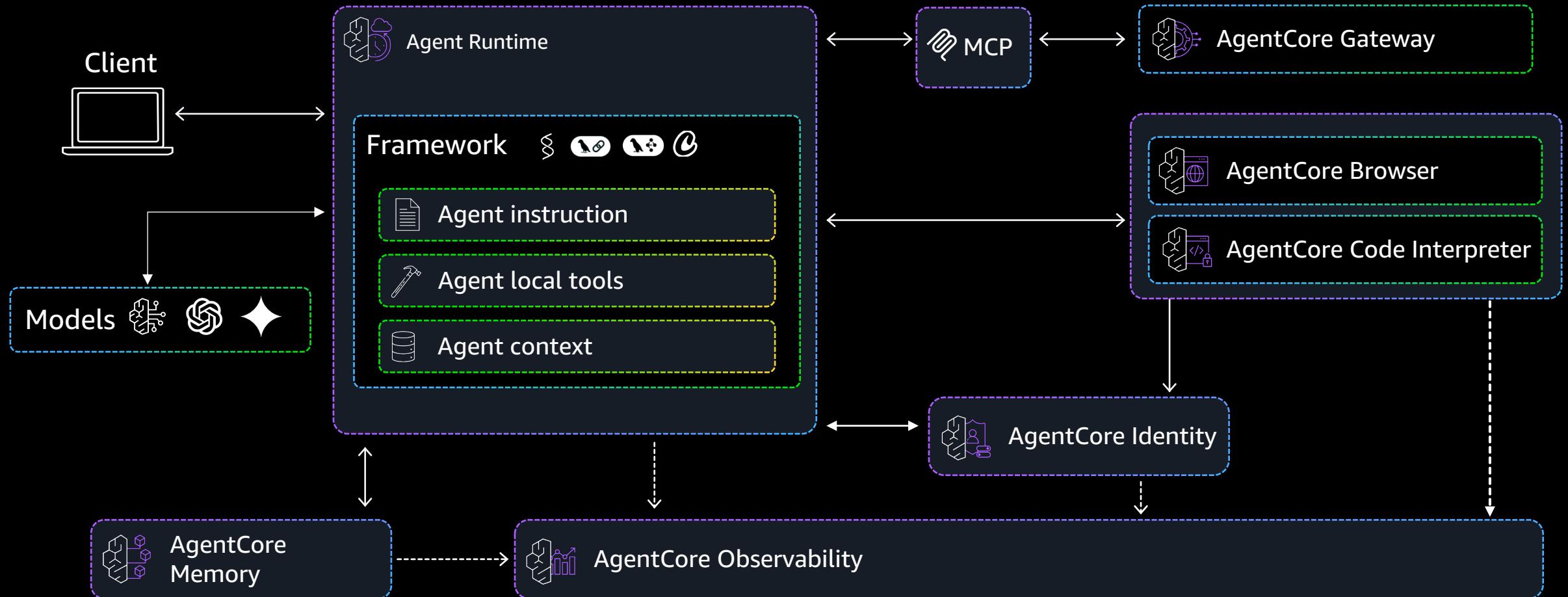


Browser Tool



Observability

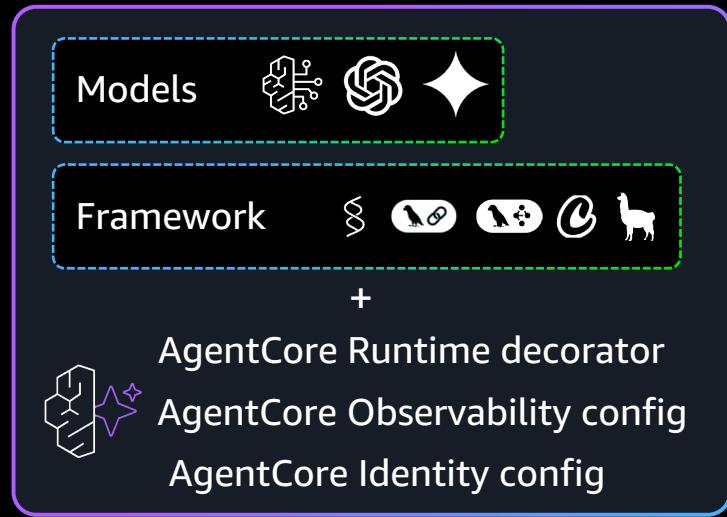
AgentCore services enabling agents at scale





AgentCore Runtime

Agent or tool code



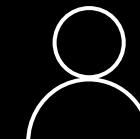
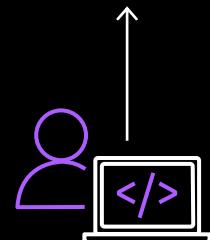
configure



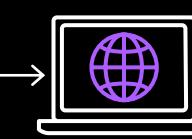
AgentCore Runtime

Amazon ECR Repository

launch



User



Application

invoke





```
from strands import Agent, tool
from strands_tools import calculator
import argparse
import json
from strands.models import BedrockModel

@tool
def weather():
    """ Get the weather """
    return "sunny"

model_id = "us.anthropic.claude-3-7-sonnet-20250219-v1:0"
model = BedrockModel(
    model_id=model_id,
)
agent = Agent(
    model=model,
    tools=[calculator, weather],
    system_prompt="You're a helpful assistant. You can perform simple math calculations and tell the weather."
)

def strands_agent_bedrock(payload):
    """
    Invoke the agent with a payload
    """
    user_input = payload.get("prompt")
    response = agent(user_input)
    return response.message['content'][0]['text']
```

```
from strands import Agent, tool
from strands_tools import calculator
import argparse
import json
from strands.models import BedrockModel
from bedrock_agentcore.runtime import BedrockAgentCoreApp ←

app = BedrockAgentCoreApp() ←

@tool
def weather():
    """ Get the weather """
    return "sunny"

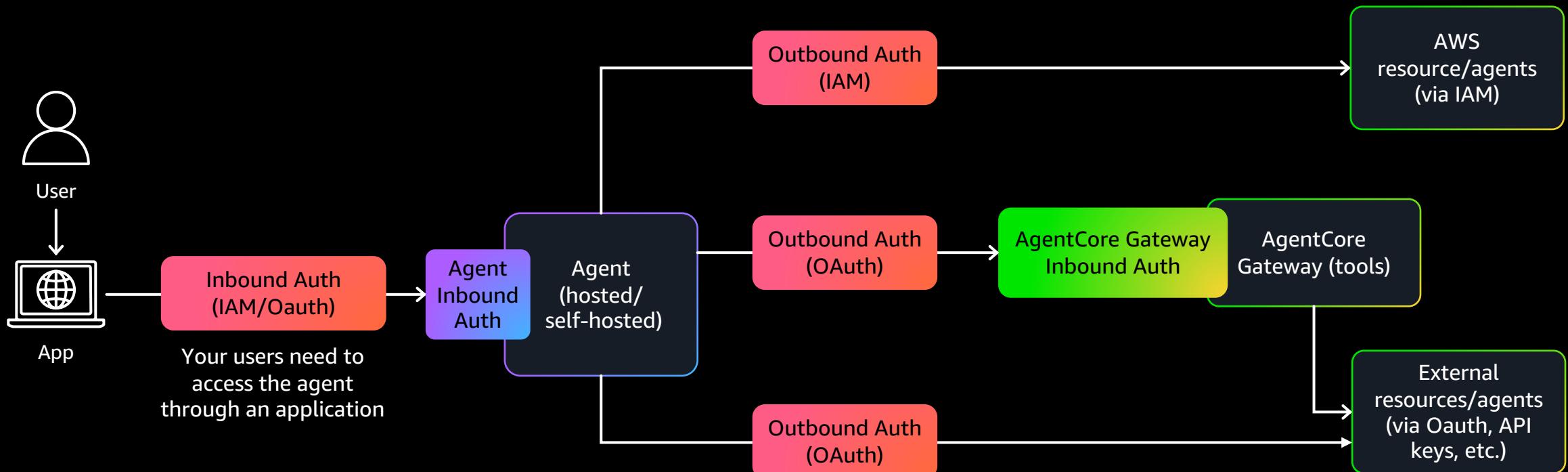
model_id = "us.anthropic.claude-3-7-sonnet-20250219-v1:0"
model = BedrockModel(
    model_id=model_id,
)
agent = Agent(
    model=model,
    tools=[calculator, weather],
    system_prompt="You're a helpful assistant. You can perform simple math calculations and tell the weather."
)

@app.entrypoint ←
def strands_agent_bedrock(payload):
    """
    Invoke the agent with a payload
    """
    user_input = payload.get("prompt")
    response = agent(user_input)
    return response.message['content'][0]['text']

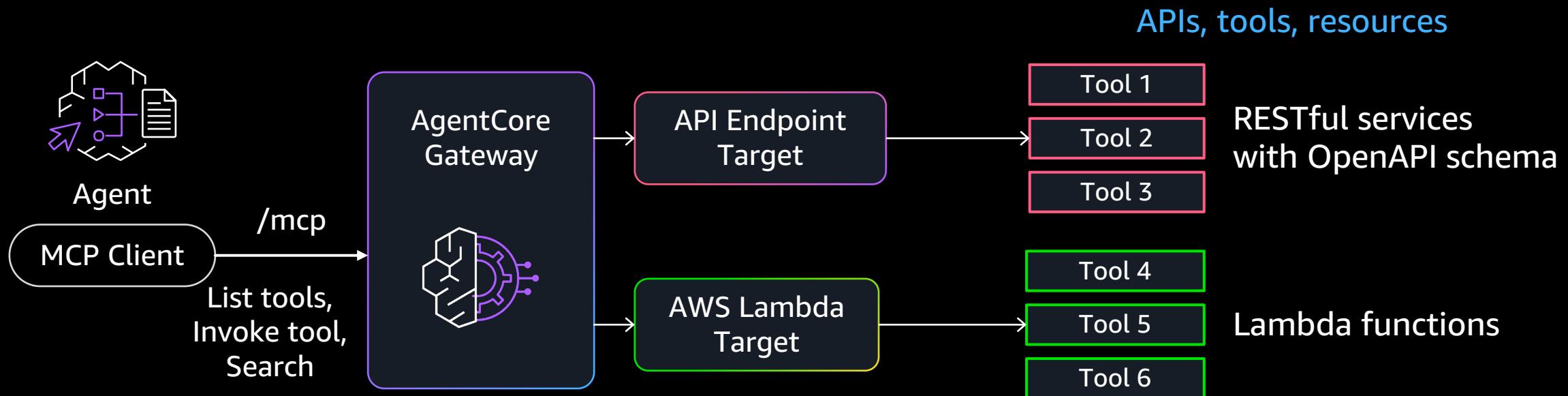
if __name__ == "__main__":
    app.run() ←
```



AgentCore Identity

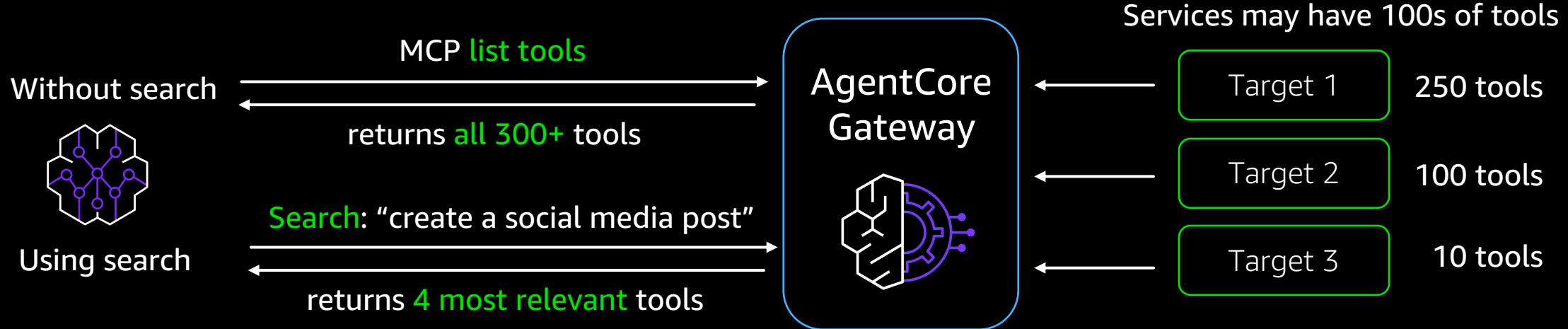


AgentCore Gateway



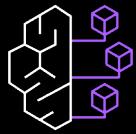


AgentCore Gateway semantic search

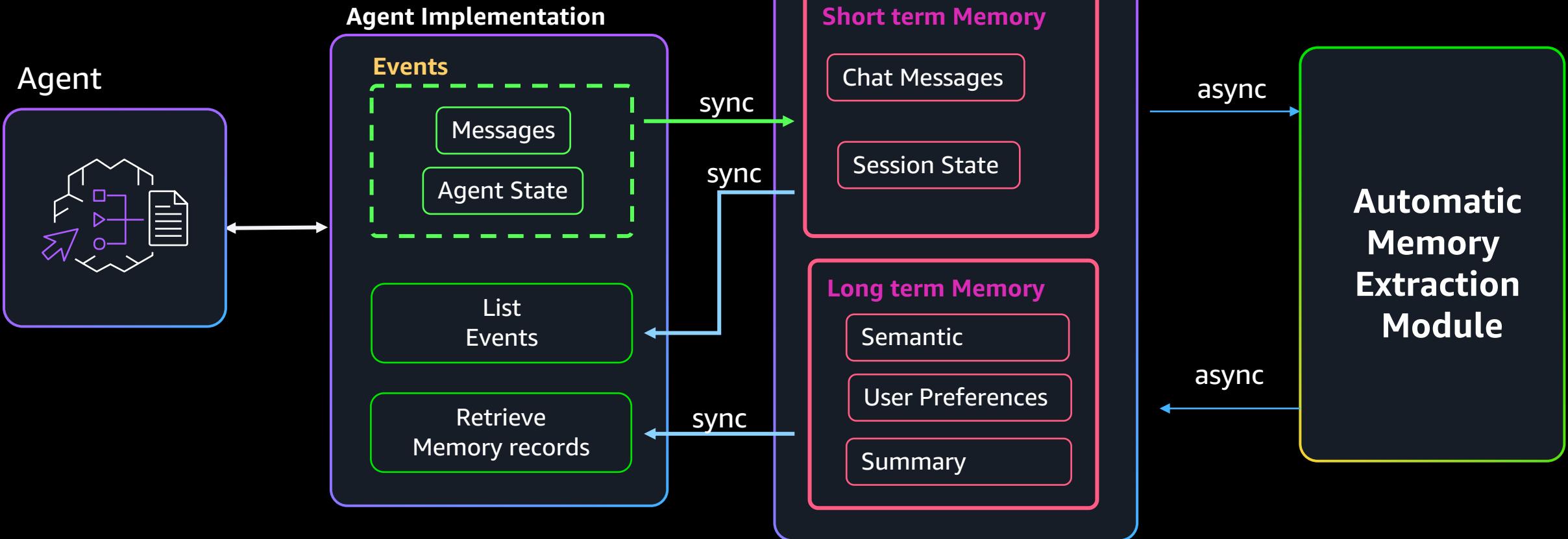


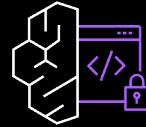
Benefits

- 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

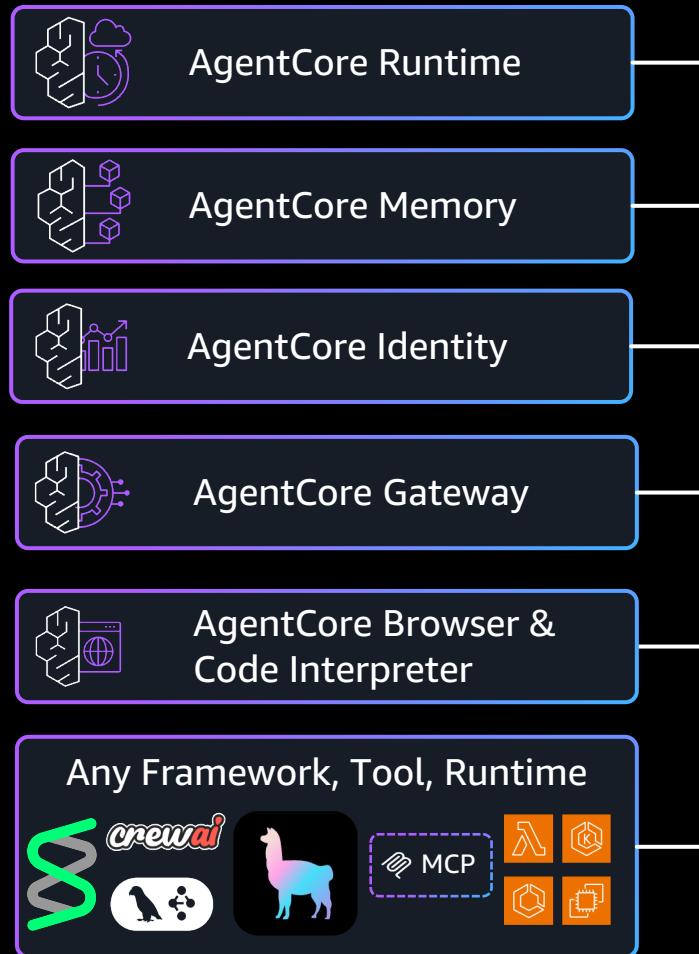


AgentCore Memory

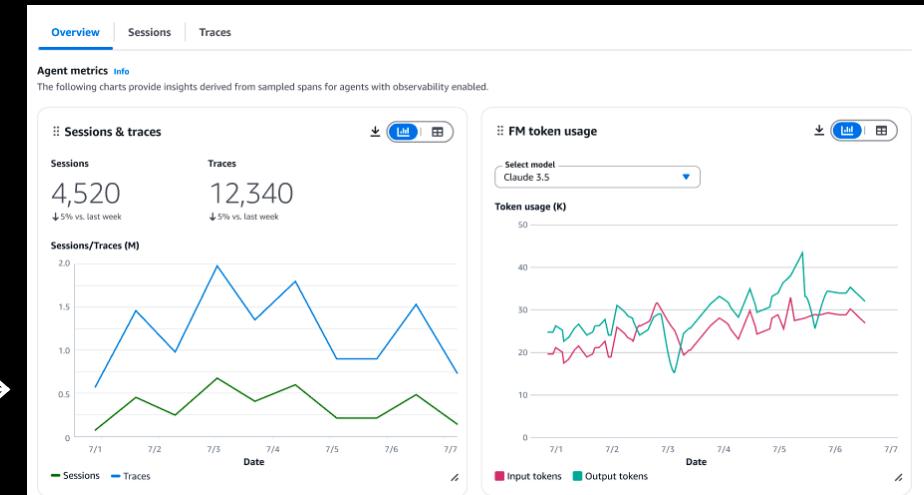




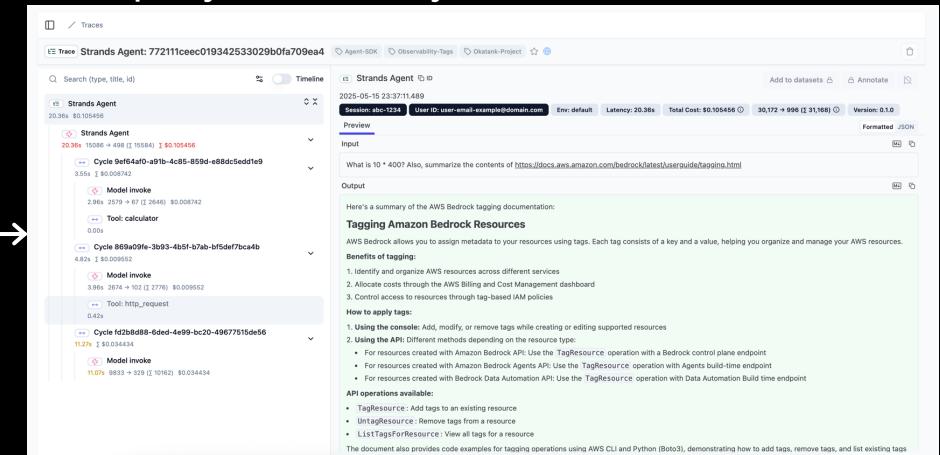
AgentCore Observability



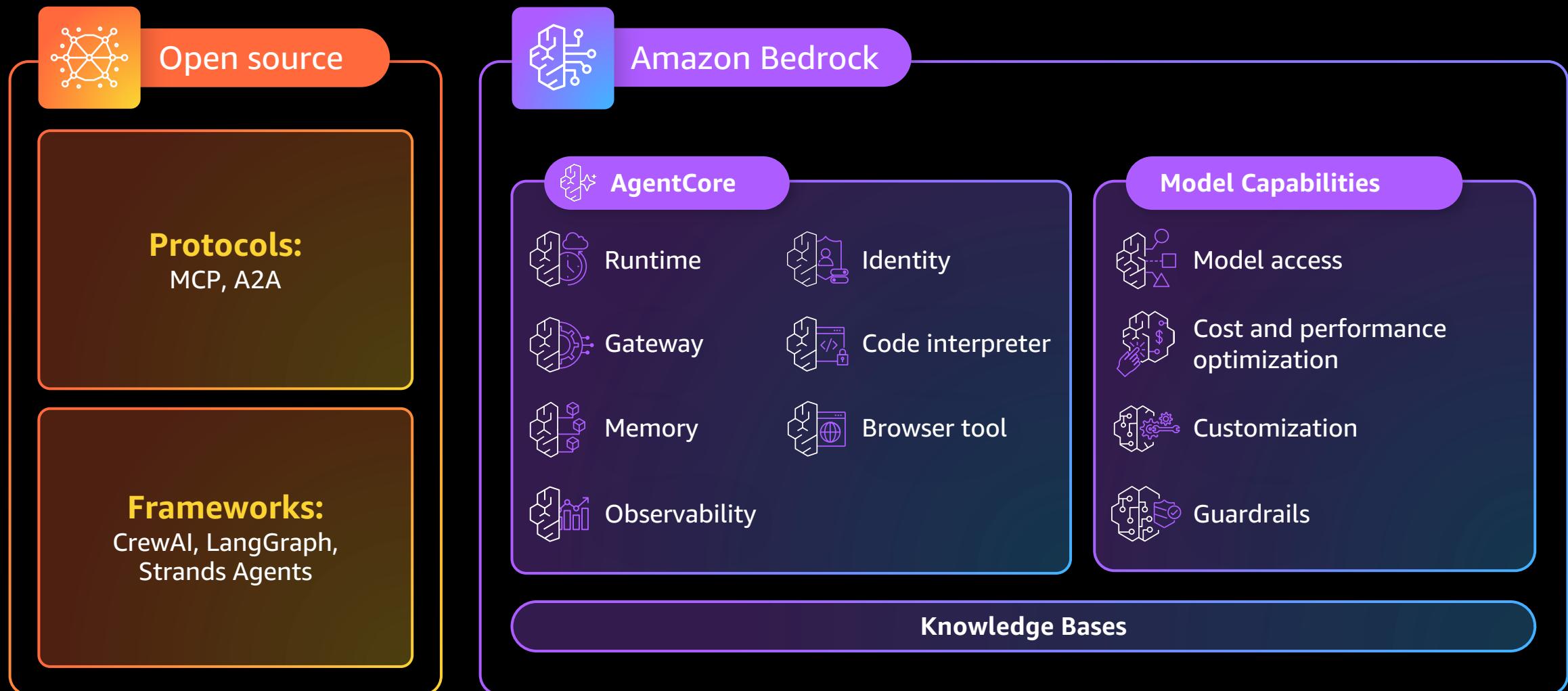
AgentCore Observability dashboards

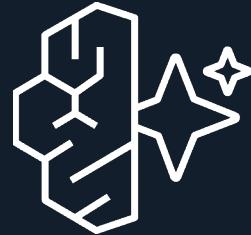


Third-party observability dashboards



Amazon Bedrock





Demo: Multi-Agent Clinical Coding Solution

Resources



© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved. Amazon Confidential and Trademark.

Get started with Strands Agents

Start your agentic AI journey today



Docs

Review the user guide, quick start, API documentation for best practices



SDK

Get started building agents and contribute to the open-source project on GitHub



Tools

Access pre-built tools to get started experimenting with agents



Samples

Discover sample agents in our samples repository

Get started now!



AgentCore
Overview



AgentCore
Samples



AgentCore
Documentation



AgentCore
blog post

Public github repo of samples and tutorials

The screenshot shows the GitHub repository page for `aws-labs/amazon-bedrock-agentcore-samples`. The repository is public and has 118 stars, 16 forks, and 1 watching. It contains 21 branches and 0 tags. The main branch has 12 commits from user `w601sxs` made yesterday. The commits are related to adding tutorials, samples, and integrations for the Amazon Bedrock AgentCore Gateway. The repository also includes files like `CODE_OF_CONDUCT.md`, `CONTRIBUTING.md`, `LICENSE`, `NOTICE`, and `README.md`. The repository page also features an "About" section describing the purpose of the project and its tags.

About

Amazon Bedrock Agentcore accelerates AI agents into production with the scale, reliability, and security, critical to real-world deployment.

agent authentication runtime core
gateway bedrock
memory-management
identity-management agents
production-code agentic-ai

Readme
Apache-2.0 license
Code of conduct
Security policy
Activity
Custom properties
118 stars
1 watching
16 forks

<https://github.com/awslabs/amazon-bedrock-agentcore-samples>

Developer guide and API reference docs

The screenshot shows the AWS developer documentation for Amazon Bedrock AgentCore. At the top, there's a navigation bar with links for 'Get started', 'Service guides', 'Developer tools', 'AI resources', a search bar, and buttons for 'Return to the Console' and 'Feedback'. Below the header, the main content area has a dark header with the title 'Amazon Bedrock AgentCore' and a subtext: 'Amazon Bedrock AgentCore enables you to deploy and operate highly capable agents securely, at scale using any framework and model.' The main content is organized into three sections: 'Amazon Bedrock AgentCore Developer Guide', 'Amazon Bedrock AgentCore Data Plane API Reference', and 'Amazon Bedrock AgentCore Control Plane API Reference'. Each section includes a brief description and links to 'HTML' and 'PDF' versions. On the right side of the content area, there's a sidebar titled 'On this page' which is currently empty. At the bottom of the page, there are links for 'Privacy', 'Site terms', 'Cookie preferences', and a copyright notice: '© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.'

Explanation of each service, key concepts, getting started, advanced topics, detailed API specs, ...

<https://docs.aws.amazon.com/bedrock-agentcore/>



FAQ's

Amazon Bedrock AgentCore (Preview) FAQs

Page topics

General	11
Runtime	2
Gateway	6
Memory	2
Code Interpreter	2
Browser Tool	2

General

What is Amazon Bedrock AgentCore?

AgentCore enables developers to accelerate AI agents into production with the scale, reliability, and security, critical to real-world deployment. AgentCore provides services and tools to make agents more effective and capable, purpose-built infrastructure to securely scale agents, and controls to operate trustworthy agents. AgentCore services and tools are composable and work with any open-source frameworks and any model, so you don't have to choose between open-source flexibility and enterprise-grade security and reliability.

<https://aws.amazon.com/bedrock/agentcore/faqs/>



AgentCore videos

7-minute demo video



<https://youtu.be/usFlb9aEd1U>

2-minute explainer video



<https://youtu.be/YDqTaZ4dpXc>

Thank you!

