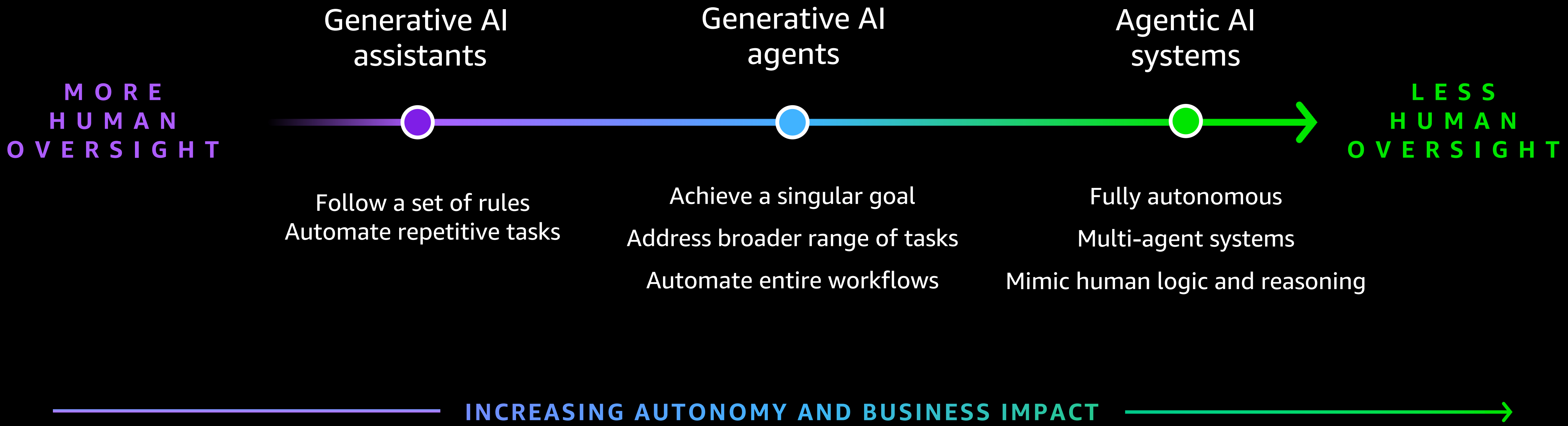
Amazon Bedrock AgentCore

Dr Anil Giri

Solution Architect ISV

The evolution into Agentic AI



MORE HUMAN OVERSIGHT

Generative AI assistants

Follow a set of rules
Automate repetitive tasks

Generative AI agents

Achieve a singular goal
Address broader range of tasks
Automate entire workflows

Agentic AI systems

Fully autonomous
Multi-agent systems
Mimic human logic and reasoning

LESS HUMAN OVERSIGHT

INCREASING AUTONOMY AND BUSINESS IMPACT

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

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

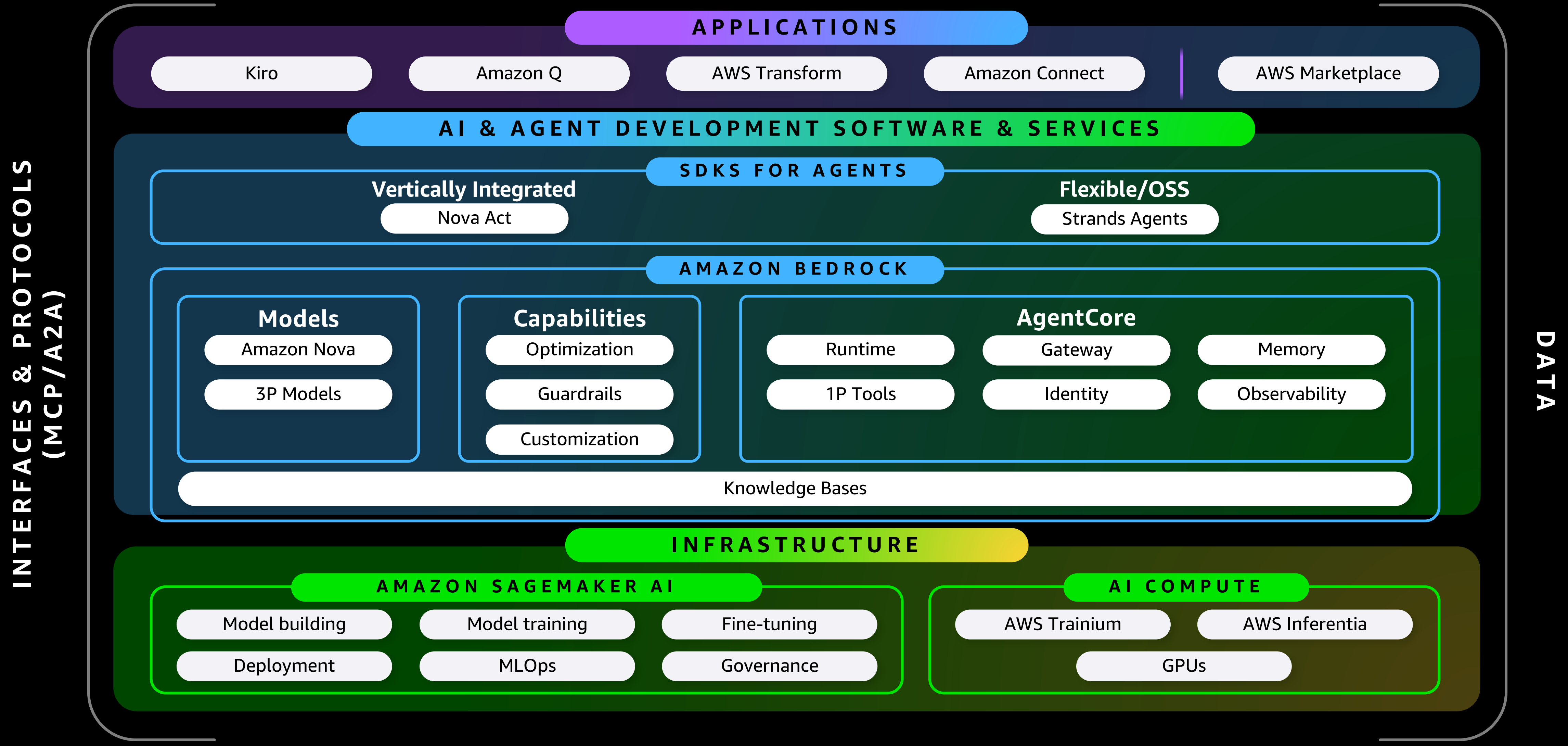
Tools for Building AI Agents

- Marketplace
- Bedrock Agents
- Strands
- Nova Act SDK

Agentic AI Foundation

Amazon Bedrock AgentCore

AWS AI Stack



INTERFACES & PROTOCOLS (MCP/A2A)

DATA

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

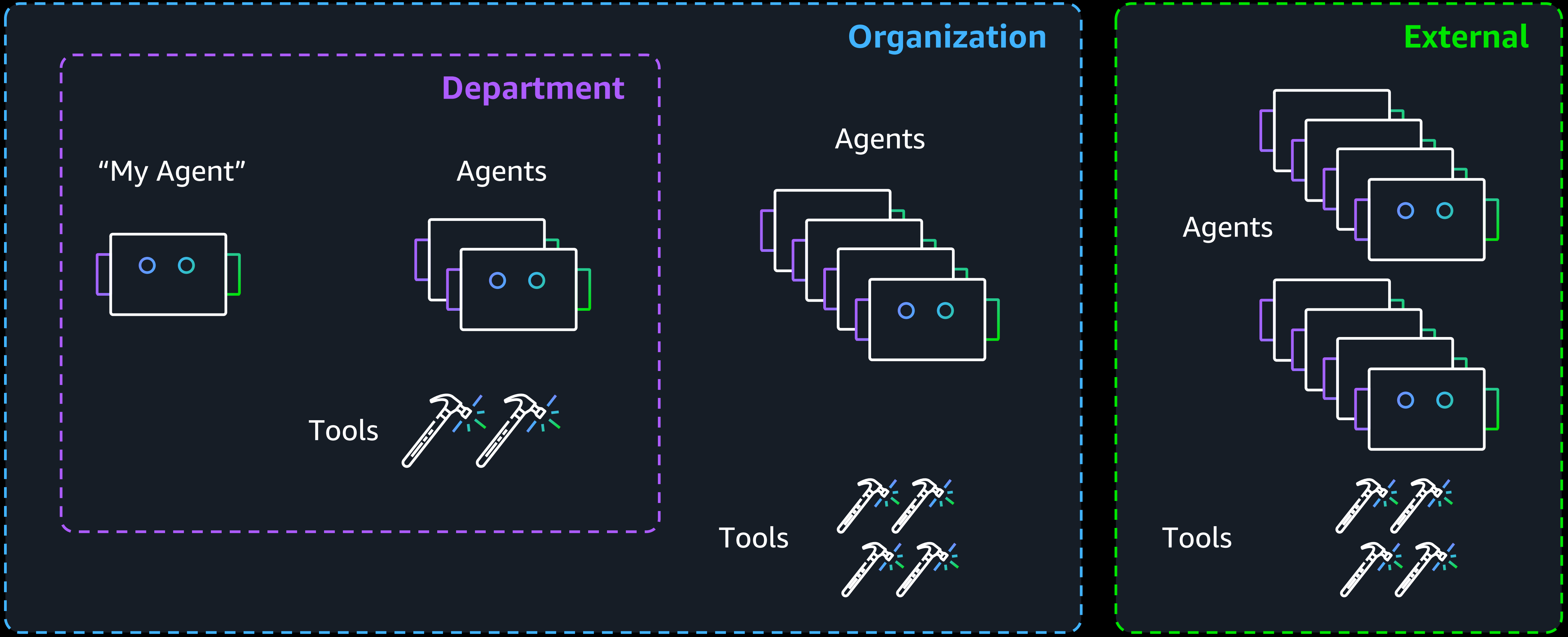
AI COMPUTE

AWS Trainium | AWS Inferentia

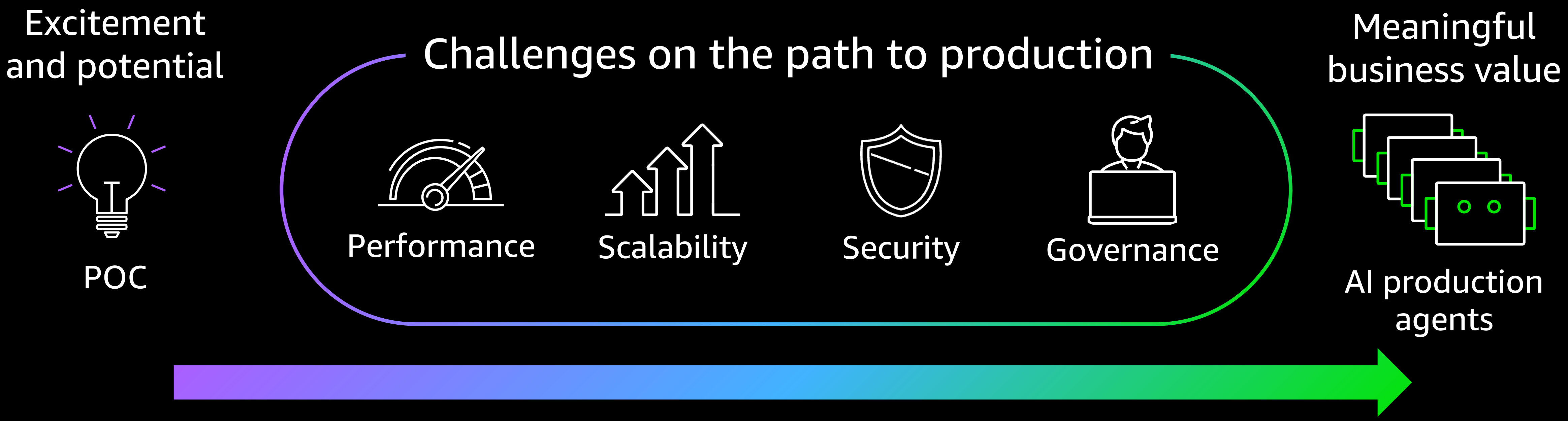
GPUs

Amazon Bedrock AgentCore

Building an agent is the start, production scale is the goal



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

Deploy and operate highly capable agents securely, at scale using any framework and model

TIME TO VALUE

Build powerful AI agents without the infrastructure and operational headaches

FLEXIBLE

Create agents with any framework or model

TRUSTED

Deploy secure, scalable, and reliable agents your organization can trust

Foundational services for running highly capable agents, securely at scale

Deploy securely
at scale

Enhance with tools
and memory

Monitor

Secure, scalable runtime for agents

Deploy securely at scale

AgentCore Runtime

AgentCore Identity

Enhance with tools and memory

Monitor

Essential tools and capabilities to build highly effective agents

Deploy securely at scale

Enhance with tools and memory

- AgentCore Gateway
- AgentCore Memory
- AgentCore Browser
- AgentCore Code Interpreter

Monitor

Visibility to operate agents you can trust

Deploy securely at scale

Enhance with tools and memory

Monitor

AgentCore Observability

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

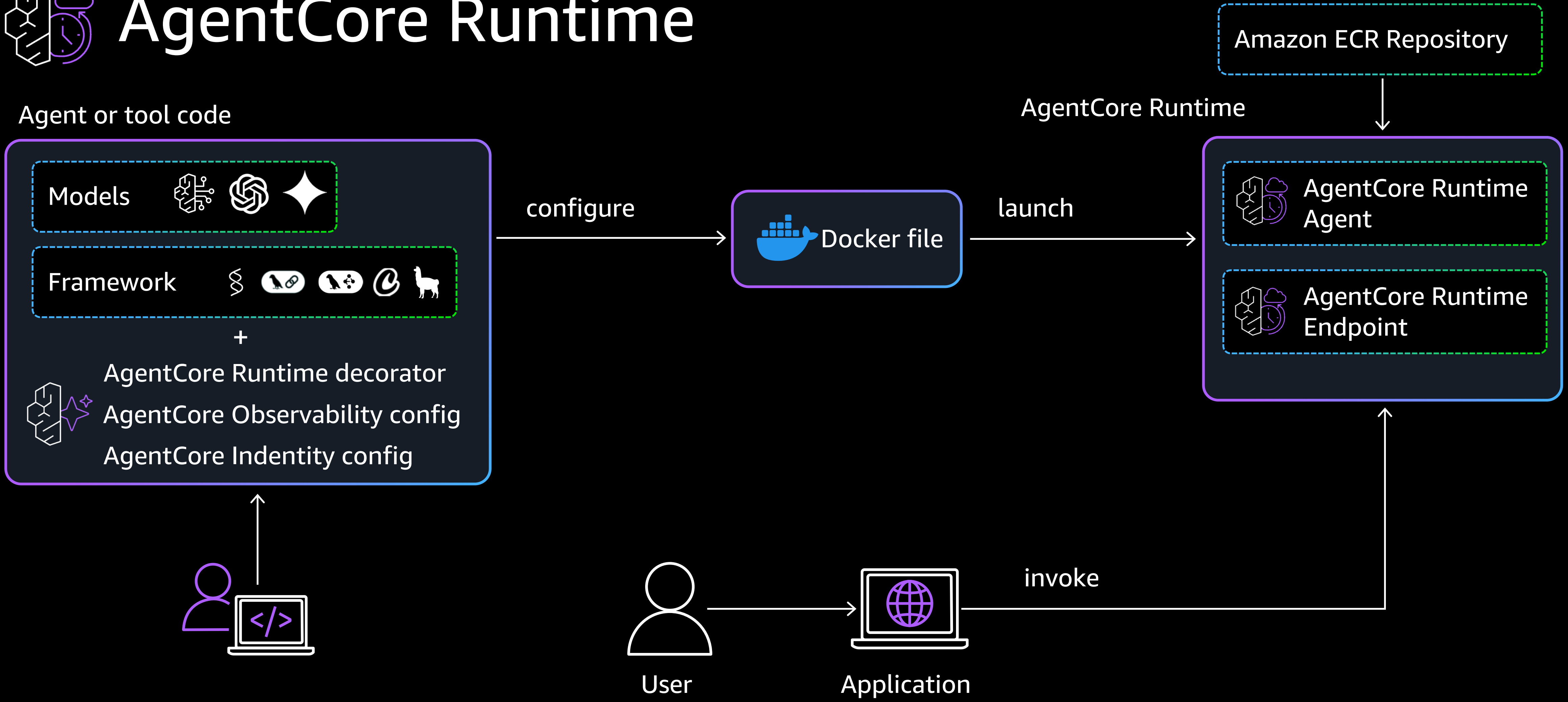
Accelerate time to market

- 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



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

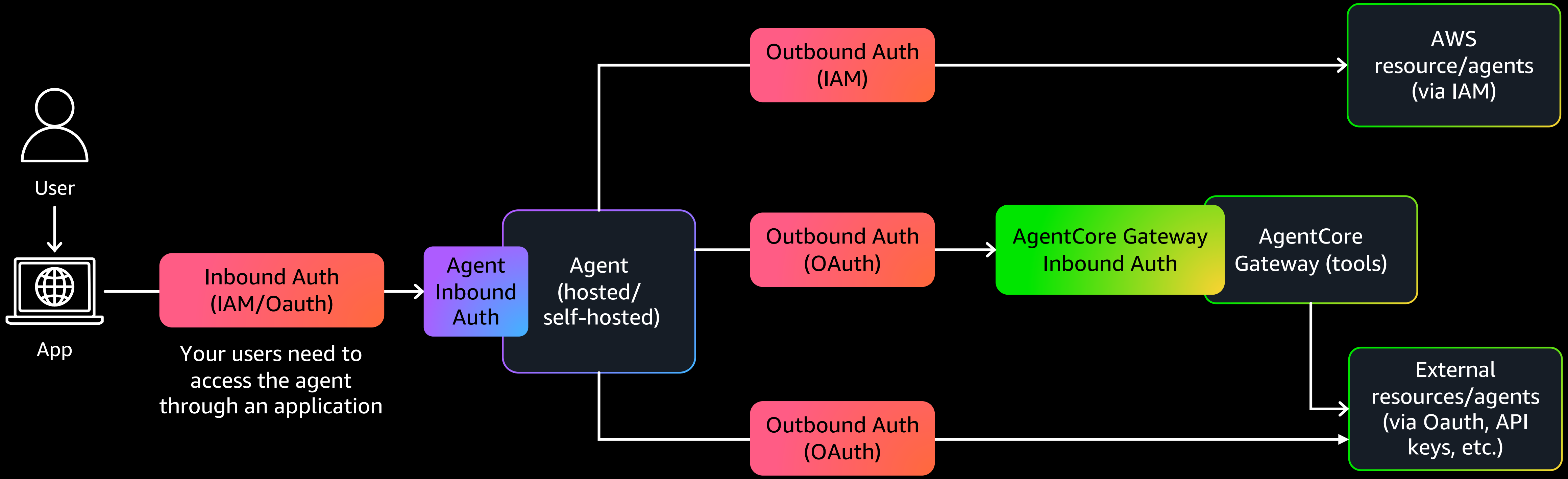
Build streamlined AI agent experiences

- 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


User
App
Inbound Auth (IAM/Oauth)
Agent Inbound Auth
Agent (hosted/ self-hosted)
Outbound Auth (IAM)
AWS resource/agents (via IAM)
Outbound Auth (OAuth)
AgentCore Gateway Inbound Auth
AgentCore Gateway (tools)
Outbound Auth (OAuth)
External resources/agents (via Oauth, API keys, etc.)

Your users need to access the agent through an application

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

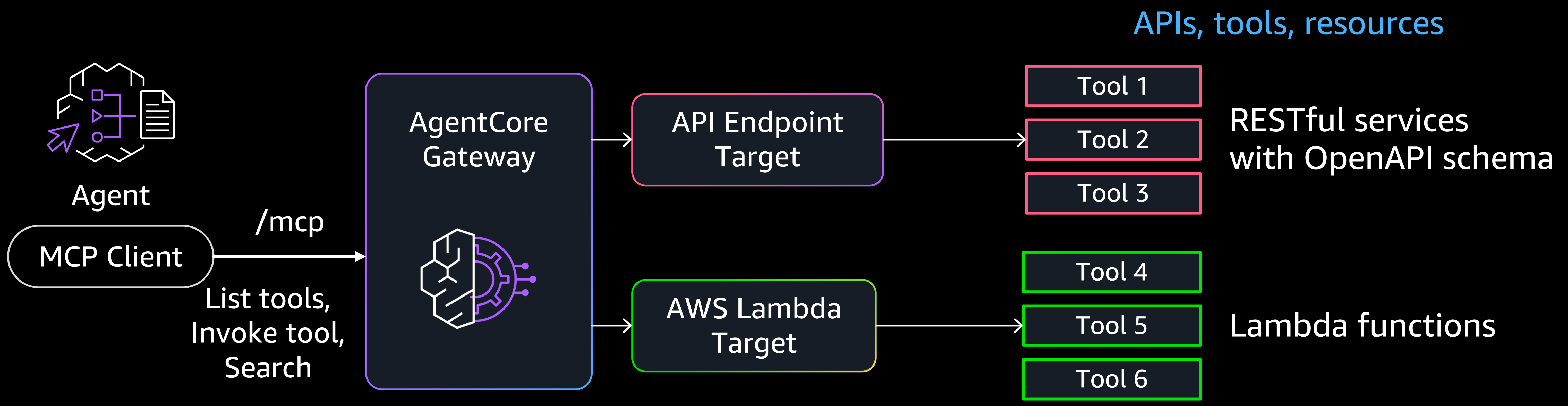
Secure and unified access

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

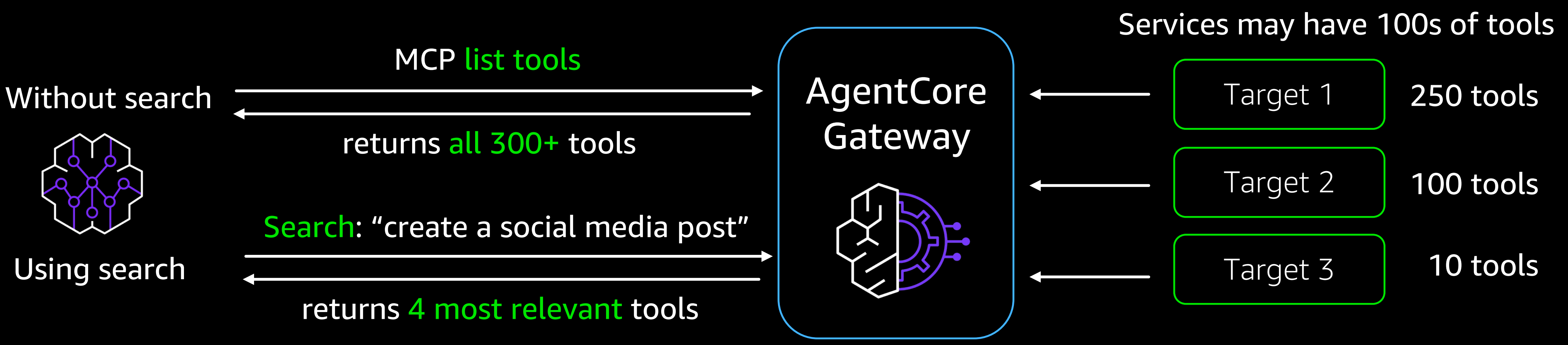
Intelligent tool discovery

- 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

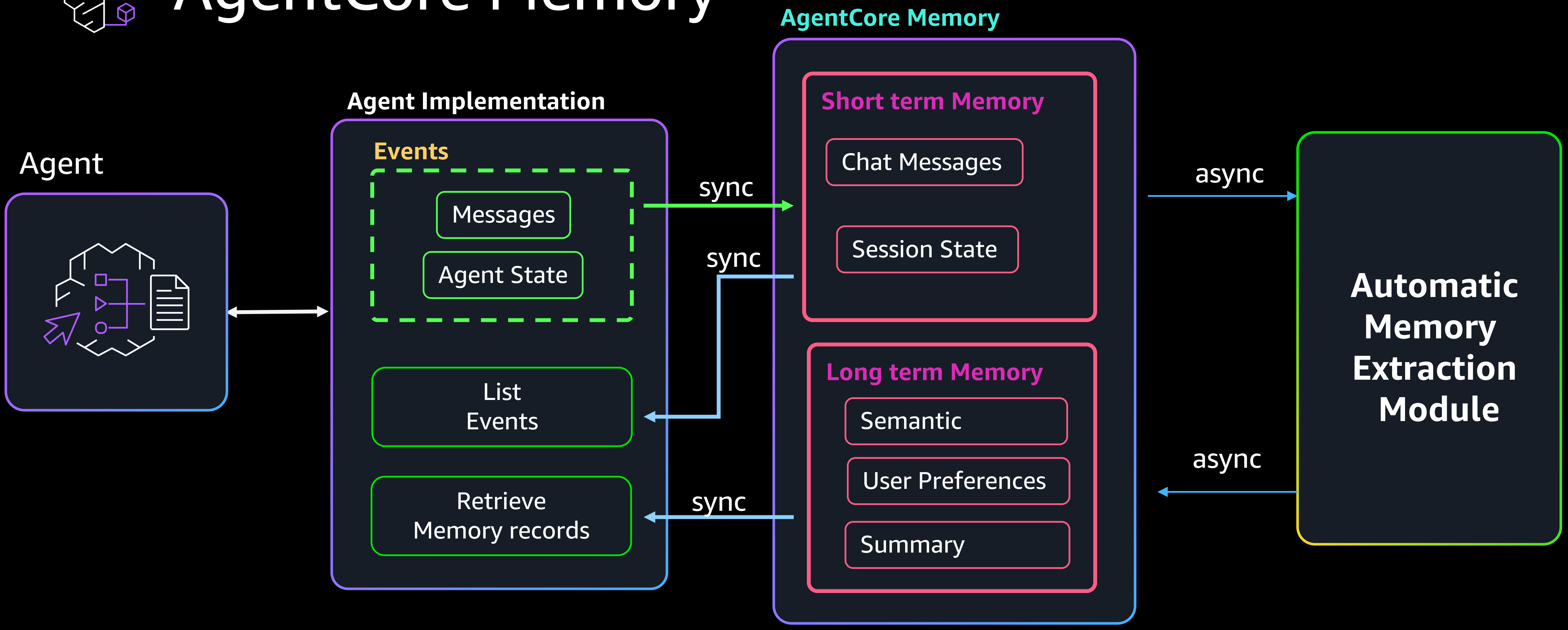
Enterprise-grade

- 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

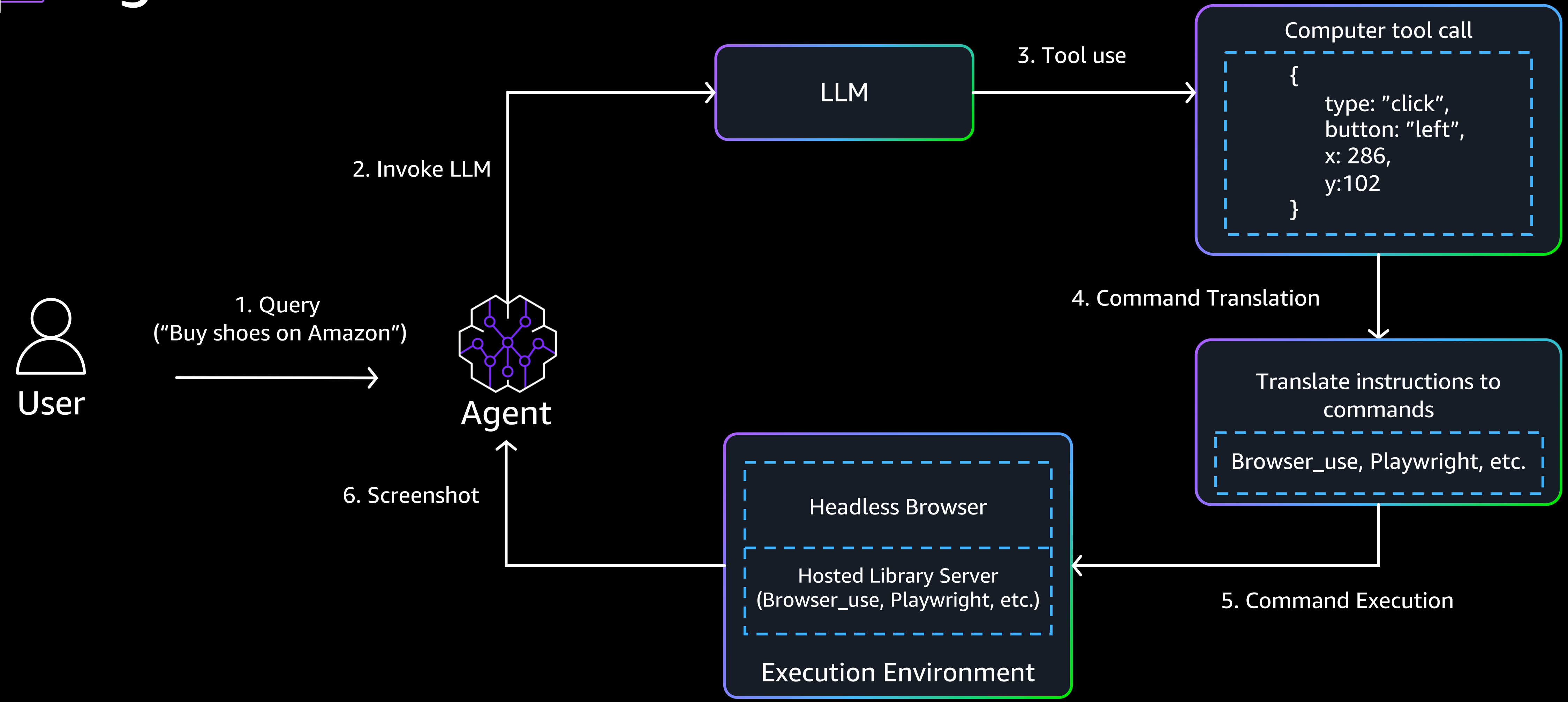
Enterprise-grade security

- 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



```
{
    type: "click",
    button: "left",
    x: 286,
    y:102
}
```

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

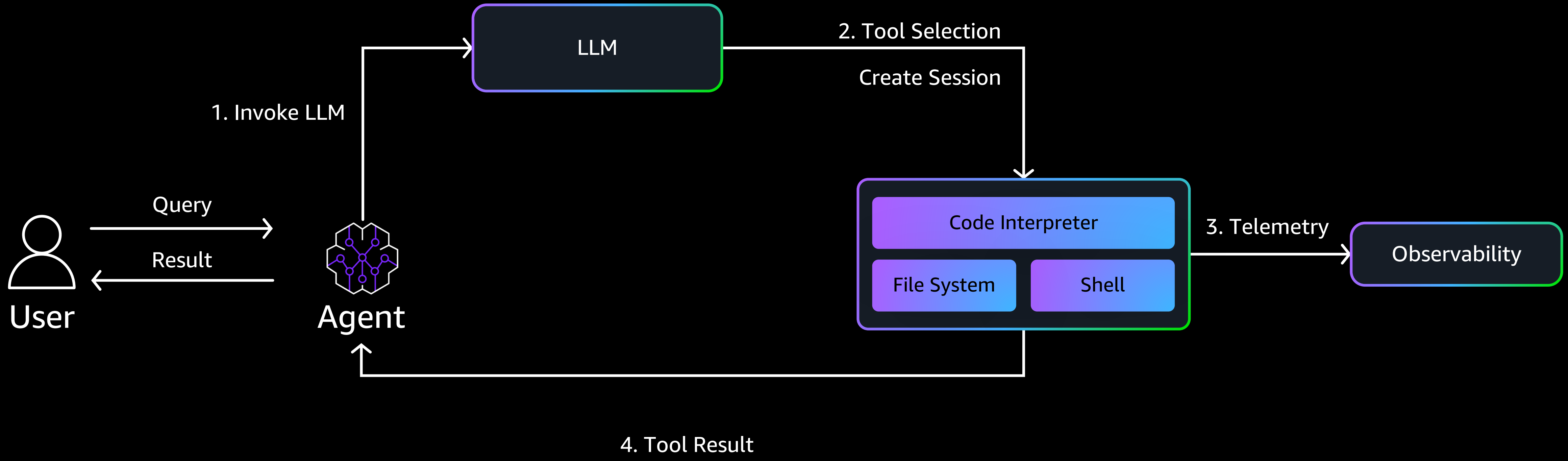
Large-scale data processing

- 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

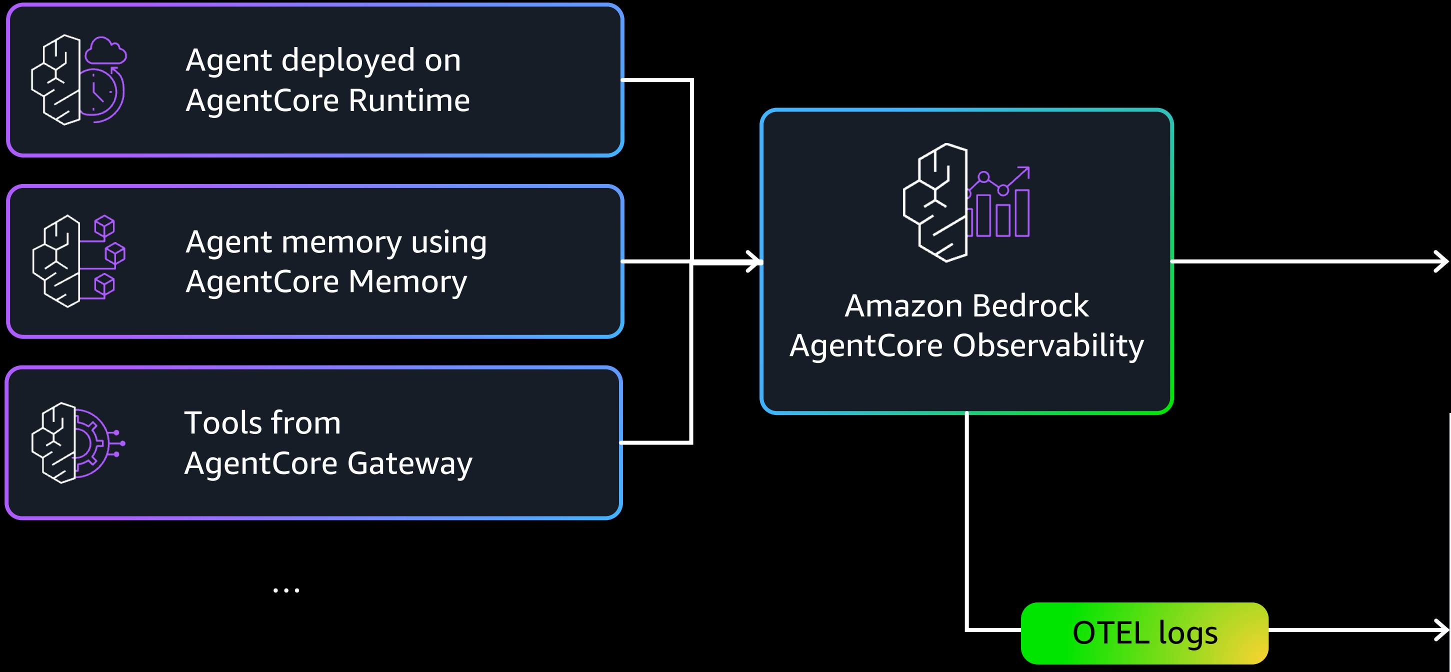
Maintain quality and trust

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

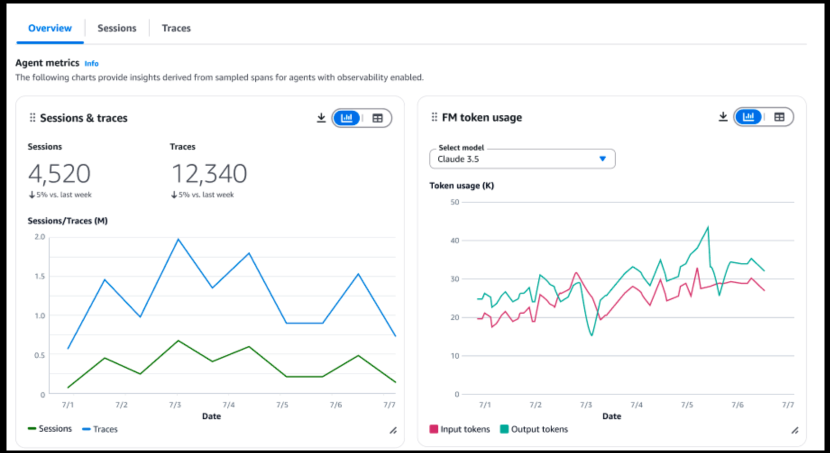
Integrate with 3P observability tools

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

AgentCore Observability


Agent deployed on AgentCore Runtime
Agent memory using AgentCore Memory
Tools from AgentCore Gateway
...
Amazon Bedrock AgentCore Observability
OTEL logs

AgentCore Observability dashboards



Agent metrics

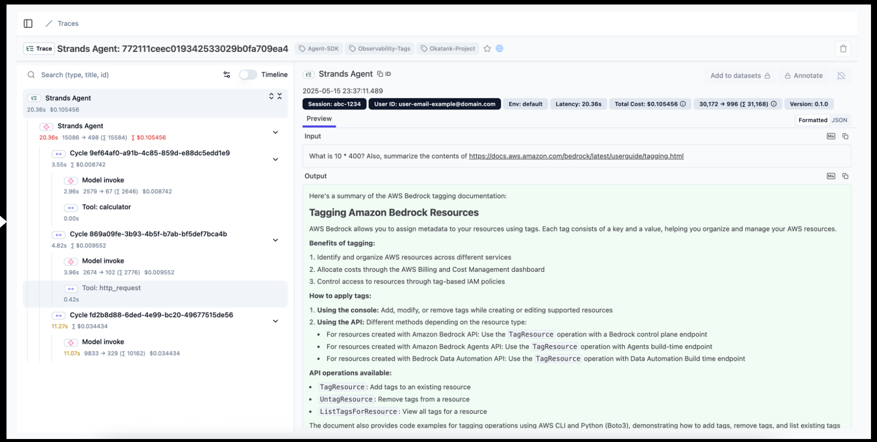
The following charts provide insights derived from sampled spans for agents with observability enabled.

Sessions: 4,520 ↓5% vs. last week

Traces: 12,340 ↓5% vs. last week

Select model: Claude 3.5

Third-party observability dashboards



Strands Agent: 772111ceec019342533029b0fa709ea4

Strands Agent
20.36s $0.105456

Strands Agent
20.36s 15086 → 498 (∑ 15584) ∑ $0.105456

Cycle 9ef64af0-a91b-4c85-859d-e88dc5edd1e9
3.55s ∑ $0.008742

Model invoke
2.96s 2579 → 67 (∑ 2646) $0.008742

Tool: calculator
0.00s

Cycle 869a09fe-3b93-4b5f-b7ab-bf5def7bca4b
4.82s ∑ $0.009552

Model invoke
3.96s 2674 → 102 (∑ 2776) $0.009552

Tool: http_request
0.42s

Cycle fd2b8d88-6ded-4e99-bc20-49677515de56
11.27s ∑ $0.034434

Model invoke
11.07s 9833 → 329 (∑ 10162) $0.034434

Strands Agent

2025-05-15 23:37:11.489

Session: abc-1234 | User ID: user-email-example@domain.com | Env: default | Latency: 20.36s | Total Cost: $0.105456 | 30,172 → 996 (∑ 31,168) | Version: 0.1.0

Input

What is 10 * 400? Also, summarize the contents of https://docs.aws.amazon.com/bedrock/latest/userguide/tagging.html

Output

Here's a summary of the AWS Bedrock tagging documentation:

Tagging Amazon Bedrock Resources

AWS Bedrock allows you to assign metadata to your resources using tags. Each tag consists of a key and a value, helping you organize and manage your AWS resources.

Benefits of tagging:

1. Identify and organize AWS resources across different services
2. Allocate costs through the AWS Billing and Cost Management dashboard
3. Control access to resources through tag-based IAM policies

How to apply tags:

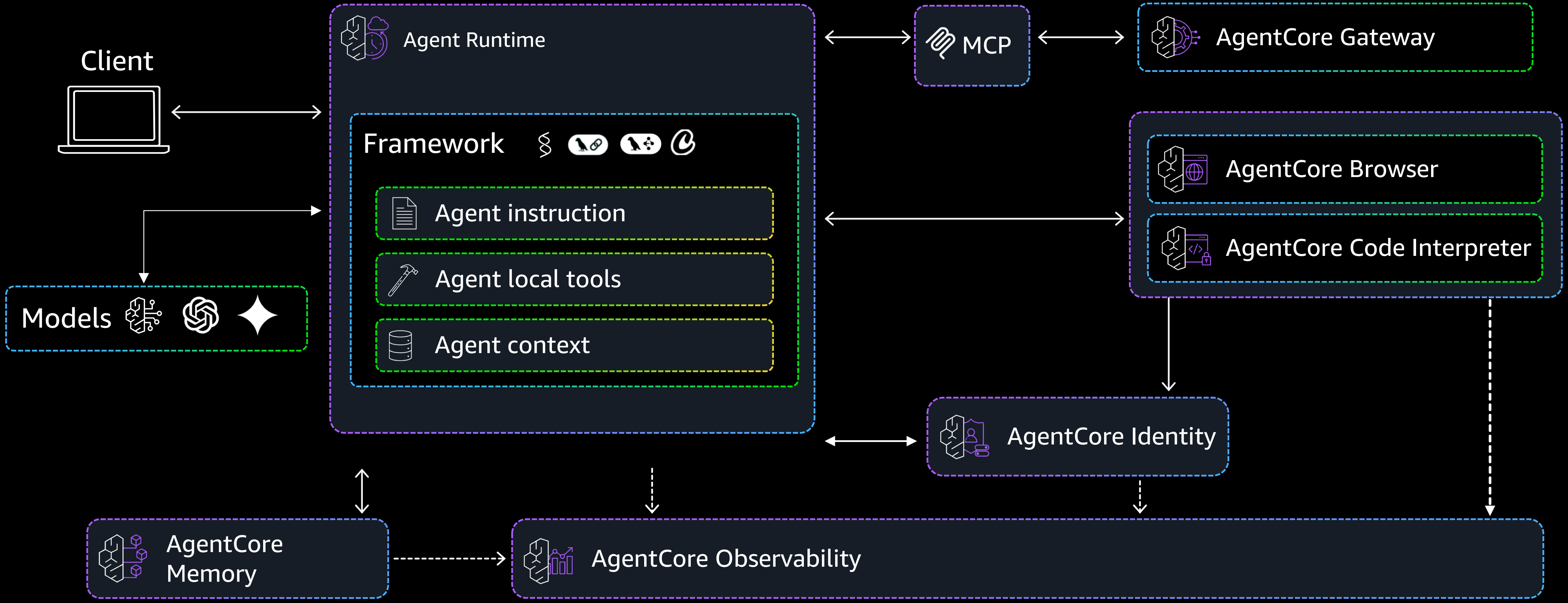
1. **Using the console:** Add, modify, or remove tags while creating or editing supported resources
2. **Using the API:** Different methods depending on the resource type:
   - For resources created with Amazon Bedrock API: Use the `TagResource` operation with a Bedrock control plane endpoint
   - For resources created with Amazon Bedrock Agents API: Use the `TagResource` operation with Agents build-time endpoint
   - For resources created with Bedrock Data Automation API: Use the `TagResource` operation with Data Automation Build time endpoint

API operations available:

- `TagResource`: Add tags to an existing resource
- `UntagResource`: Remove tags from a resource
- `ListTagsForResource`: View all tags for a resource

The document also provides code examples for tagging operations using AWS CLI and Python (Boto3), demonstrating how to add tags, remove tags, and list existing tags

AgentCore capabilities enabling agents at scale



AI agents and tools in AWS Marketplace

PRE-BUILT AGENTS

ANTHROP\C

AUTOMATION™ ANYWHERE

karini

AGENT TOOLS

brave

CONFLUENT

Saviynt

tavily

AGENT DEV SOLUTIONS

accenture

Agentforce

lyzr

PROFESSIONAL SERVICES

Deloitte.

genpact

IBM

SOFTWARE SOLUTIONS

nlx

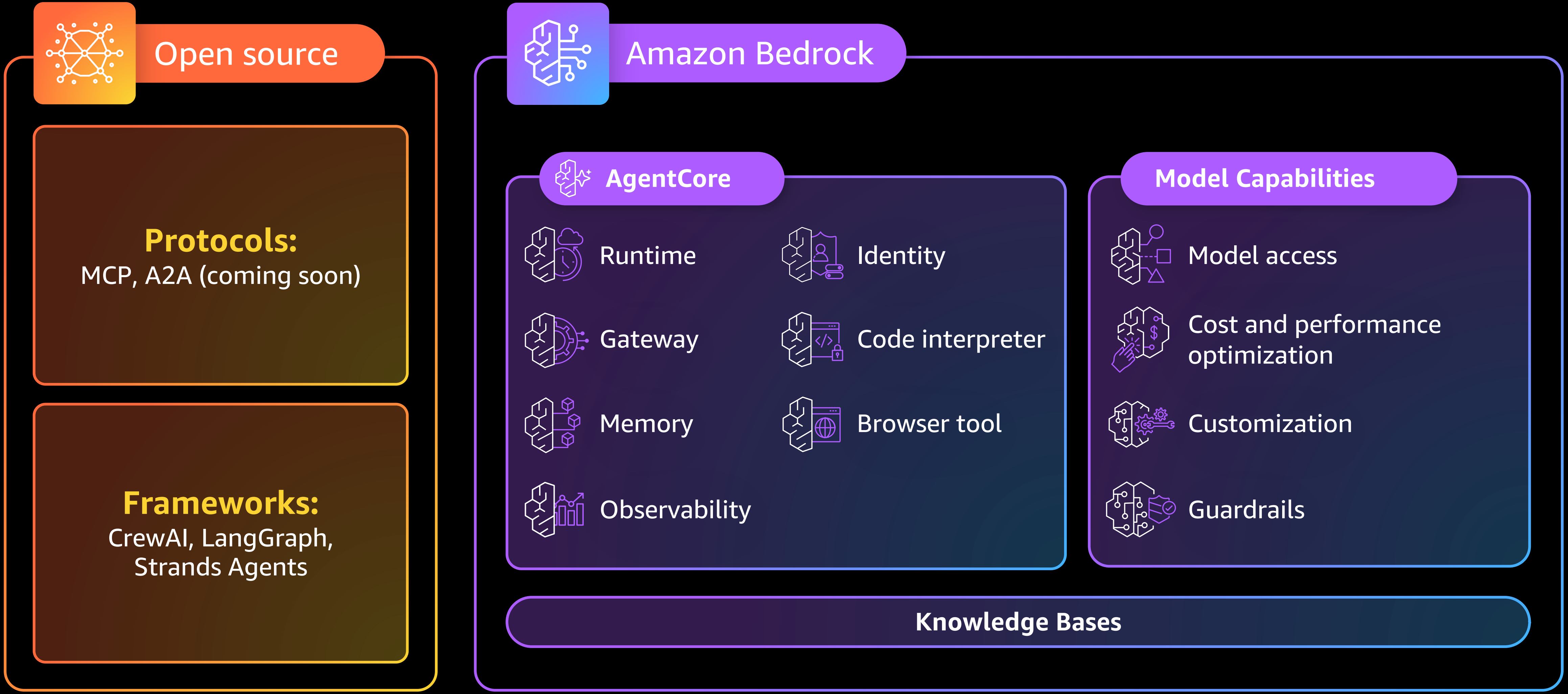
servicenow

snowflake

aws

This is not a complete list. To view all products listed in AWS Marketplace, visit AWS Marketplace.
This list is current as of July 15, 2025.

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

Thank you!