

# Agentic AI on AWS

*Amazon Bedrock AgentCore Workshop*

Daniel Pruessner ([dpkey@amazon.com](mailto:dpkey@amazon.com)), Solution Architect covering BNZ

Rob van der Harst ([rharst@amazon.com](mailto:rharst@amazon.com)), Account Director covering BNZ



Whātuia te rangi e tū nei	Tie together all that is above
Whātuia te papa e takoto nei	Tie together all that is from below
Korowaitia ki te kahu ora	Nurture with the cloak of life
Korowaitia ngā pito mata, ngā kura huna	Nurture, All Potential and aspiration, knowledge and x-factor
Kia toitū	
Kia toiroa	To be permanent
Kia toi kairangi	To be sustainable
Kia puta, kia ora!	To be excellent
Hui e, tāiki e!	Release and thrive! Bind and confirm!

wh = fa

ng = like the sound in sing

r = rolled r, like a soft d

a – as in but

e – as in vet

i – as in beat

o – as in walk

u – as in to

ā – as in bar

ē – as in dairy

ī – as in peel

ō – as in bore

ū – as in too



# Agenda

Topic	Time	Duration(min)	Type
Welcome and Introduction (people arrive and have coffee)	9am	30	
Overview of Agentic AI	9:30am	15	Presentation
Use Case – Mortgage Assistant and Introduction to Strands	9:45am	15	Presentation
Running Strands Agent on local environment	10am	30	Lab
Morning Tea	10:30am	20	
Introduction to AgentCore Runtime	10:50am	15	Presentation
AgentCore Runtime Lab	11:05am	25	Lab
Introduction to AgentCore Identity	11:30am	15	Presentation
Introduction to AgentCore Gateway	11:45am	15	Presentation
Lunch	12pm	60	
AgentCore Gateway and Identity Lab	1pm	25	Lab
Introduction to AgentCore Memory	1:45pm	20	Presentation
AgentCore Memory Lab	2:05pm	20	Lab
Introduction to AgentCore Observability	2:40pm	15	Presentation
AgentCore Observability Lab	2:55pm	15	Lab
Introduction to AgentCore Tools	3:10pm	10	Presentation
AgentCore Tools Lab	3:25pm	20	Lab
Trivia/Quiz	3:45pm	10	

# Organizations are creating value with agentic AI

## Workplace productivity



Ex. Knowledge worker productivity,  
Software development

## Business workflows



Ex. Customer experience, incident  
management, demand forecasting

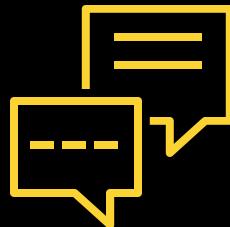
## Innovation and research



Ex. Automate complex data analysis  
and simulations



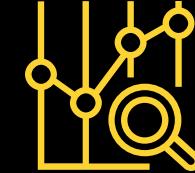
# What is Agentic AI?



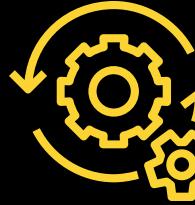
Access to  
enterprise data



Intelligent,  
autonomous  
systems

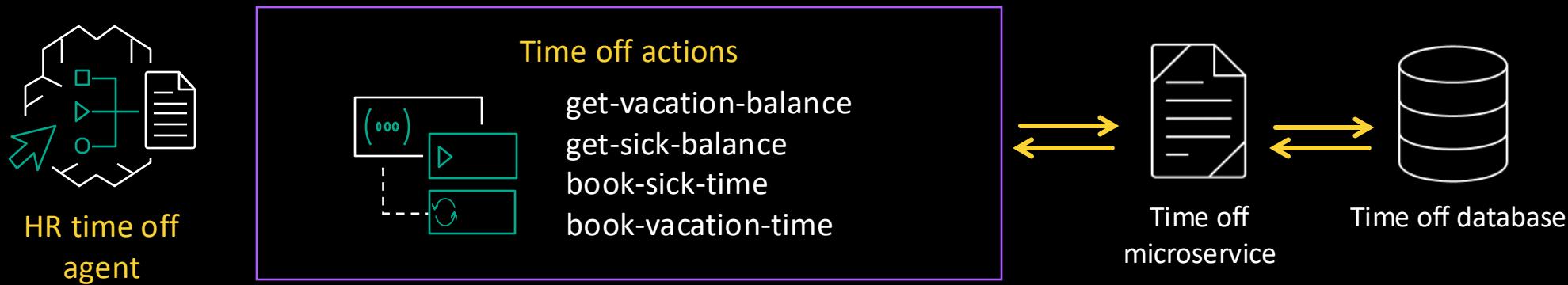


Plan, reason,  
and act



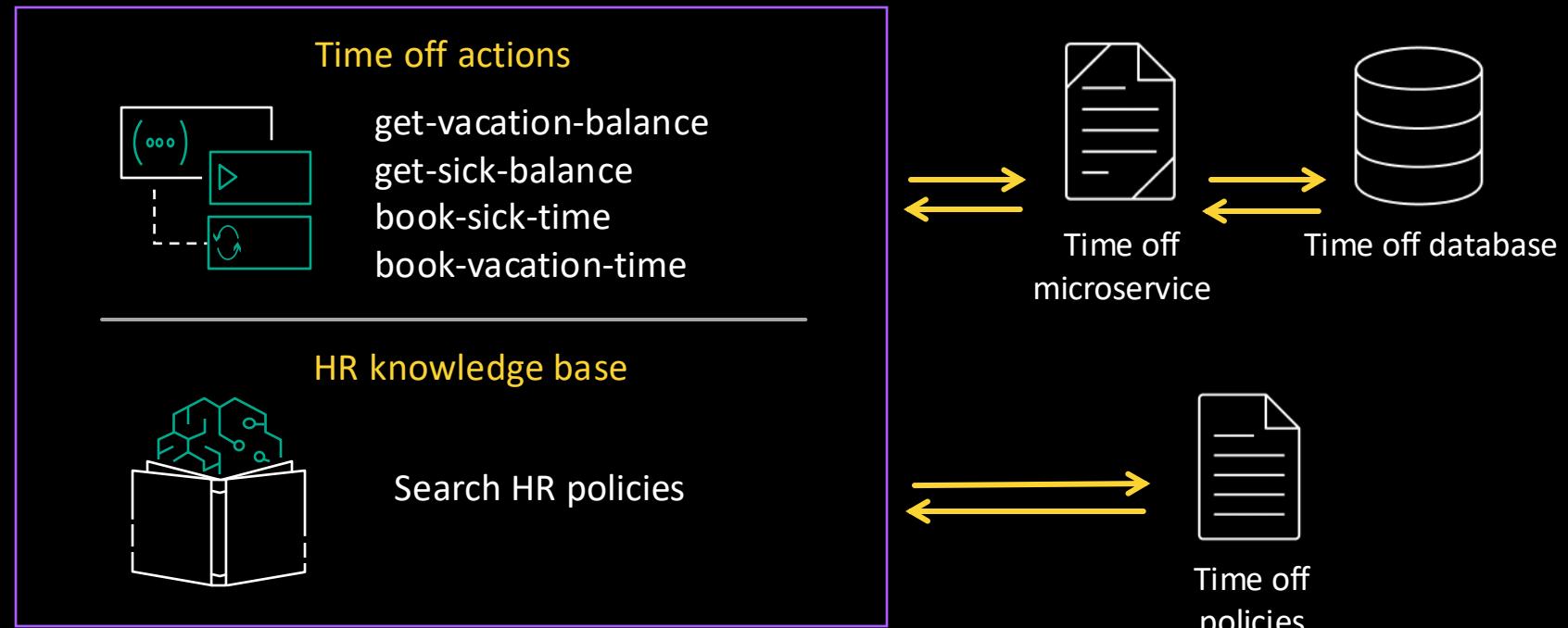
Ability to use  
tools

# Agents can start small and focused . . .

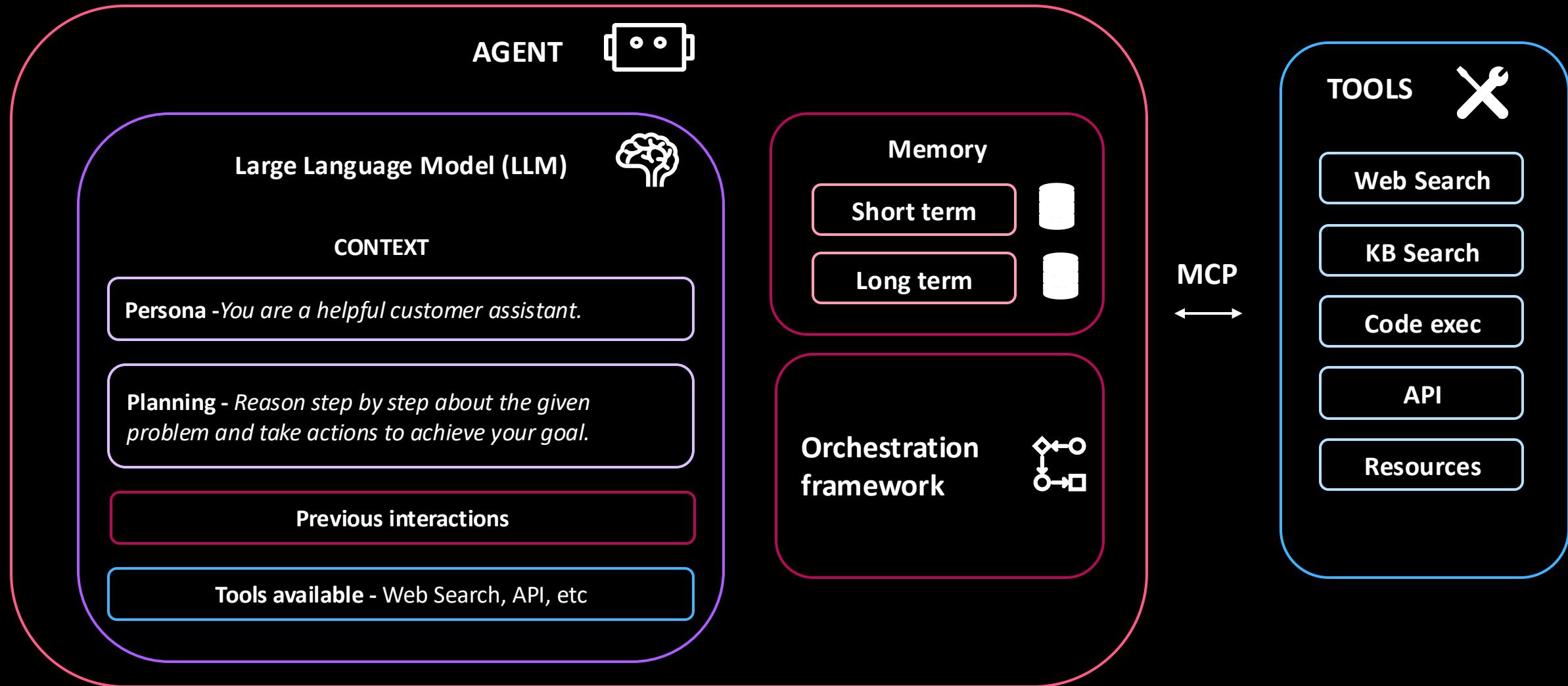


... and can be easily expanded

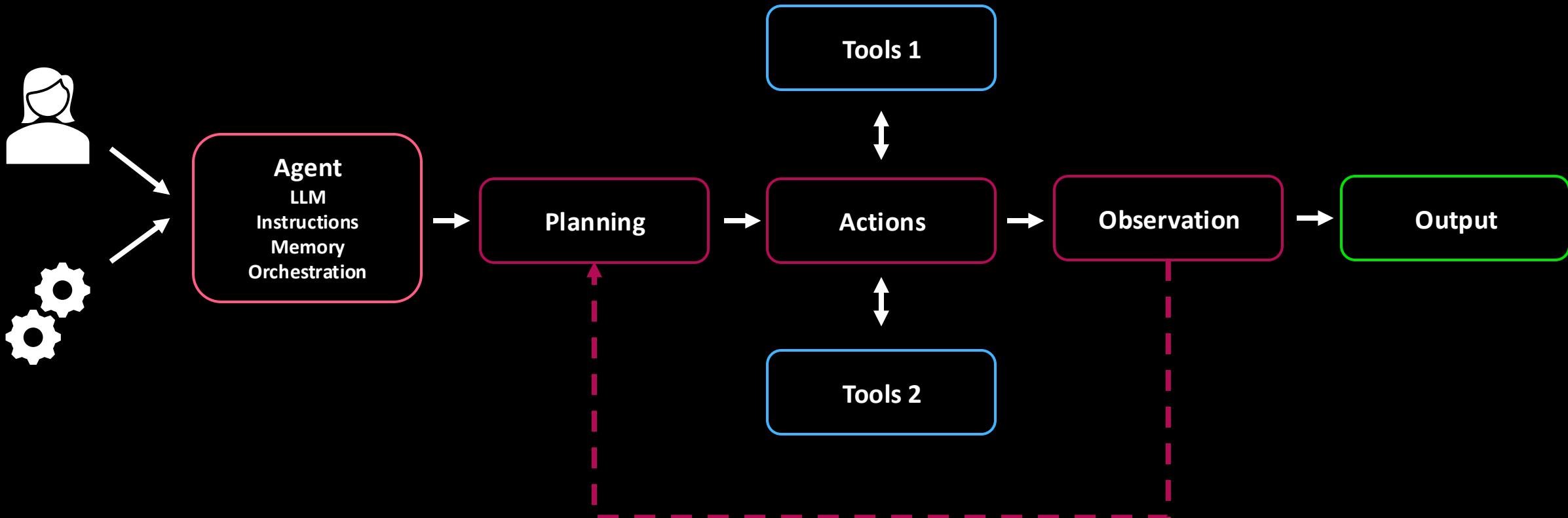
V2



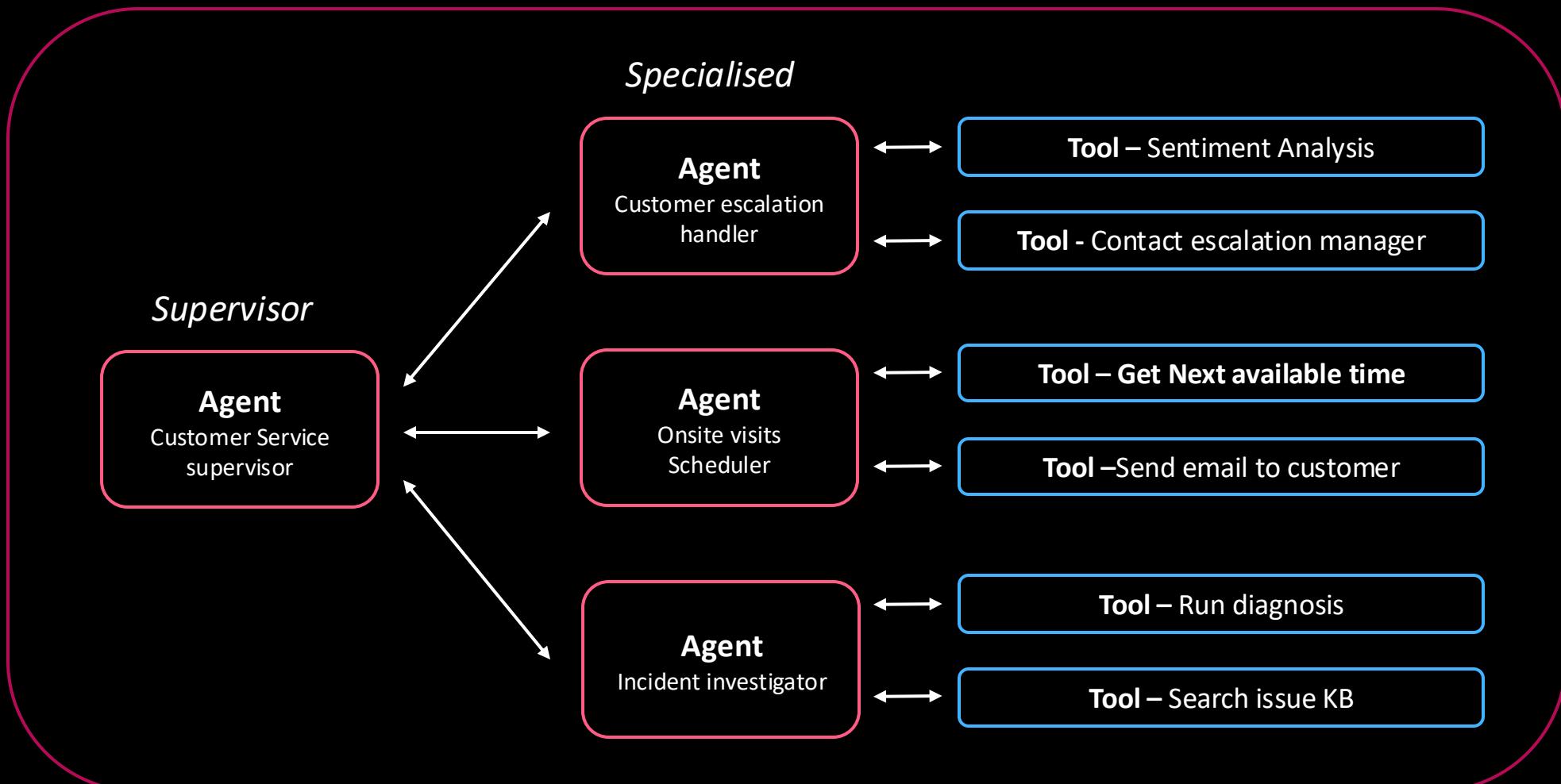
# Inside view of an agentic system



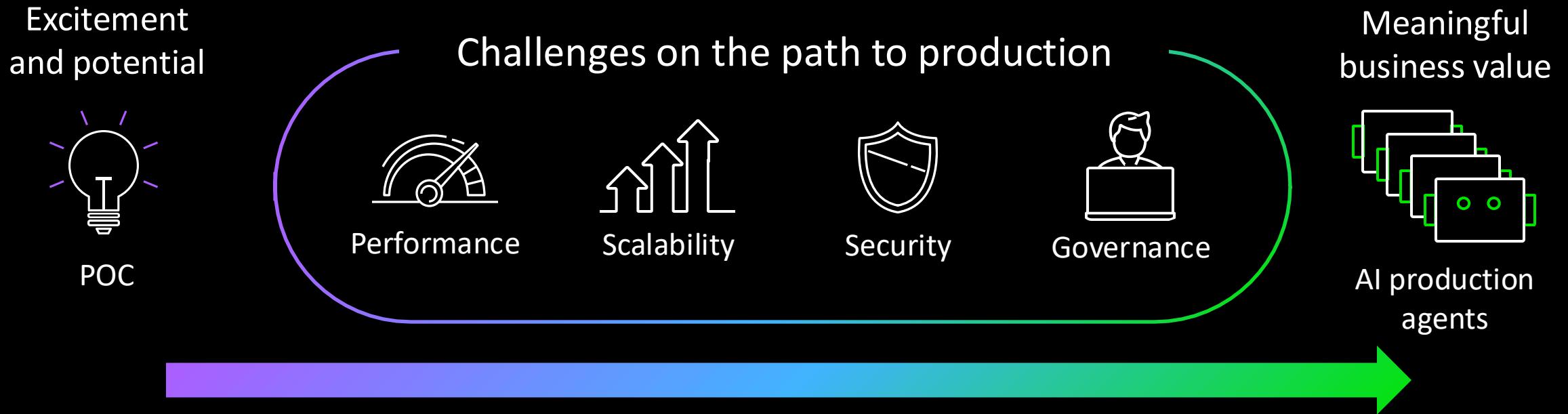
# Simple execution of a single agent



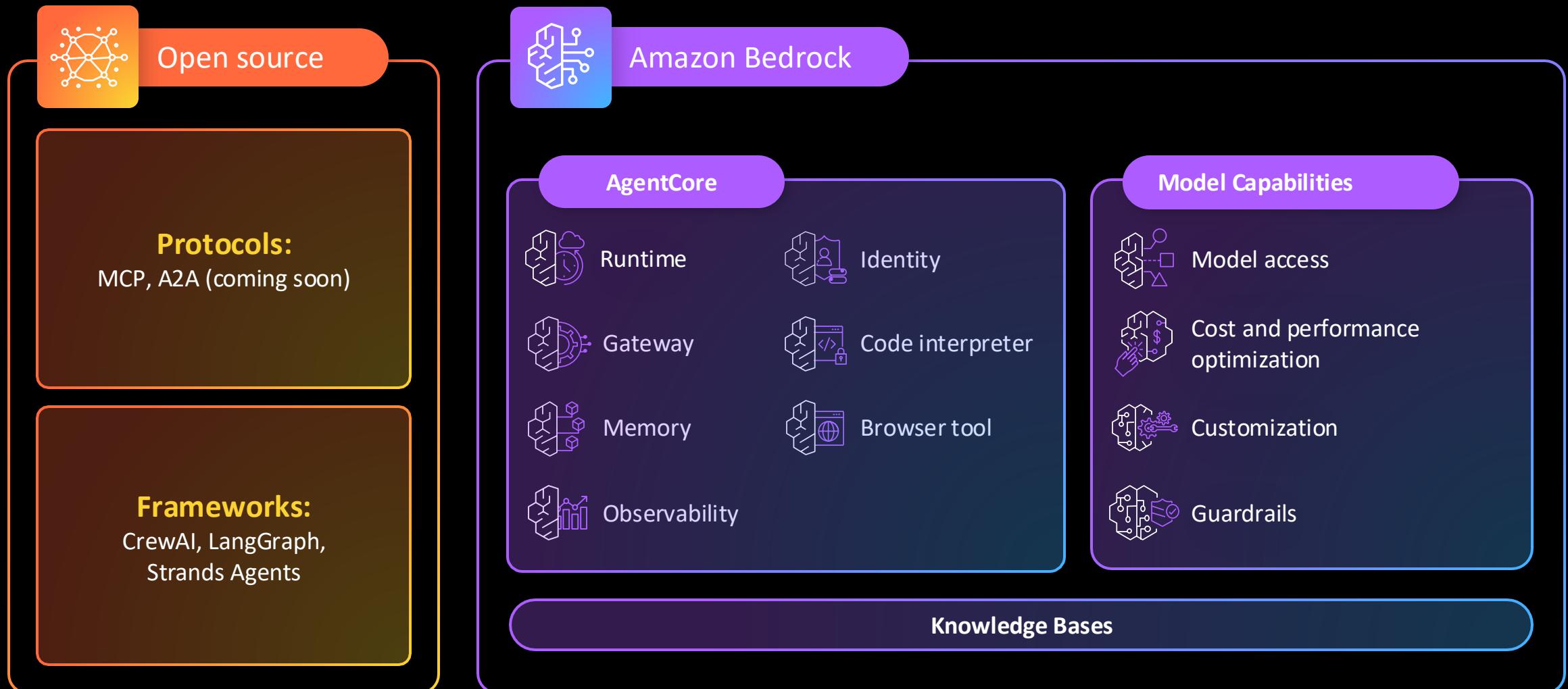
# Multi-Agents system examples



# The prototype to production “chasm”



# Agentic AI Primitives

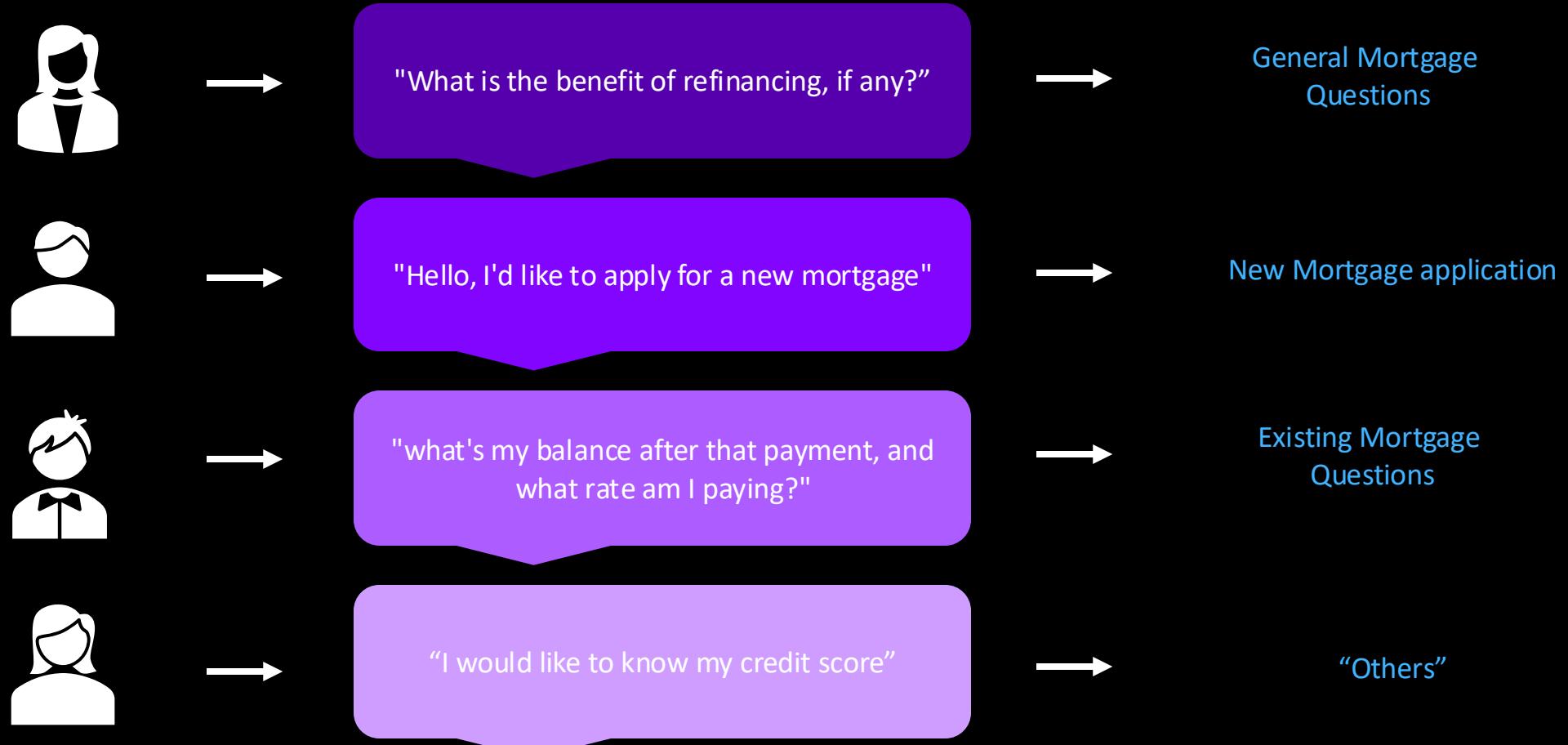




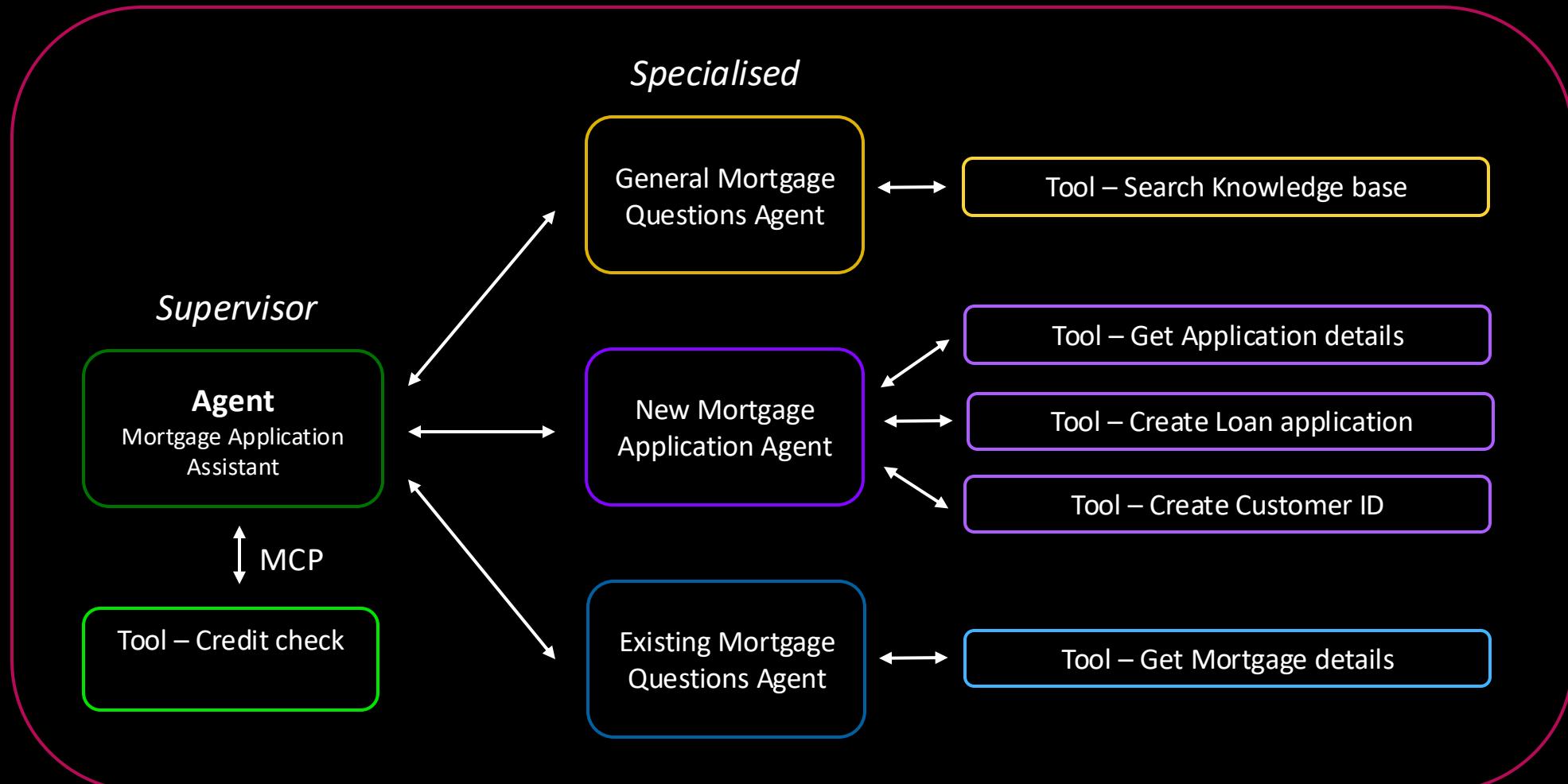
# The use case we're solving for

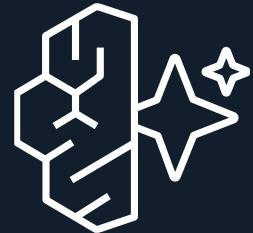
# Mortgage Chatbot Assistant

The bank's Mortgage department receives a flood of questions daily.

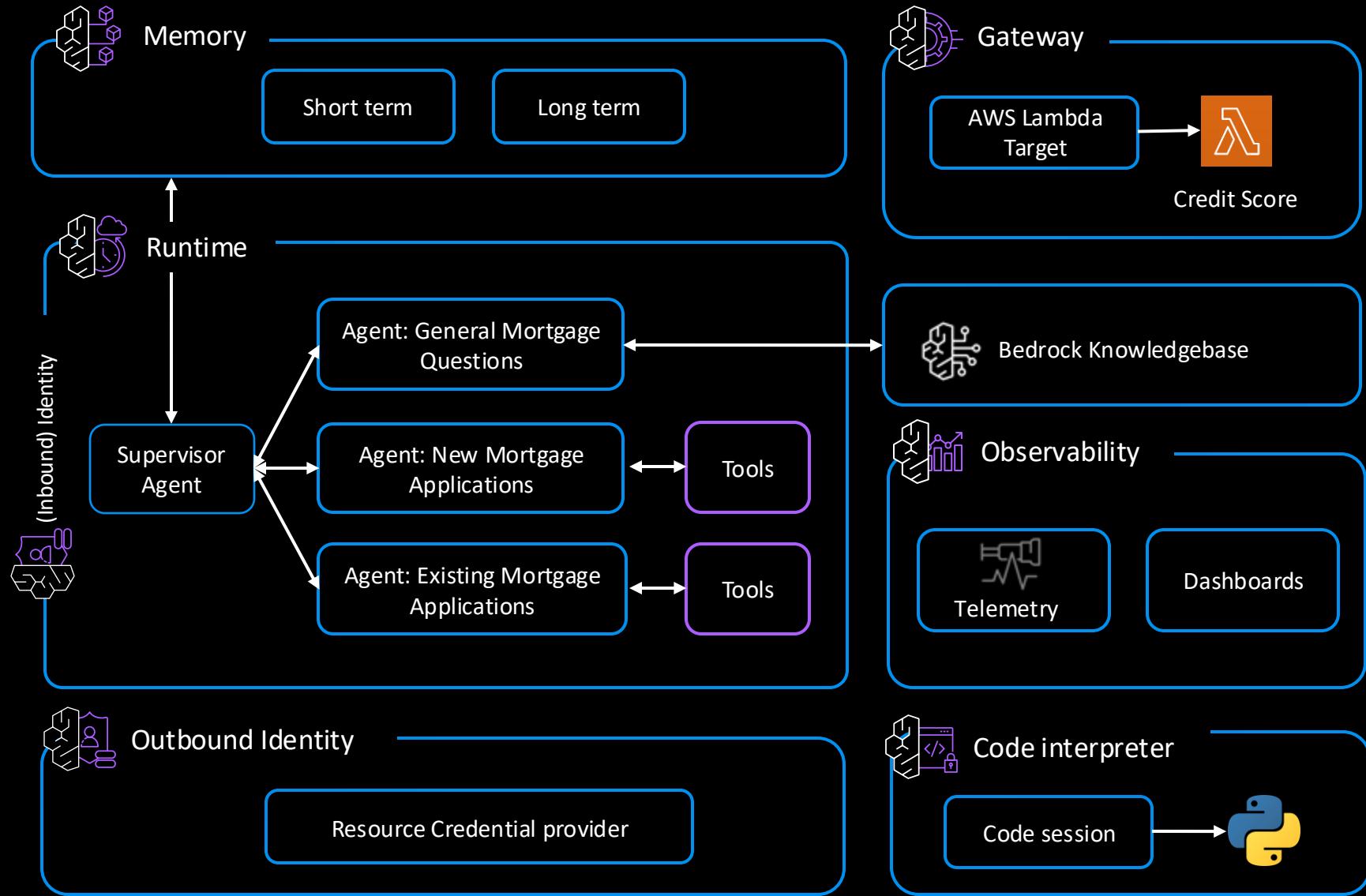


# Our Multi-Agents design – Agents as Tools pattern

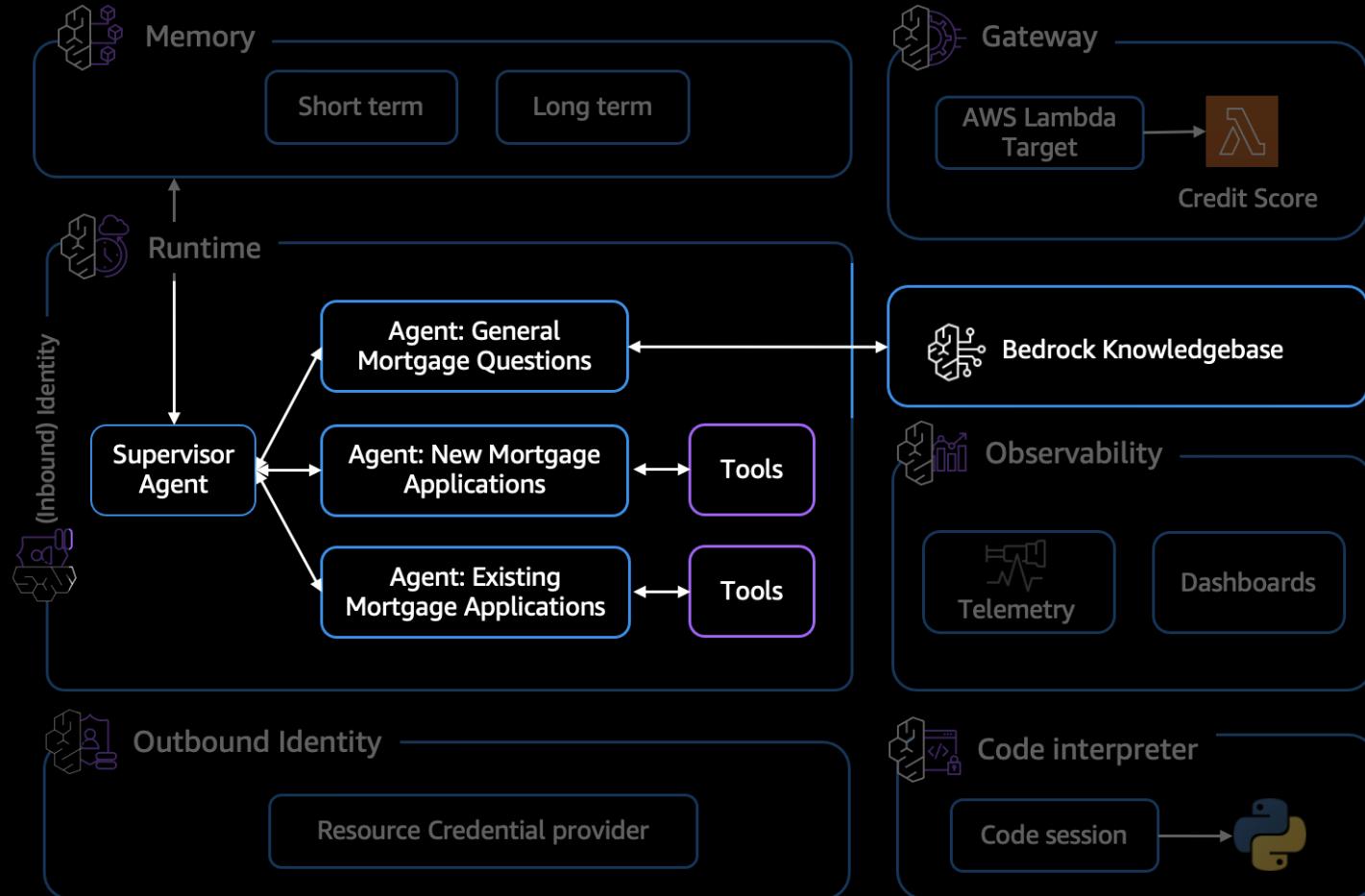




# What we're building – step by step



- Step 1 – Set up, KB
- Step 2 – Strands Agent
- Step 3 – Runtime
- Step 4 – Gateway
- Step 5 - Identity
- Step 6 – Memory
- Step 7 – Observability
- Step 8 – Tools



Step 1 – Set up, KB

Step 2 – Strands Agent

Step 3 – Runtime

Step 4 – Gateway

Step 5 (Optional)– Identity

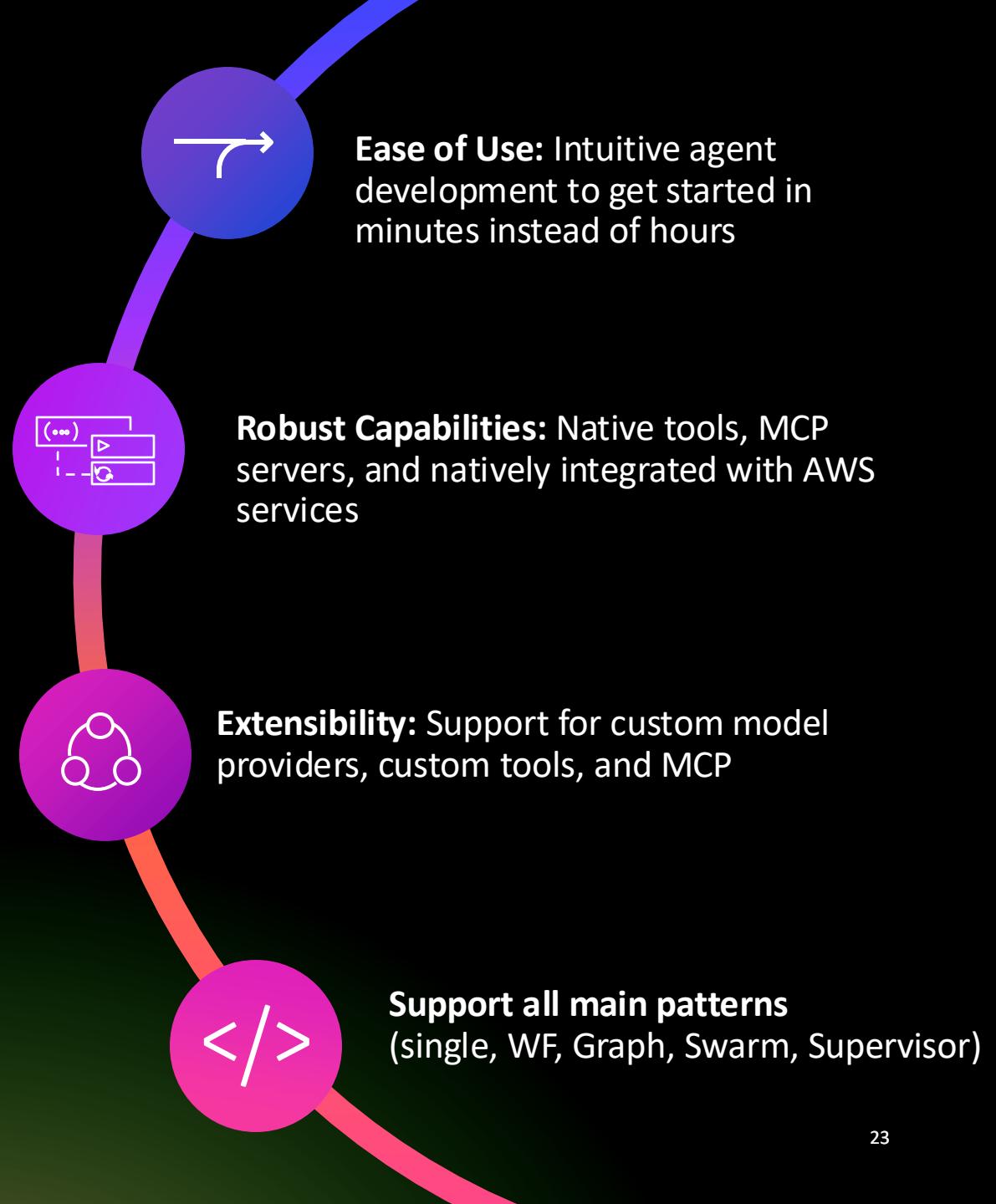
Step 6 – Memory

Step 7 – Observability

Step 8 (Optional) – Tool

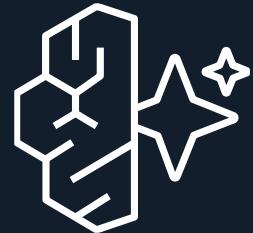
# Strands Agents

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





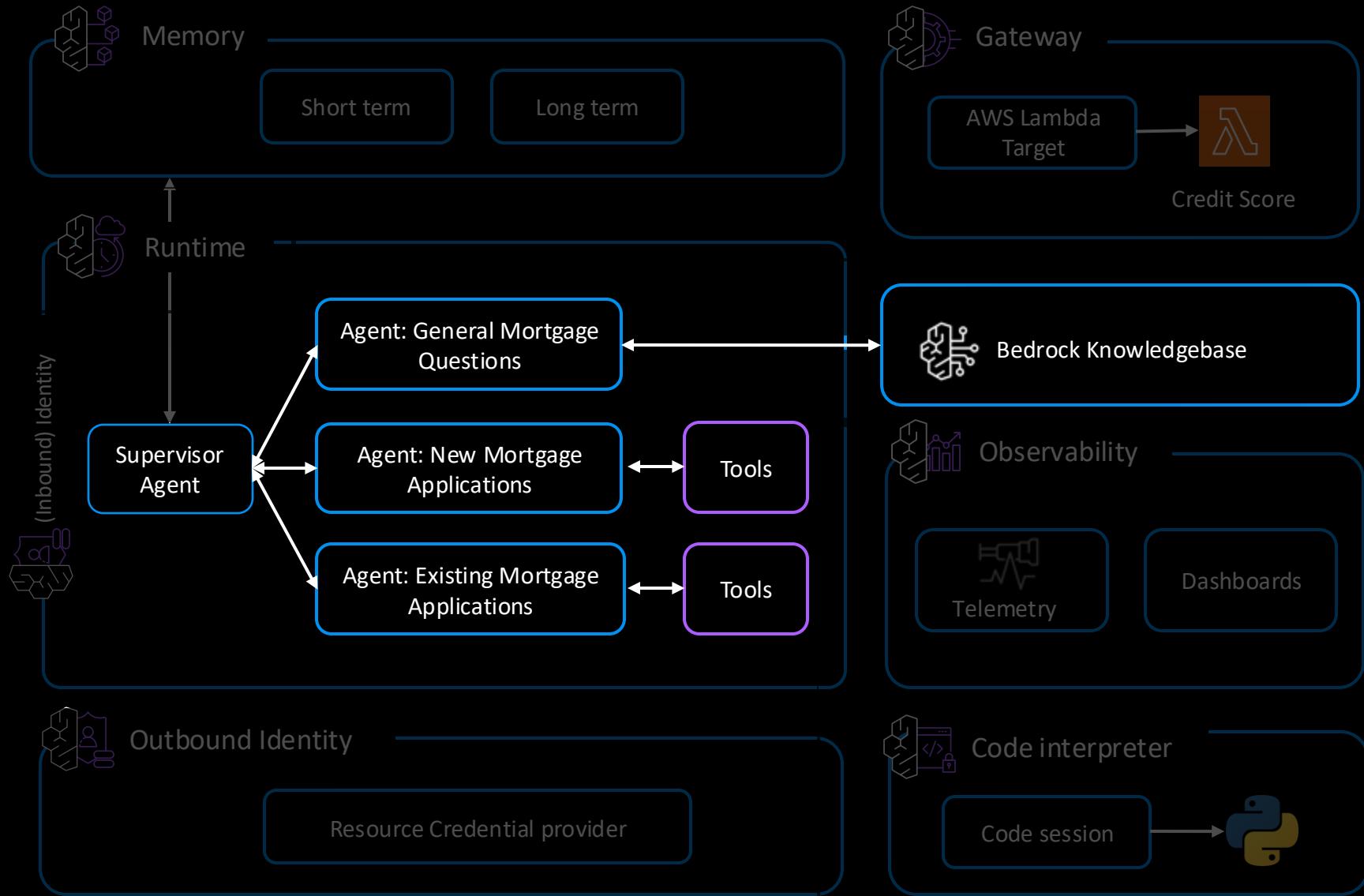
# Lab 01 and Lab 02



# AgentCore Runtime



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



Step 1 – Set up, KB

Step 2 – Strands Agent

Step 3 – Runtime

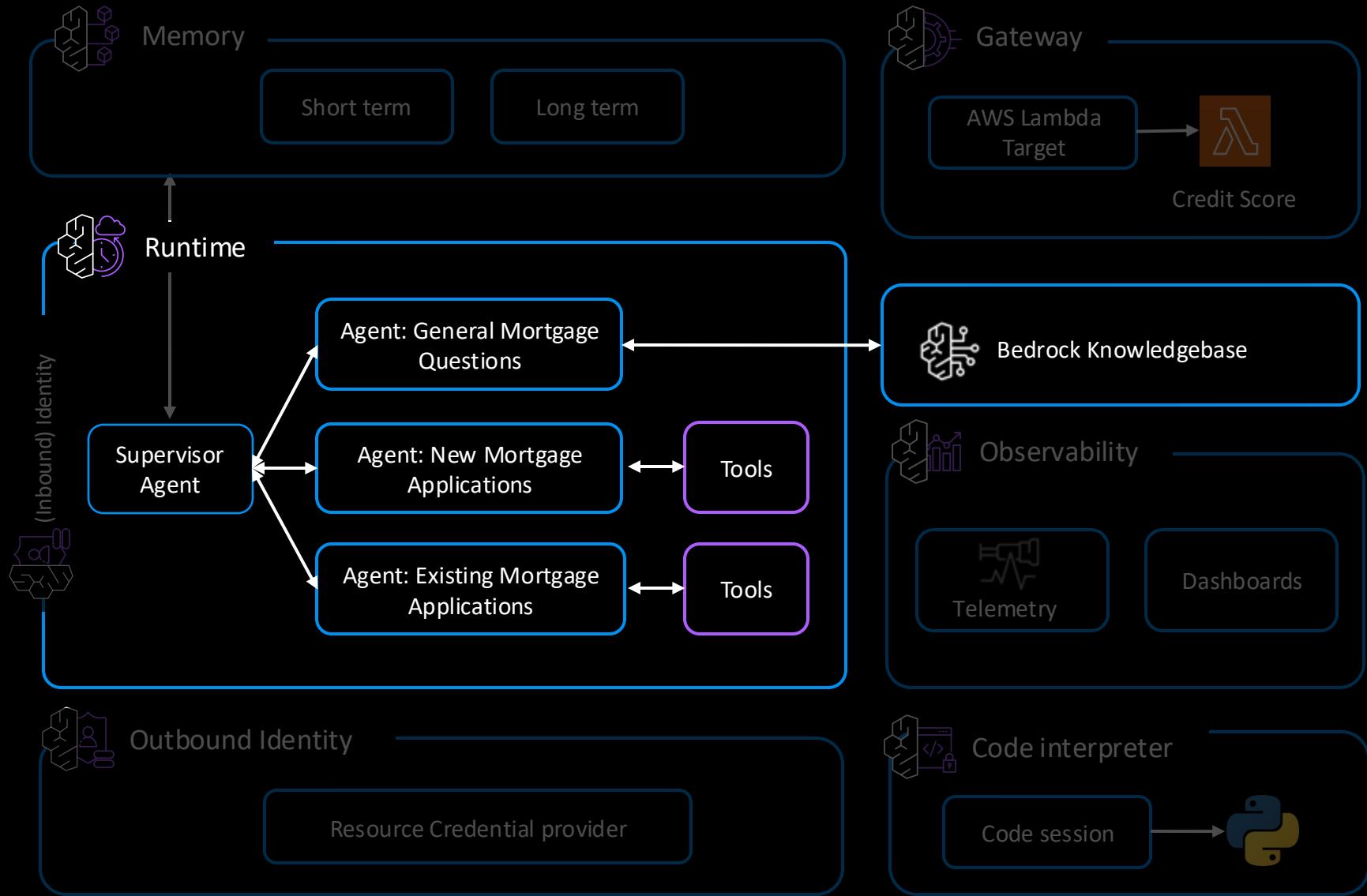
Step 4 – Gateway

Step 5 - Identity

Step 6 – Memory

Step 7 – Observability

Step 8 – Tools



Step 1 – Set up, KB

Step 2 – Strands Agent

Step 3 – Runtime

Step 4 – Gateway

Step 5 - Identity

Step 6 – Memory

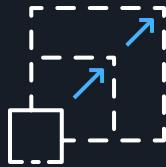
Step 7 – Observability

Step 8 – Tools



# AgentCore Runtime

Scale from real-time to multi-hour workloads



- Multi-modal
- Real time & long running (8h)
- Auto-scaling

Accelerate time to market



- Any frameworks
- Any models
- Starter toolkit to speed up deployment

Secure workload with enterprise-grade isolation

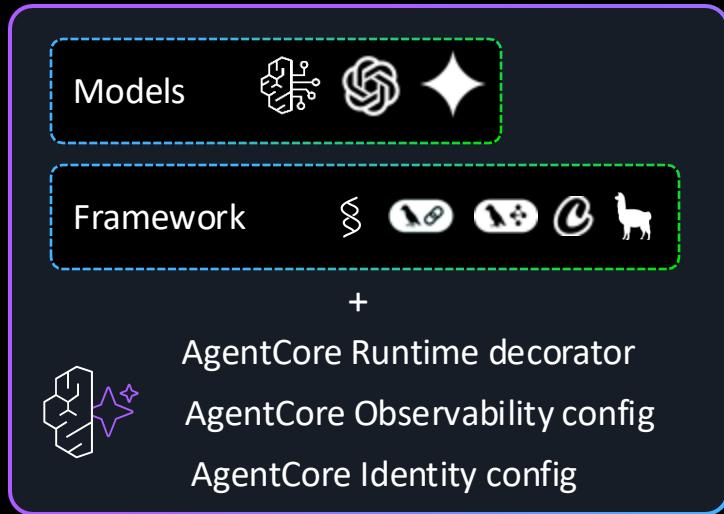


- MicroVMs
- True session isolation to protect your data
- Integrates with existing identity providers



# AgentCore Runtime

Agent or tool code

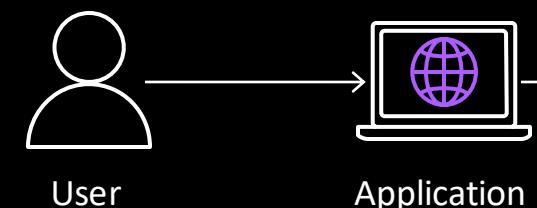
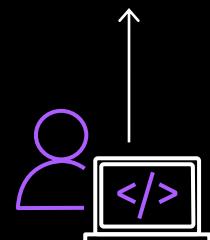
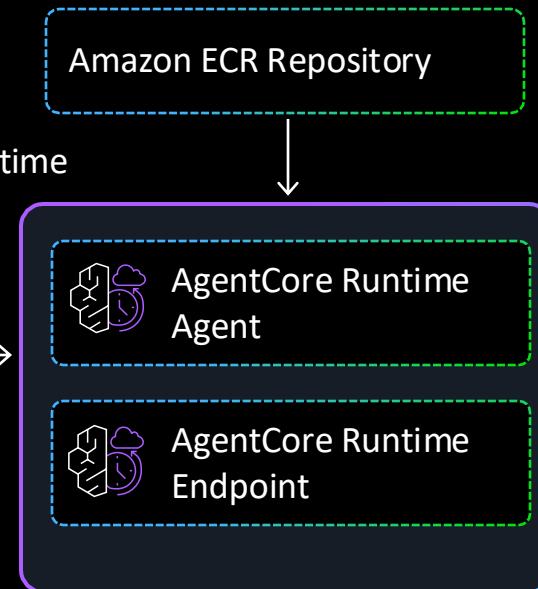


configure



AgentCore Runtime

launch



invoke





# AgentCore Runtime

```
● ● ●

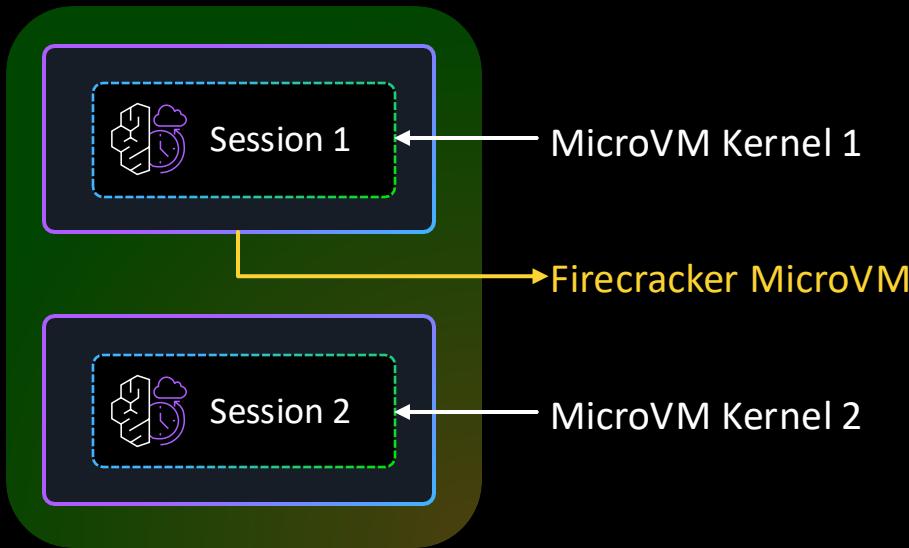
# Build your agent with the SDK
from bedrock_agentcore import BedrockAgentCoreApp

app = BedrockAgentCoreApp( )

@app.entrypoint
def my_agent(request):
    # Implement agent logic
    return response

app.run()
```

# AgentCore Runtime *True* Session Isolation



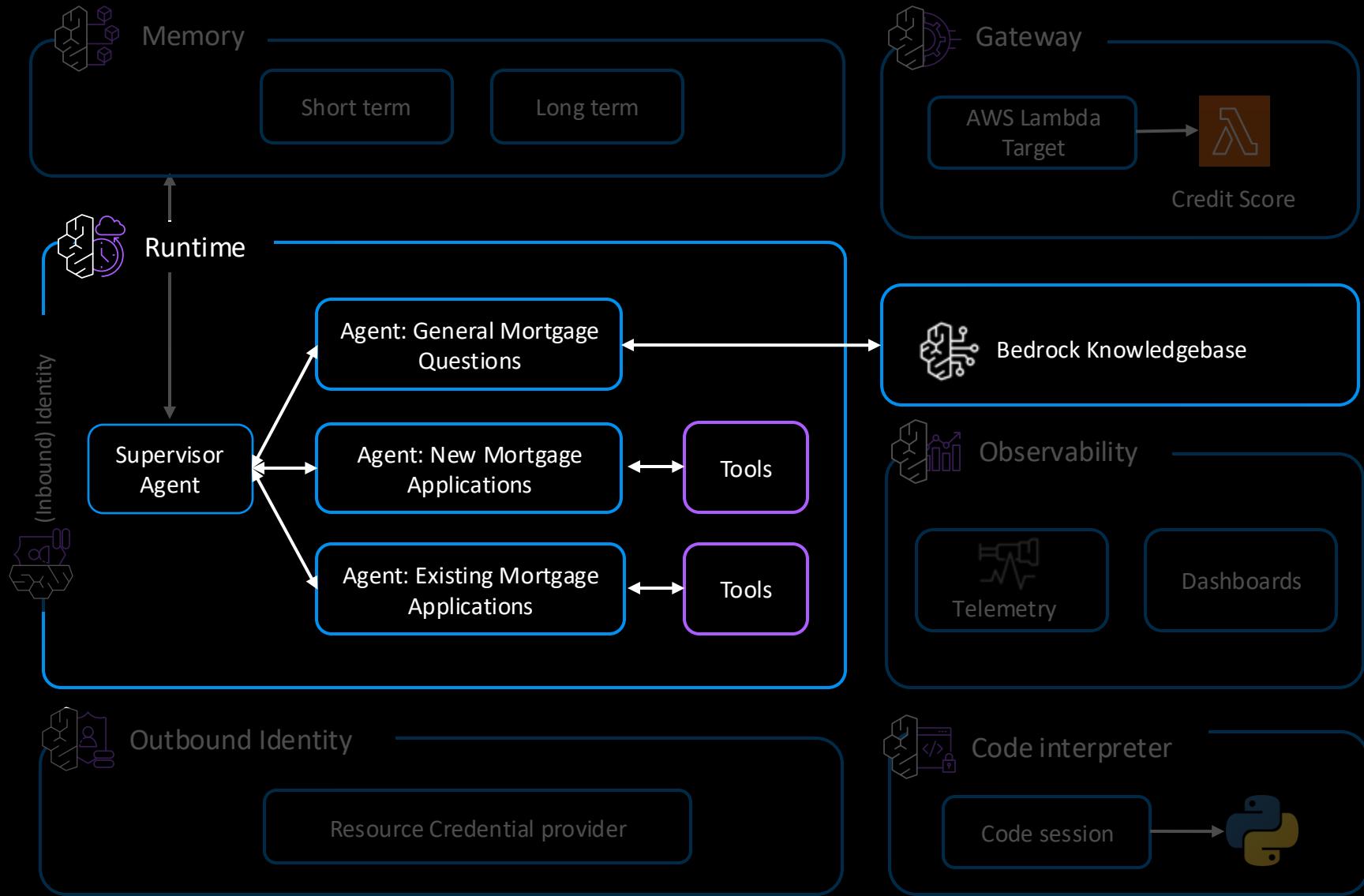
- Each session runs in a completely isolated microVM (compute + memory + filesystem)
- With other serverless offerings, multiple sessions *may* execute in the same microVM
  - ⚠ Without session isolation, local files and state could be accessed across sessions
- Stateful - preserves state securely within a session
- Use Agentcore Memory for out-of-session state, short and long term



# LAB #03-agentcore-runtime



# AgentCore Identity



Step 1 – Set up, KB

Step 2 – Strands Agent

Step 3 – Runtime

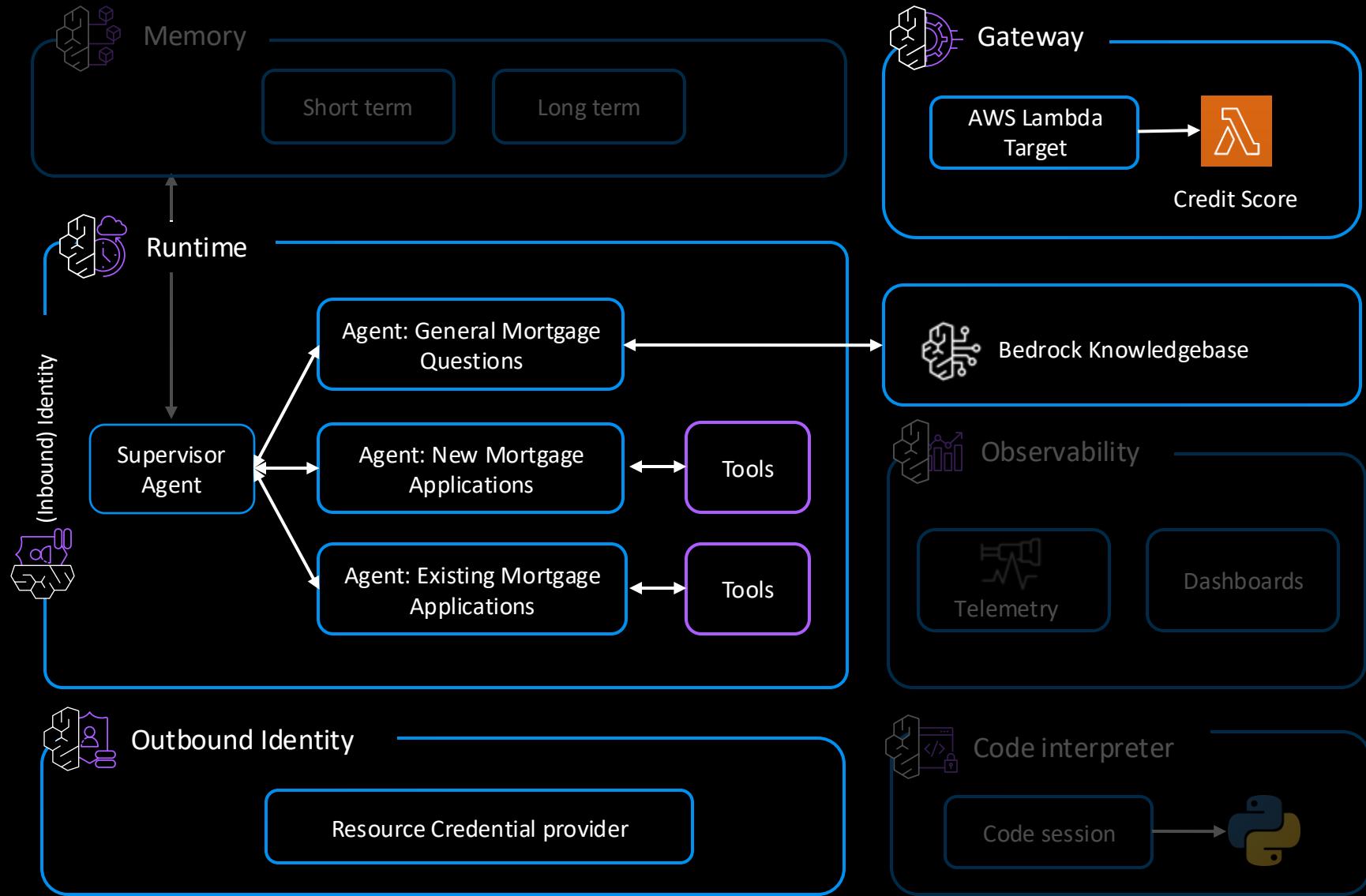
Step 4 – Gateway

Step 5 - Identity

Step 6 – Memory

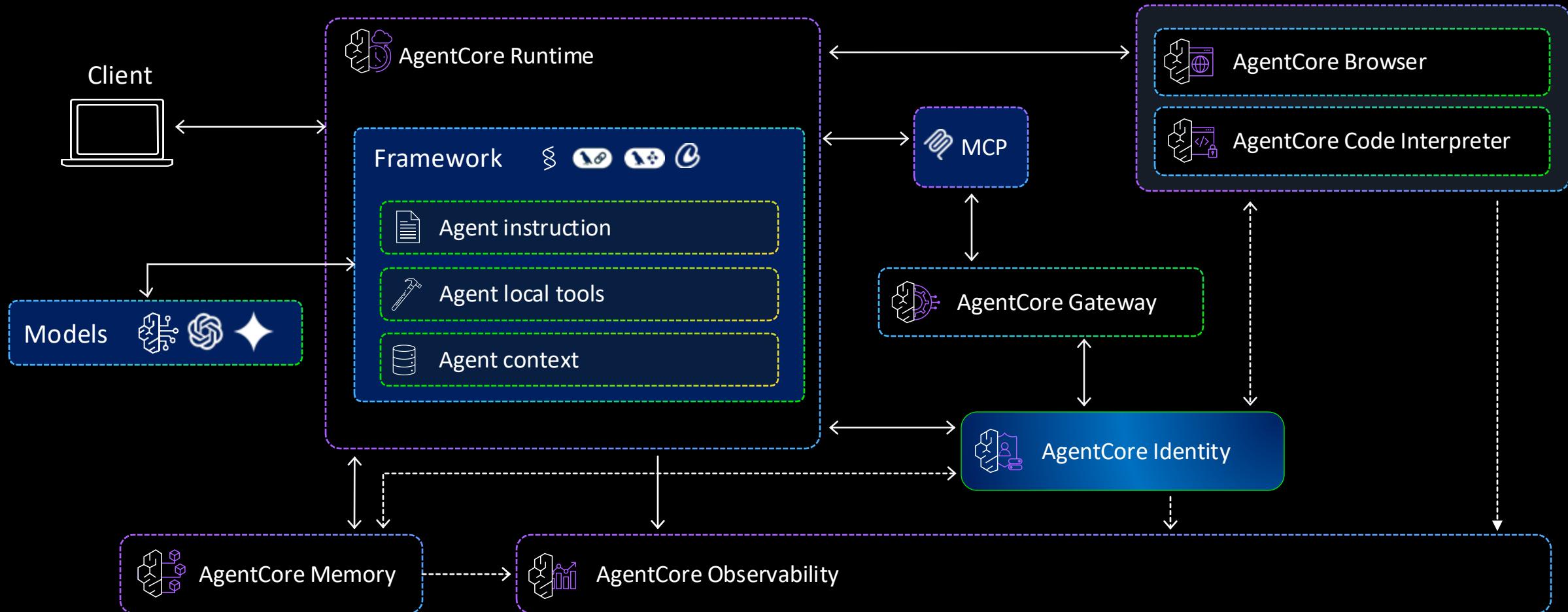
Step 7 – Observability

Step 8 – Tools



- Step 1 – Set up, KB
- Step 2 – Strands Agent
- Step 3 – Runtime
- Step 4 – Gateway
- Step 5 - Identity
- Step 6 – Memory
- Step 7 – Observability
- Step 8 – Tools

# Amazon Bedrock AgentCore



# A recap of concepts

Authentication

Proving the user/agent/entity is who they say they are

Authorisation

Verifying the user/agent/entity is allowed to perform an action

OAuth

An open standard for access delegation

JWT

JSON Web Token – contains information about identity and other attributes

IdP

Identity Provider – issues tokens

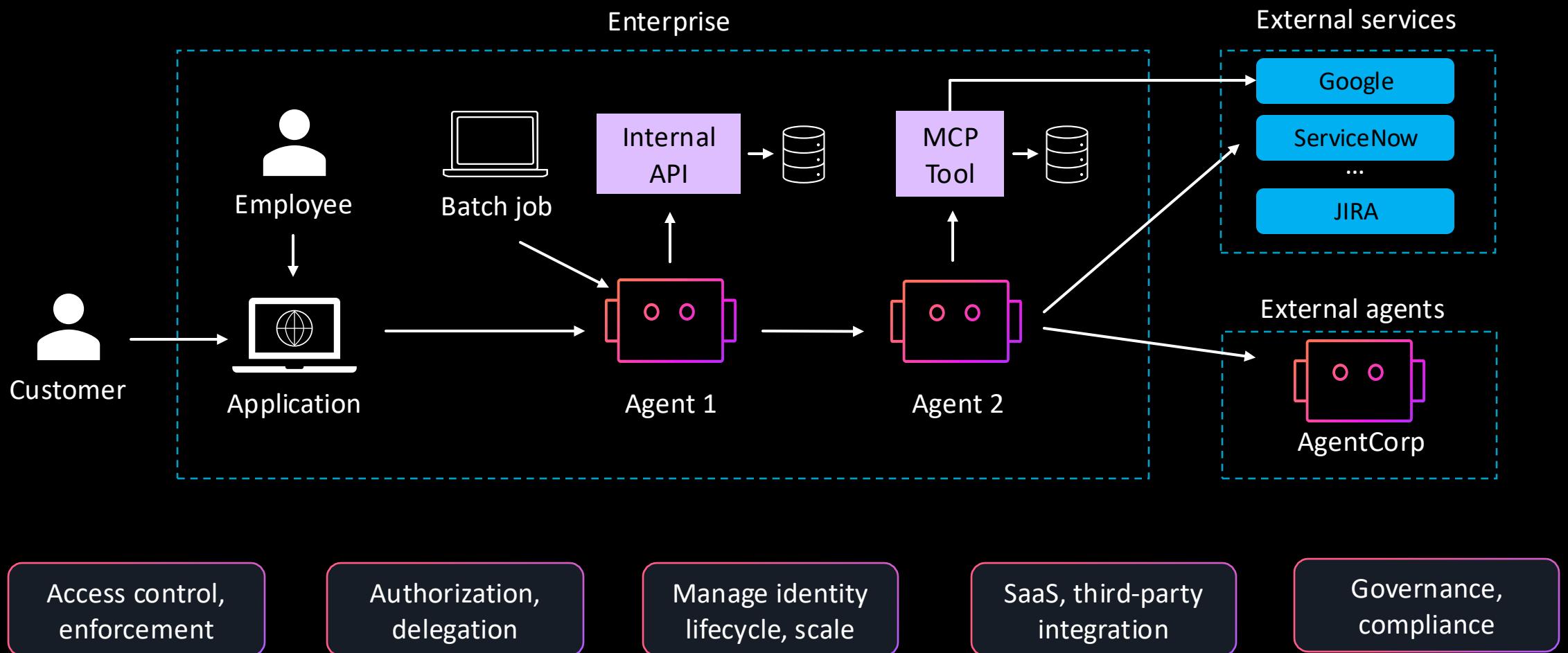
2LO

Client Credentials Flow - machine to machine auth.

3LO

Authorization Code Grant – used to auth with user context

# Agents Present Identity Challenges





# AgentCore Identity

Managed agent identity and access management

Secure, delegated access for AI agents



- Assigns distinct agent identities for secure agent operations at scale
- Enables AI agents to securely access AWS resources and third-party tools and services such as GitHub, Google, Salesforce and Slack

Build streamlined AI agent experiences



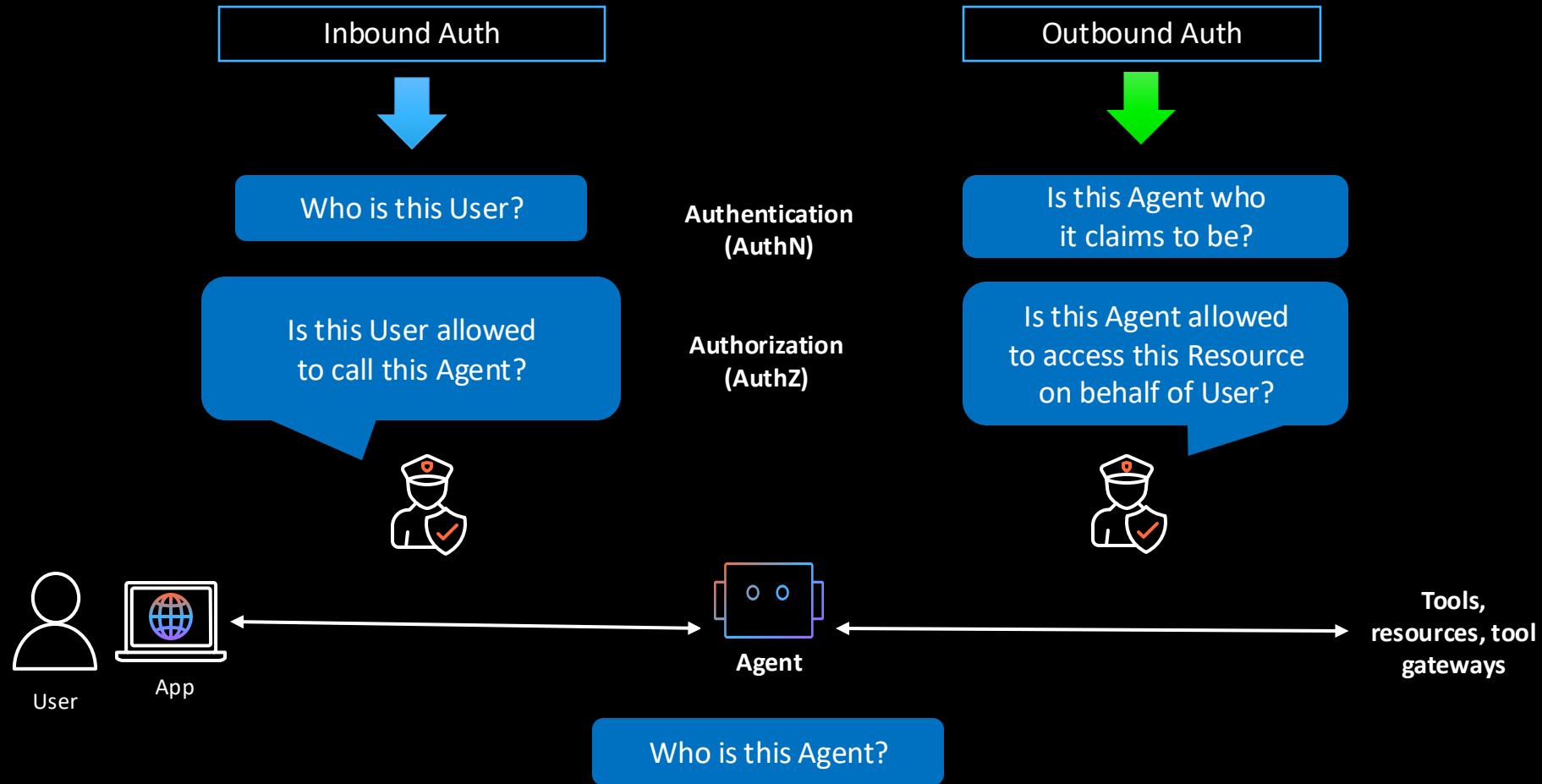
- Minimizes consent fatigue with a secure token vault
- Streamlines authentication flows and secure credential management
- Offers complete visibility through comprehensive audit trails

Accelerated AI agent development



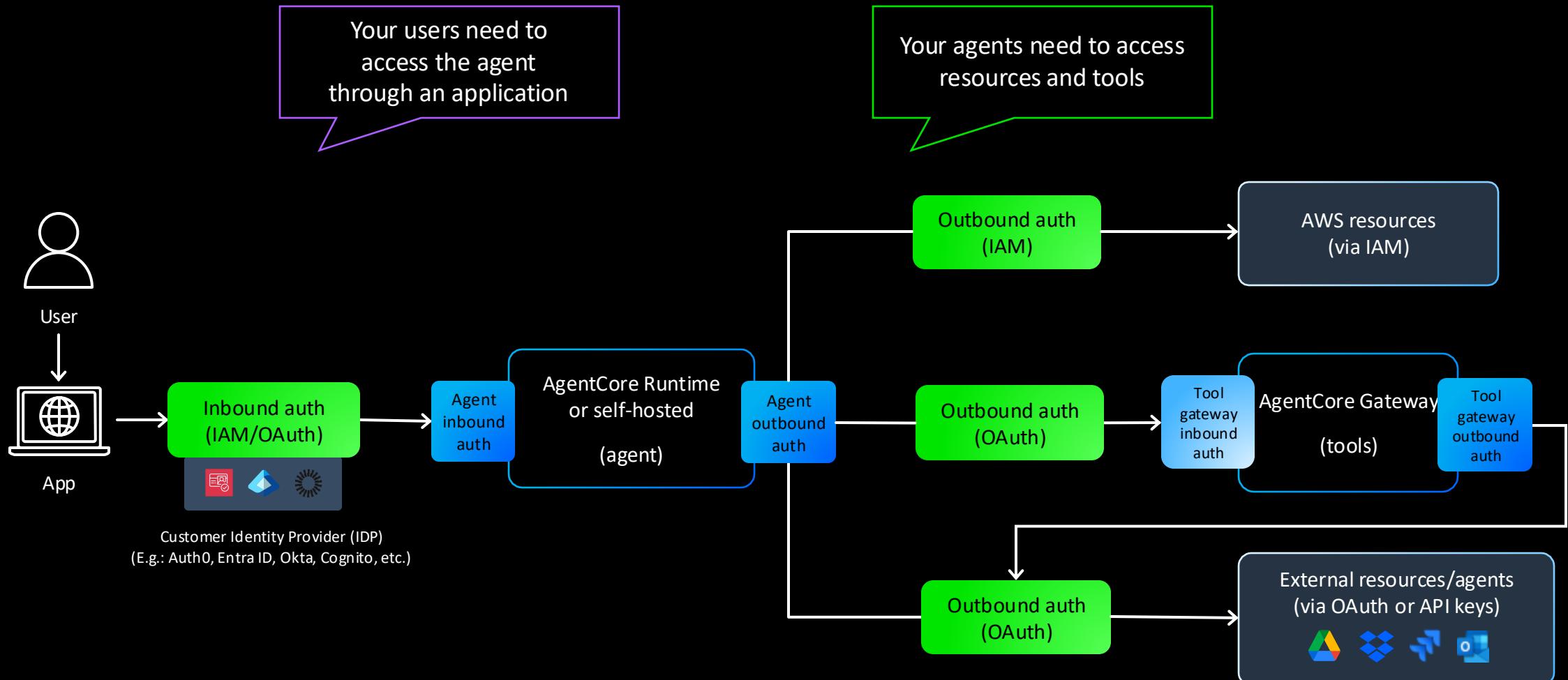
- Seamless integration with existing identity systems such as Okta, Azure Entra ID, or Amazon Cognito
- Lowers custom development efforts without need for migrating users or rebuilding authentication flows

# Agentic AI Auth Basics





# Auth with AgentCore Identity



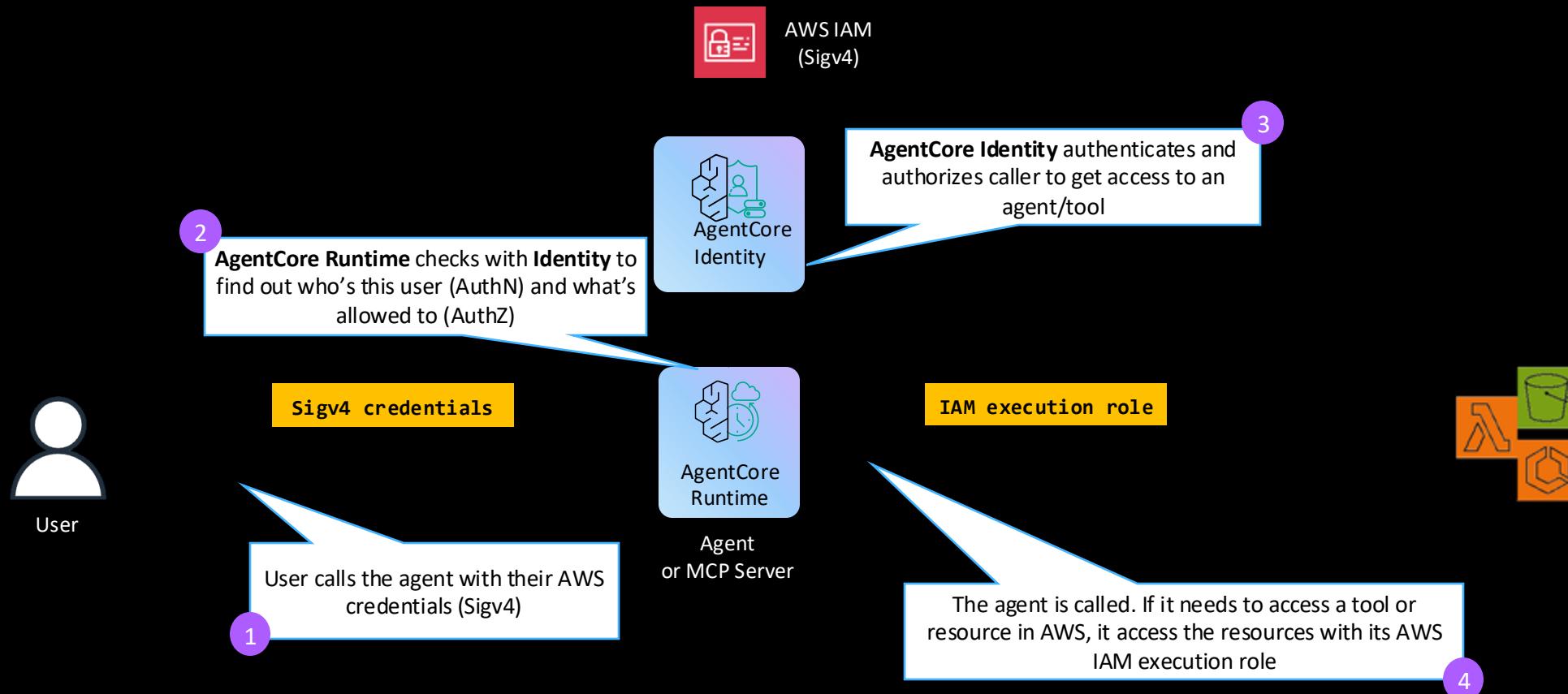
# How it works – AgentCore Identity Scenario 1



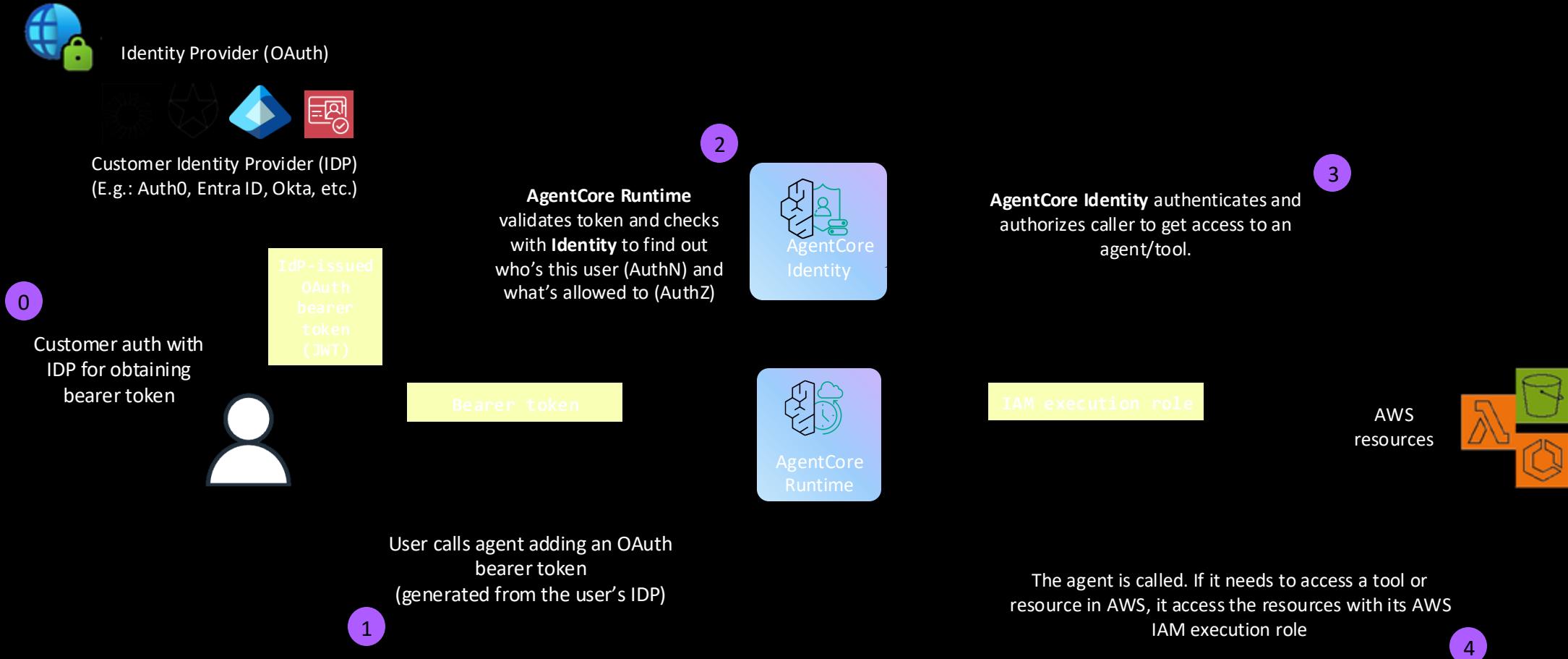
Agent/tool hosted in AgentCore Runtime accessing resources in AWS

\*Sigv4 = [AWS Signature Version 4](#)

- **Inbound auth method:** User identity authorization method AWS IAM (Sigv4 credentials)
- **Outbound auth method:** Resources access via AWS IAM role



# How it works – AgentCore Identity scenario 2





# AgentCore Gateway



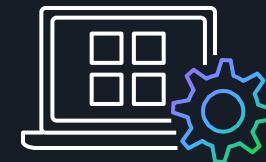
# AgentCore Gateway

## 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



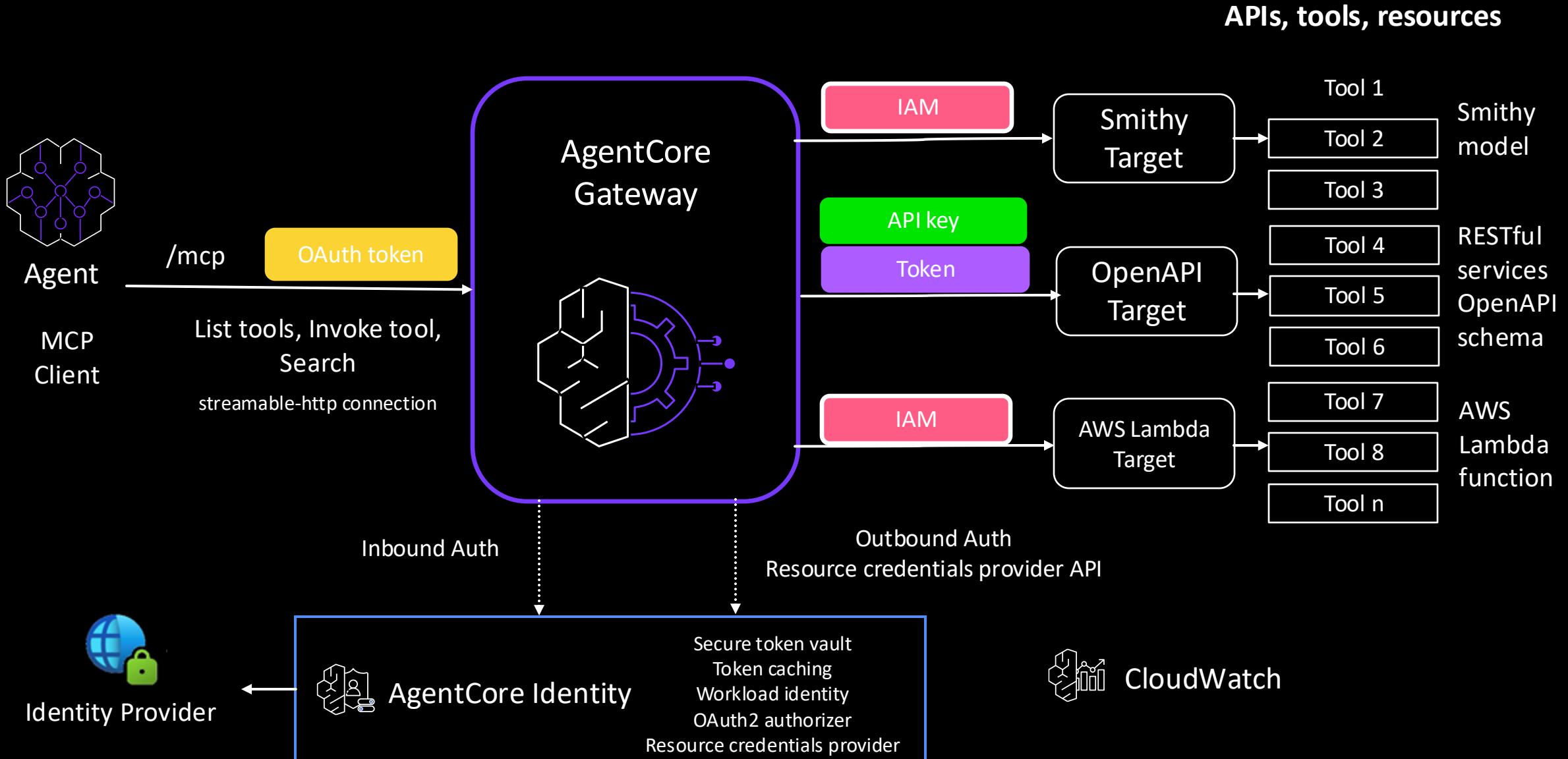
- Discover and use tools through a single, secure endpoint
- Combine multiple tools sources into one unified interface

## Intelligent tool discovery

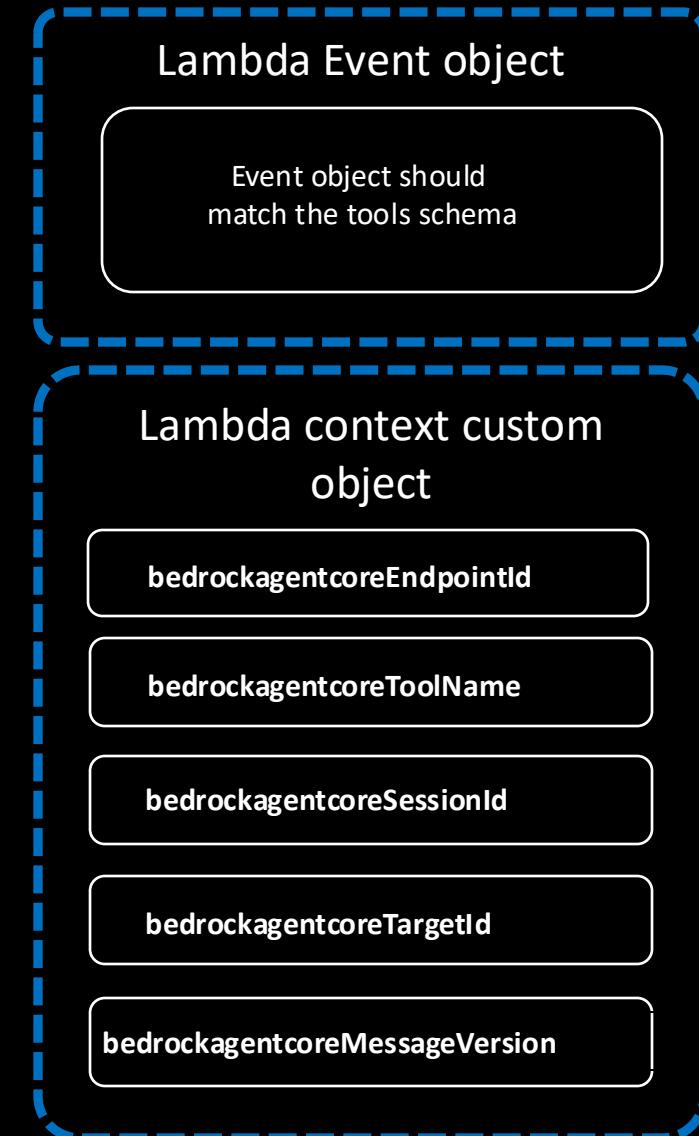
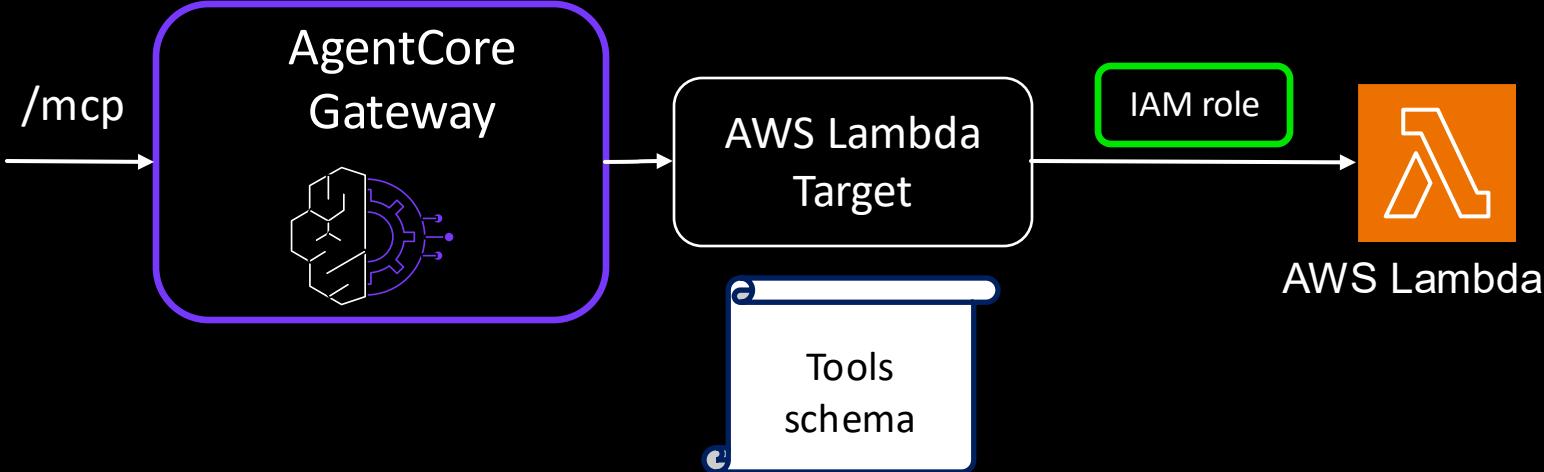


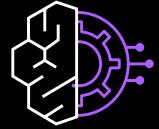
- Enable agents to find the right tools with context aware discovery
- Curated tool collections with granular permissions

# AgentCore Gateway provides secure access

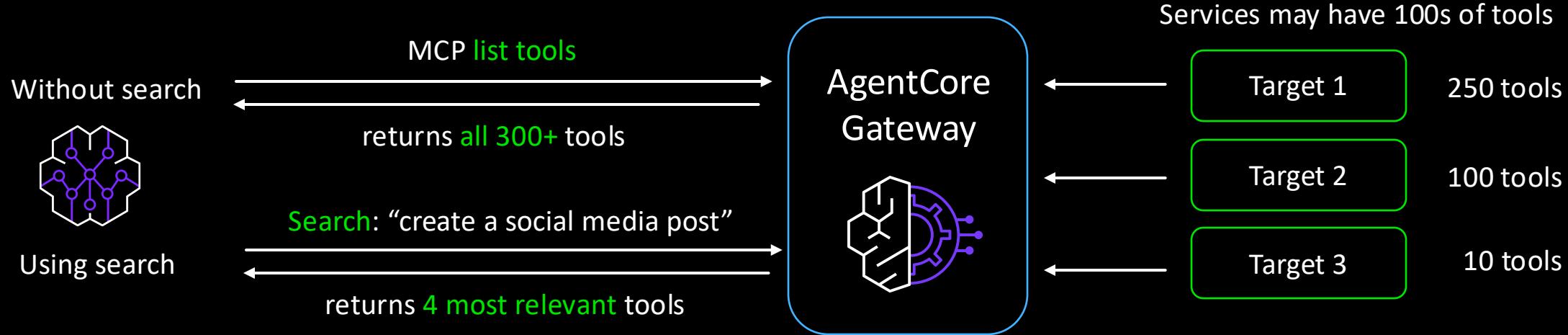


# AWS Lambda function tools for Bedrock AgentCore Gateway



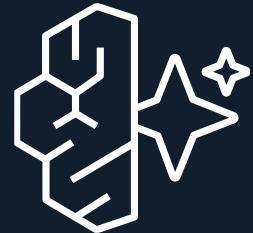


# AgentCore Gateway semantic search



## 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



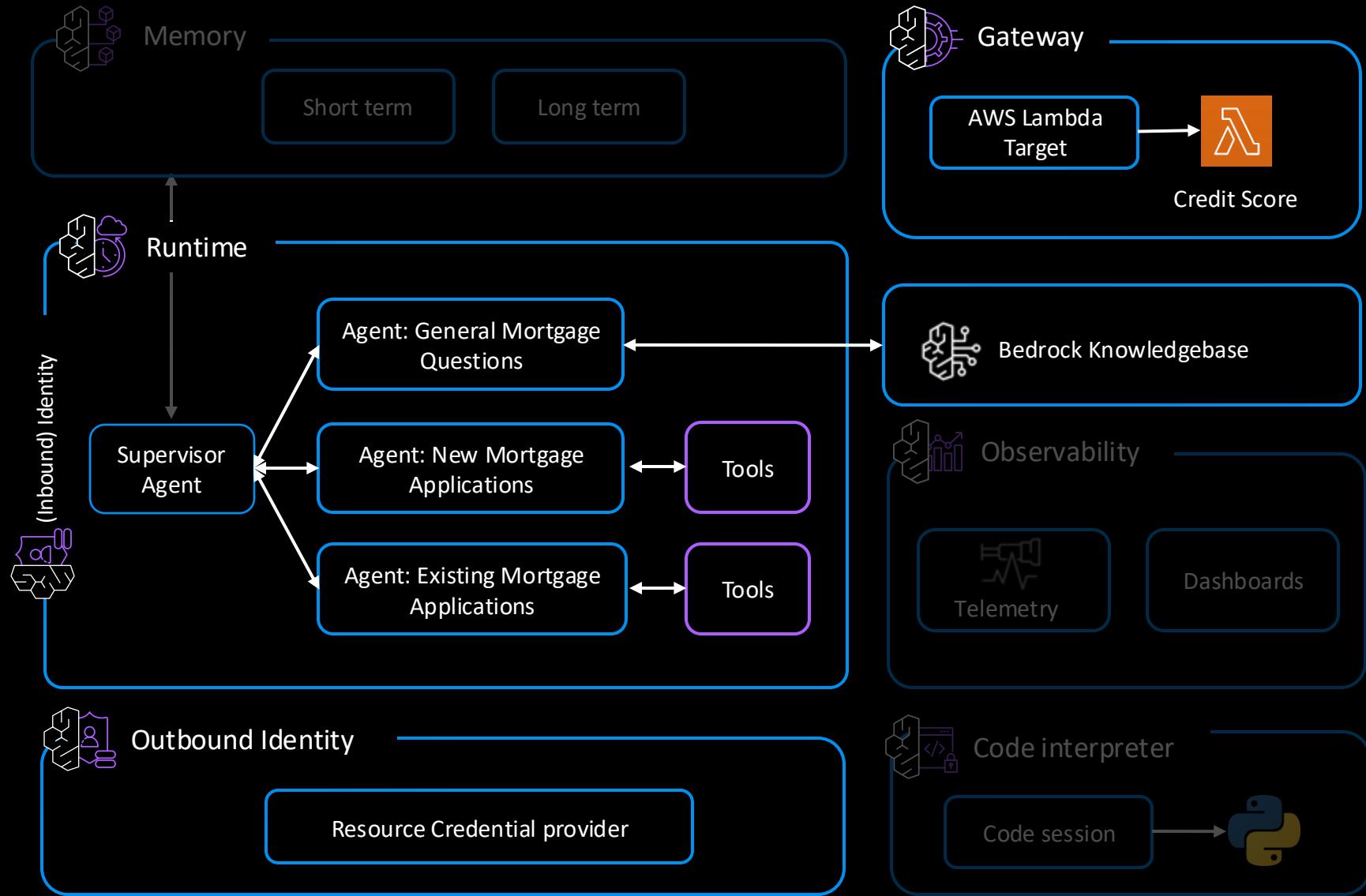
# LAB #04-agentcore-gateway



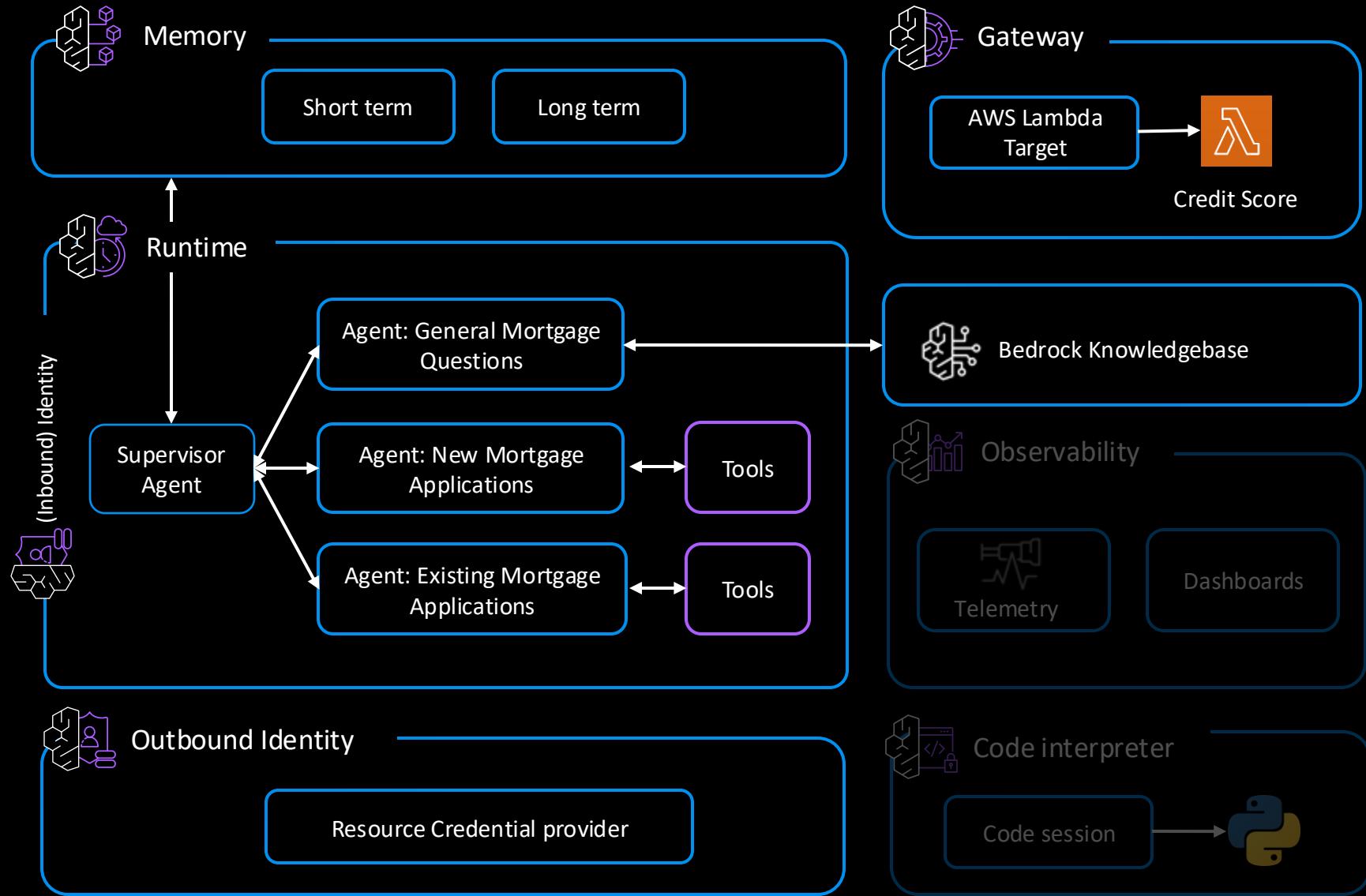
# LAB #05-agentcore-identity



# AgentCore Memory



- Step 1 – Set up, KB
- Step 2 – Strands Agent
- Step 3 – Runtime
- Step 4 – Gateway
- Step 5 - Identity
- Step 6 – Memory
- Step 7 – Observability
- Step 8 – Tools



Step 1 – Set up, KB

Step 2 – Strands Agent

Step 3 – Runtime

Step 4 – Gateway

Step 5 - Identity

Step 6 – Memory

Step 7 – Observability

Step 8 – Tools

# Agent experience without memory

User: "Hey, I'm trying to book a flight"

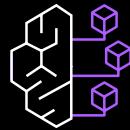
AI: "Sure I can help, what is your destination and dates?"

User: "From New York to London, 8 December"

AI: "The flight time from New York to London is 7 hours. December is a wonderful time to visit!"

User: "No, I want to *book* the flight, not get facts about it."

AI: "Sure I can help, what is your destination and dates?"



# 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 - key concepts

- **Memory**

A logical data store of memories

- **Short-term memory**

A temporary storage for memories (create, list, get)

- **Long-term memory**

Strategically stored memories that can be semantically queried

- **Strategy**

Instructions on what to move from short to long-term memory

# Logical isolation of memory data: short-term memory

**memory\_id** – the memory resource in your AWS account

**actor\_id** - entities in your system (users, agents, project, or combinations),

**session\_id** – a session of related events

# Logical isolation of memory data: long-term memory

## Namespaces

A user-defined path where memories are organised and stored **per strategy**.

Use built-in variables for dynamic namespaces:

- {actorId}
- {sessionId}
- {memoryStrategyId}

Examples:

Store everything under actorId → “/{actorId}”

Store records per actor and session → “/{actorId}/{sessionId}”

Store records food preferences per actor → “/{actorId}/food\_preferences”

# Logical isolation of memory data: long-term memory

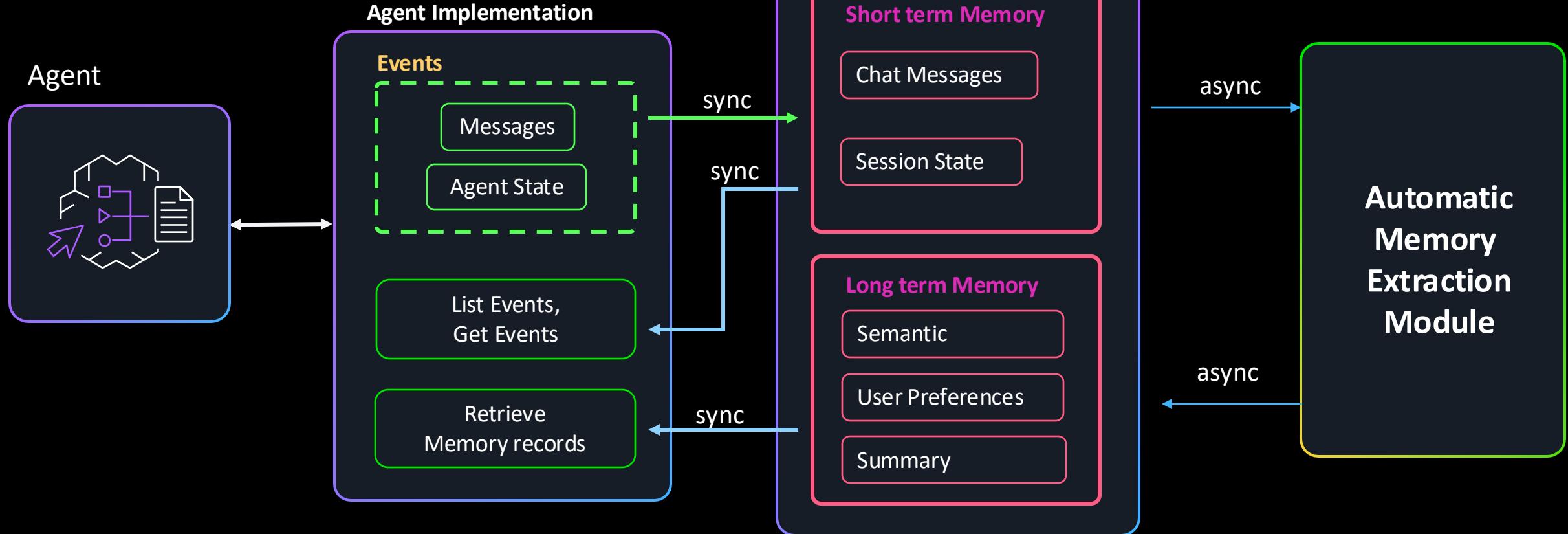
## Example: retrieve food preferences for actor (user)

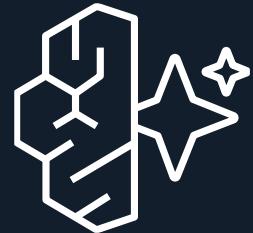
```
client = boto3.client("bedrock-agentcore")

client.retrieve_memory_records(
    memoryId="abcd1234",
    namespace="/anthonyk/food_preferences",
    searchCriteria={
        "searchQuery": "food preferences for anthony",
        "memoryStrategyId": "food-pref-strat",
    },
)
```

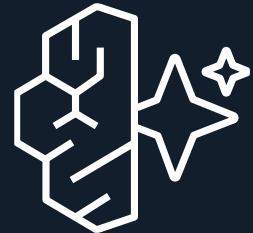


# AgentCore Memory

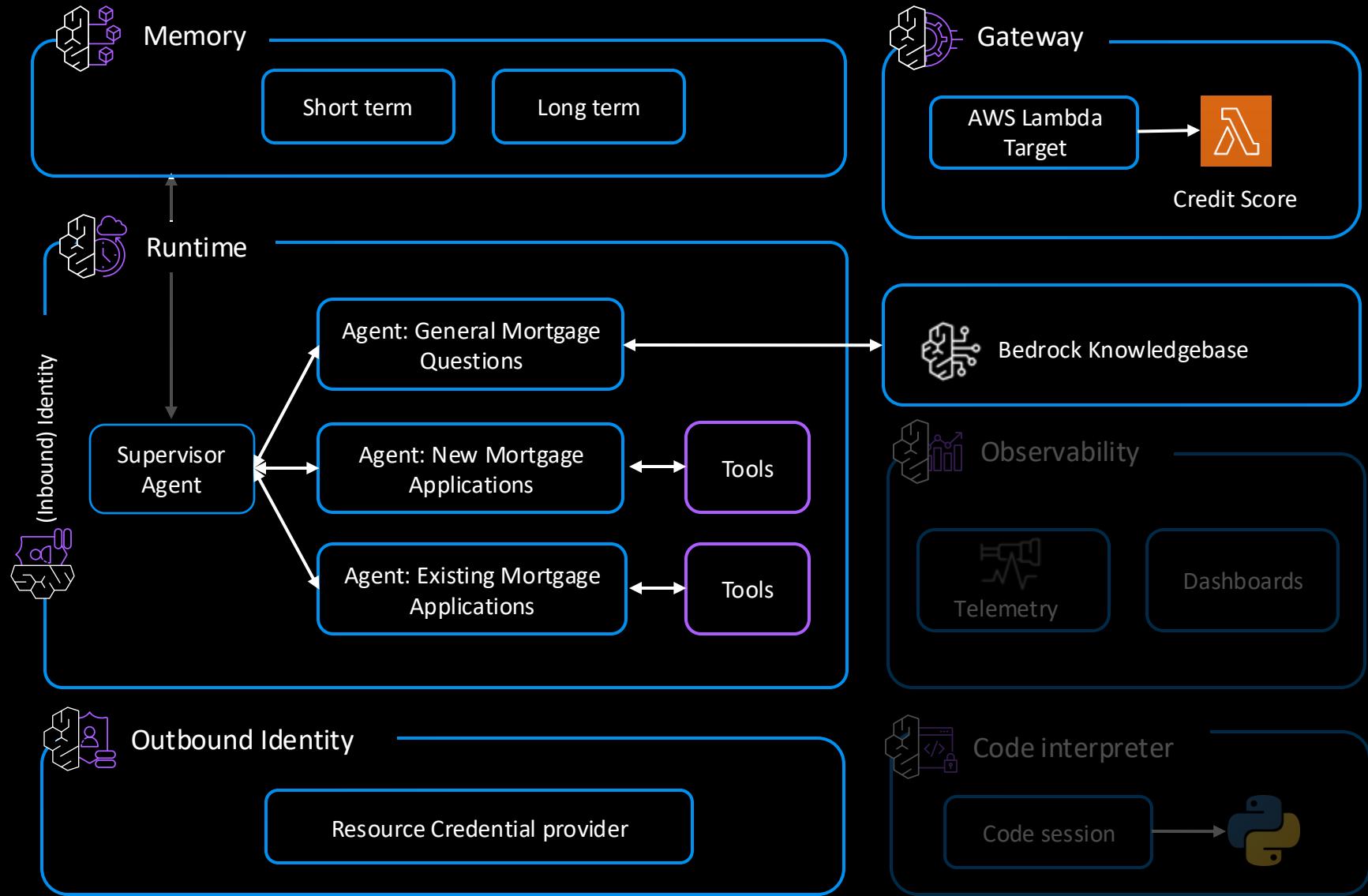




# LAB #06-agentcore-memory



# AgentCore Observability



Step 1 – Set up, KB

Step 2 – Strands Agent

Step 3 – Runtime

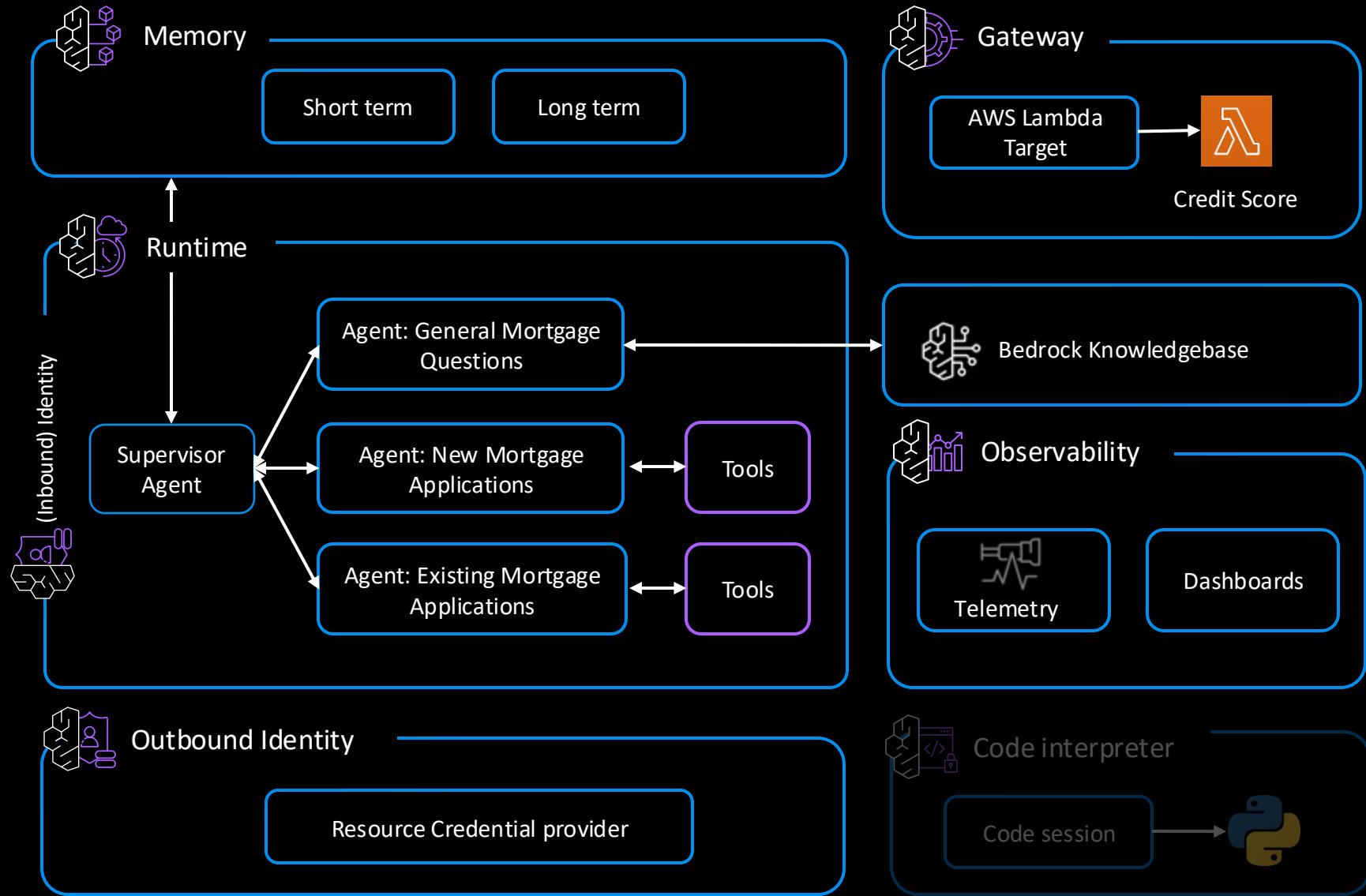
Step 4 – Gateway

Step 5 - Identity

Step 6 – Memory

Step 7 – Observability

Step 8 – Tools



- Step 1 – Set up, KB
- Step 2 – Strands Agent
- Step 3 – Runtime
- Step 4 – Gateway
- Step 5 - Identity
- Step 6 – Memory
- Step 7 – Observability
- Step 8 – Tools



# 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

Accelerate  
time to market



- Dashboards powered by CloudWatch save developers time
- Single-pane-of-glass view into agents' operational health
- Eliminate the need to manually integrate data from multiple sources

Integrate with 3P observability  
tools

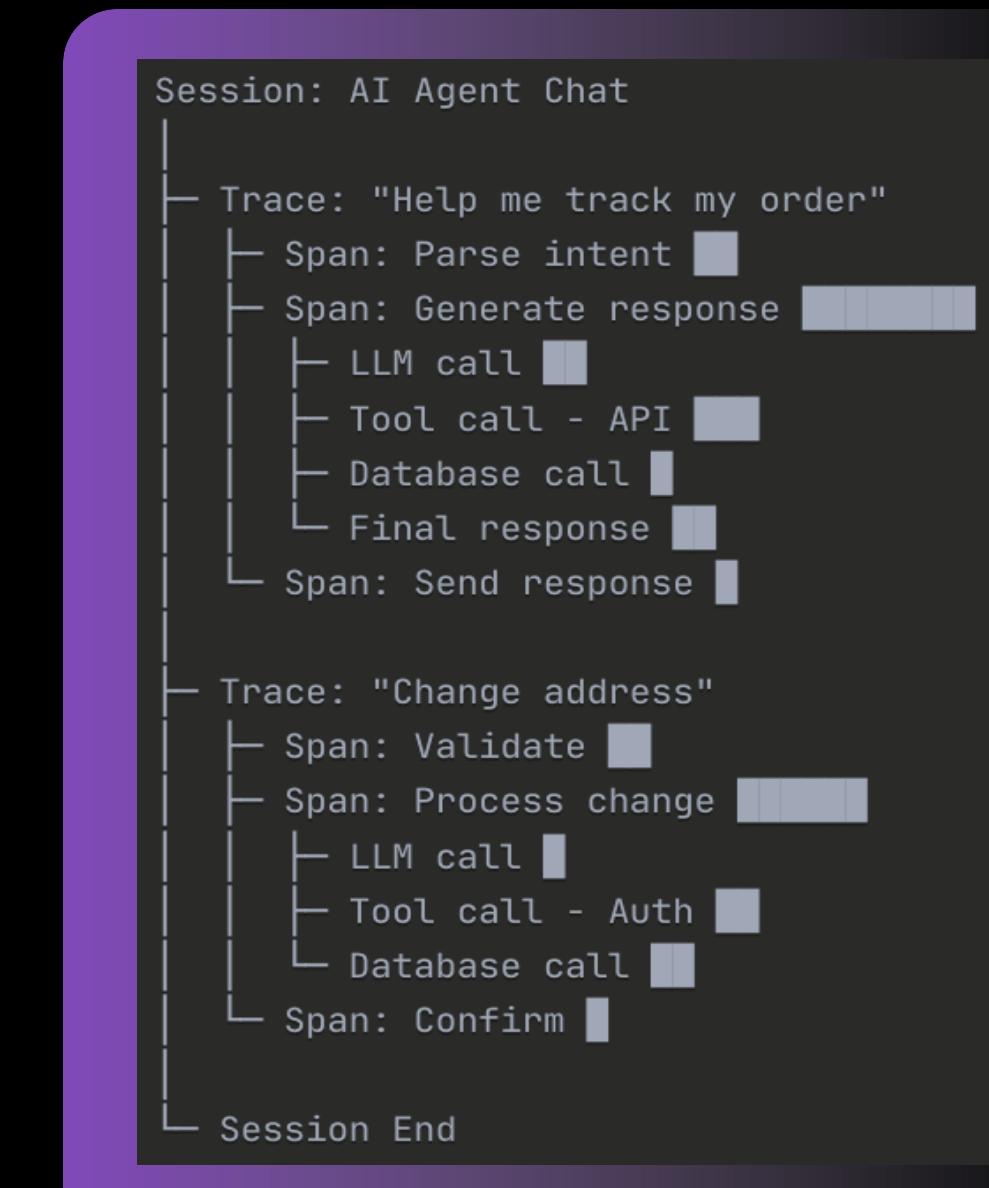


- Integration with a wide range of monitoring and observability tools, including CloudWatch
- Flexibility to leverage your existing observability stack



# Key Concepts

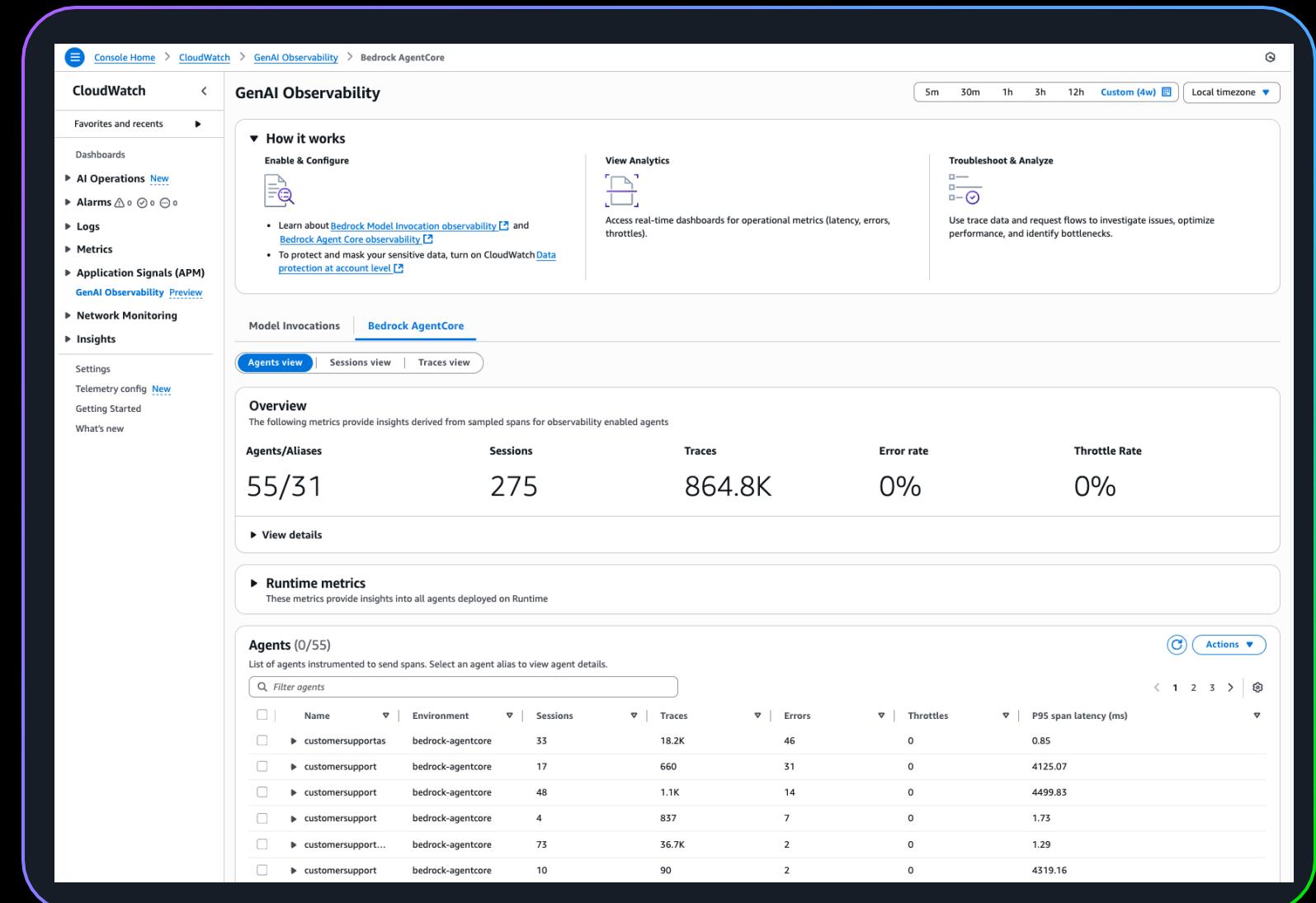
- Session
- Trace
- Span
- Sub-Span





# How it Works

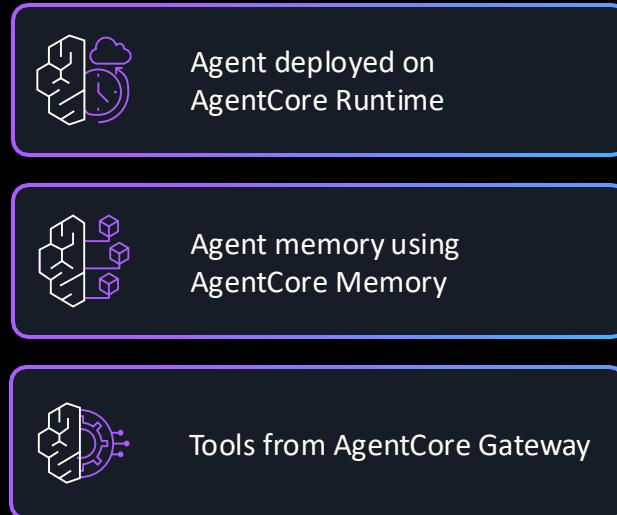
Visibility to operate agents you can trust



The screenshot shows the AWS CloudWatch GenAI Observability console. The left sidebar includes links for AI Operations, Alarms, Logs, Metrics, Application Signals (APM), GenAI Observability (Preview), Network Monitoring, and Insights. The main content area has a heading "GenAI Observability" and a "How it Works" section with "Enable & Configure" and "View Analytics" options. Below this is the "Model Invocations" section, which is currently selected. It displays metrics like Agents/Aliases (55/31), Sessions (275), Traces (864.8K), Error rate (0%), and Throttle Rate (0%). The "Agents view" tab is active, showing a table of agents with columns: Name, Environment, Sessions, Traces, Errors, Throttles, and P95 span latency (ms). The table lists several entries, including "customersupportas" and "customersupport".

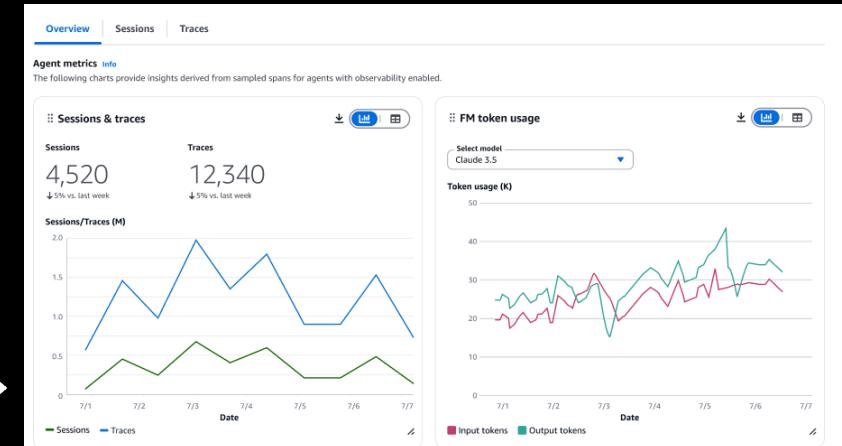


# AgentCore Observability

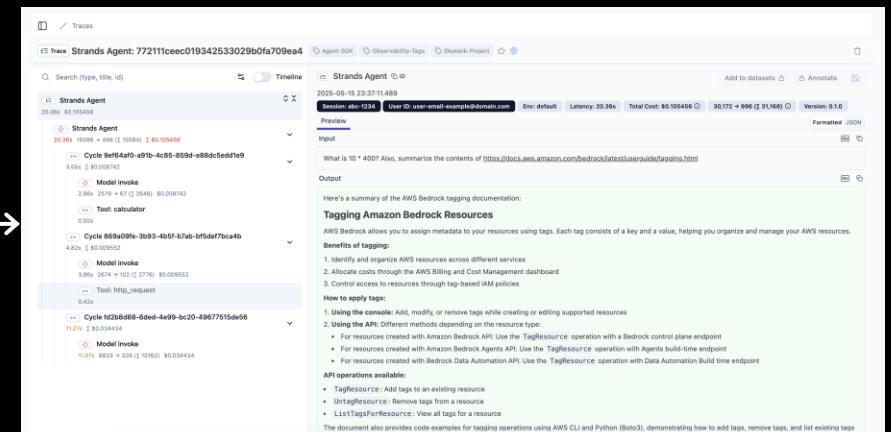


OTEL logs

AgentCore Observability dashboards



Third-party observability dashboards

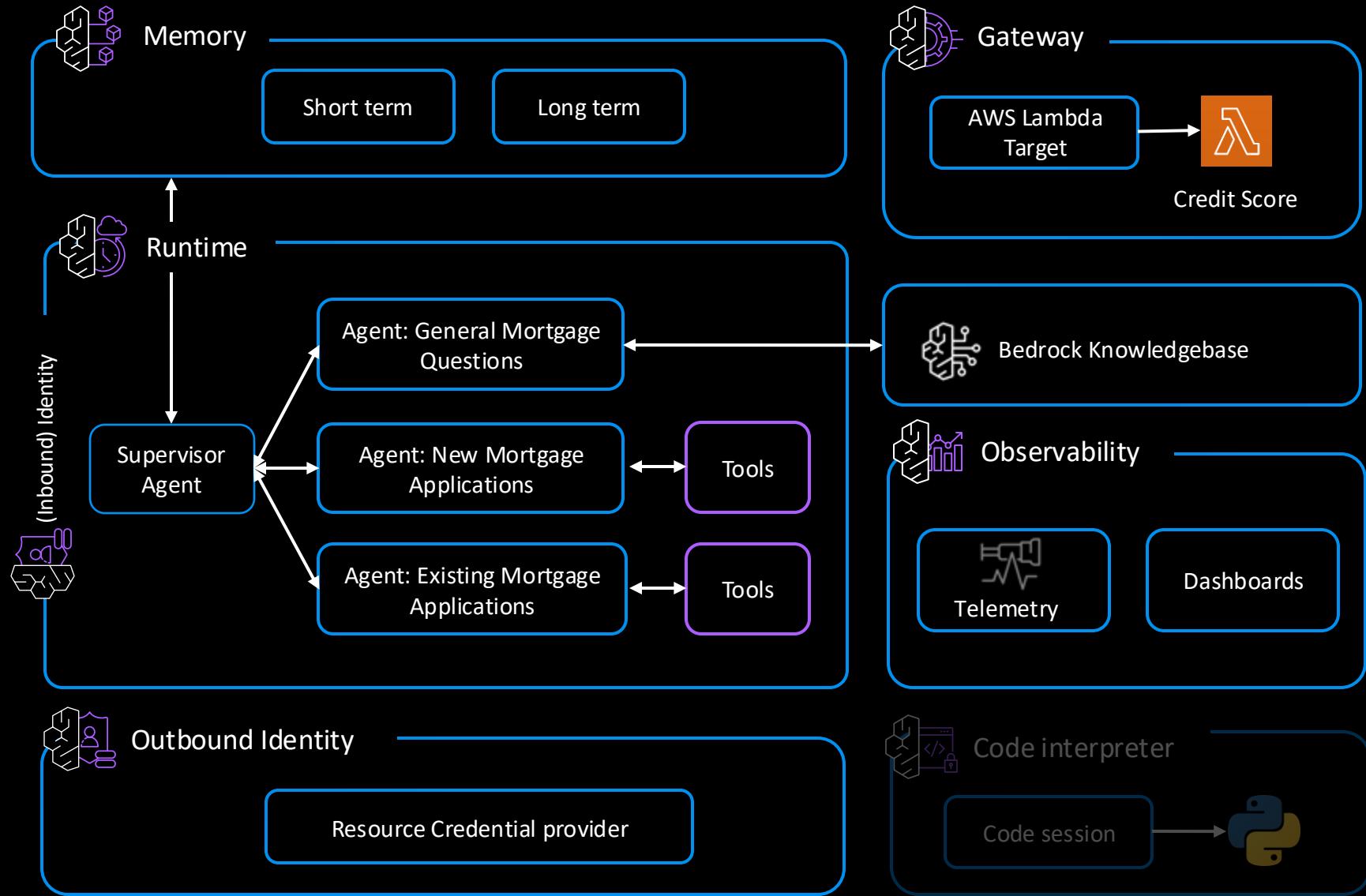




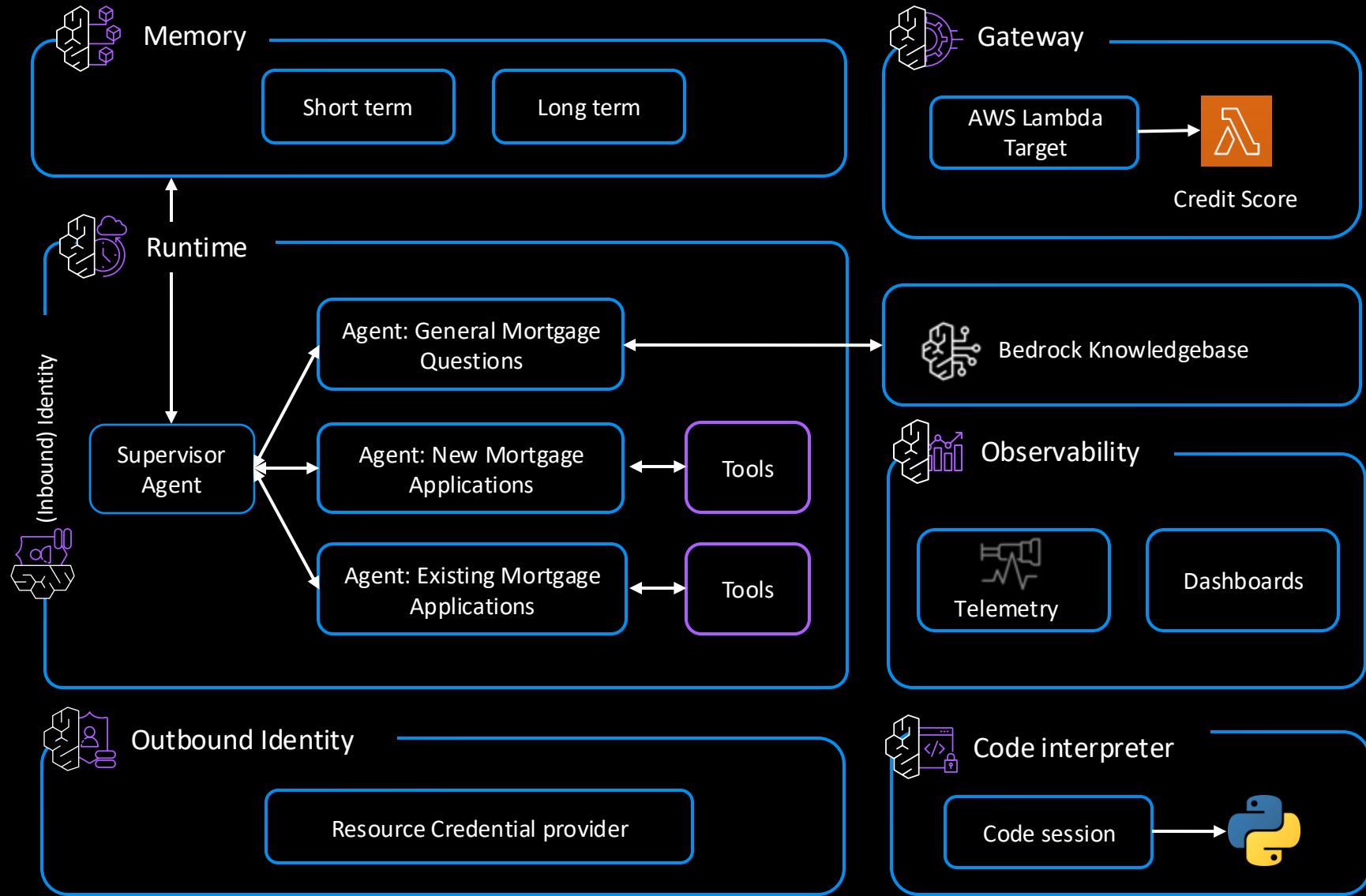
# LAB #07-agentcore-observability



# AgentCore Built-in Tools



- Step 1 – Set up, KB
- Step 2 – Strands Agent
- Step 3 – Runtime
- Step 4 – Gateway
- Step 5 - Identity
- Step 6 – Memory
- Step 7 – Observability
- Step 8 – Tools



- Step 1 – Set up, KB
- Step 2 – Strands Agent
- Step 3 – Runtime
- Step 4 – Gateway
- Step 5 - Identity
- Step 6 – Memory
- Step 7 – Observability
- Step 8 – Tools



# 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

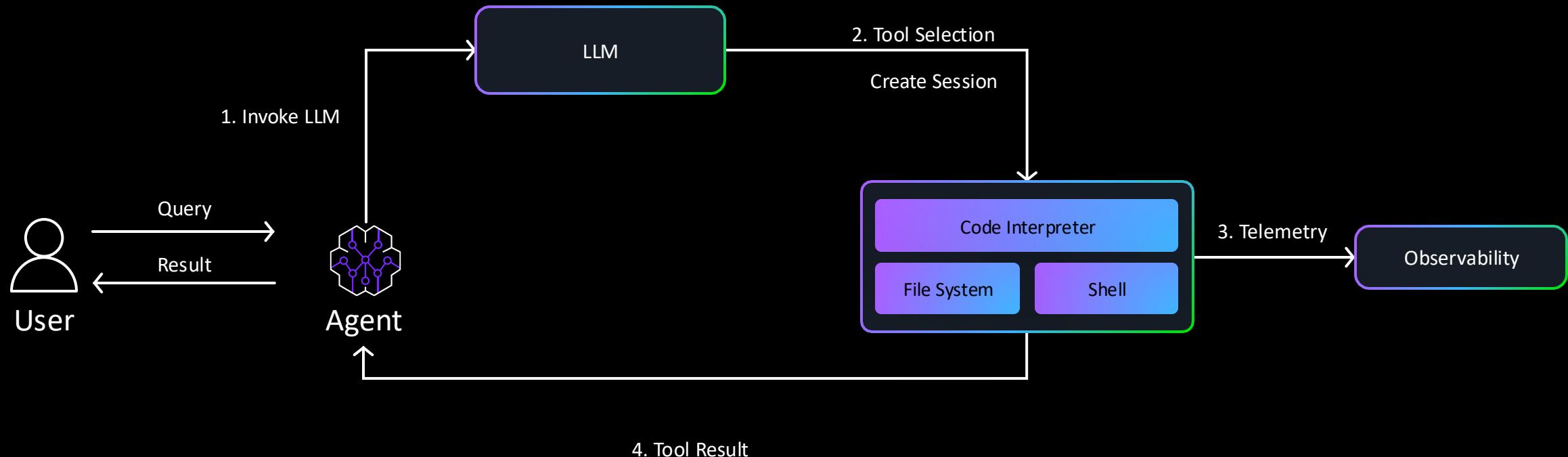
## Ease of use



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



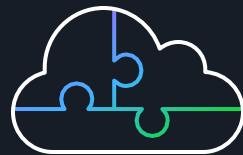
# AgentCore Code Interpreter





# AgentCore 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

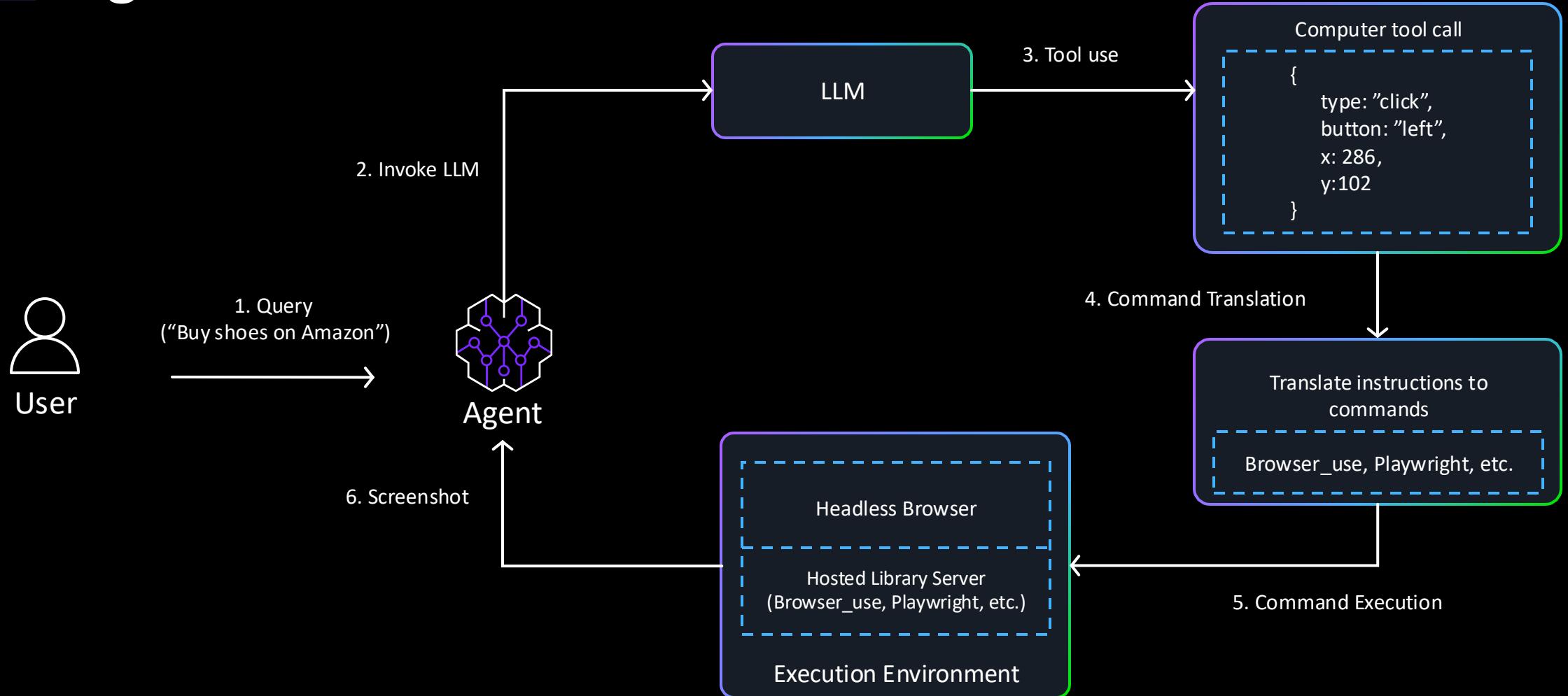
## Enterprise observability

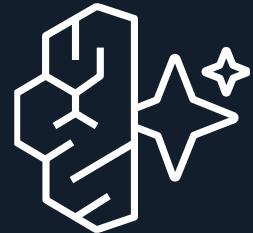


- Live streaming URLs for real-time monitoring
- Session replays for debugging
- Extensive logging of all browser commands to CloudTrail



# AgentCore Browser





# LAB #08-agentcore-tools



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

# Survey



<https://pulse.amazon/survey/BEJSYO2M>



Create your AWS  
Builder ID and claim  
your community alias  
and get access to...

Sign up today at: [bit.ly/anzbuilderid](https://bit.ly/anzbuilderid)



Amazon Q

Access to the most capable generative  
AI-powered assistant

600+

AWS Skill Builder Courses

Unlock a world of learning  
to build in-demand skills



AWS re:Post

A curated knowledge center  
and a vibrant community



Join AWS User Groups in New  
Zealand

Attend In person regular meetups  
organized by the AWS User Groups  
enabling peer to peer learning

# Join AWS User Groups

User groups are peer-to-peer communities which meet regularly to share ideas, answer questions, and learn about new services and best practices.

**17** AWS User Groups in Australia and New Zealand

**450+** AWS User Groups worldwide

Monthly/ Quarterly meetups

Find a User Group near you

[builder.aws.com/connect/community/user-groups](https://builder.aws.com/connect/community/user-groups)



Scan to join an  
AWS User Group  
In New Zealand

