



SUMMIT INDIA

OET101

Amazon Bedrock - Simplifying generative AI app development

Sahil Verma

(he/him)

Senior Solutions Architect

AWS India Pvt. Ltd.

sahver@amazon.com



Agenda

Challenges of building generative AI applications

Building generative AI applications with Amazon Bedrock

Demo

Summary

Start your generative AI journey

Building generative AI
applications **is challenging**



Accessing multiple FMs
and newer versions



Customizing FMs



Maintaining data privacy
and security



Getting FMs
to execute tasks



Connecting to
data sources



Managing infrastructure

GENERALLY AVAILABLE

Amazon Bedrock

The easiest way to build
and scale generative AI
applications with
foundation models



Access a range of leading FMs through a single API



Privately customize FMs using your organization's data



Build agents that execute complex business tasks by dynamically invoking APIs

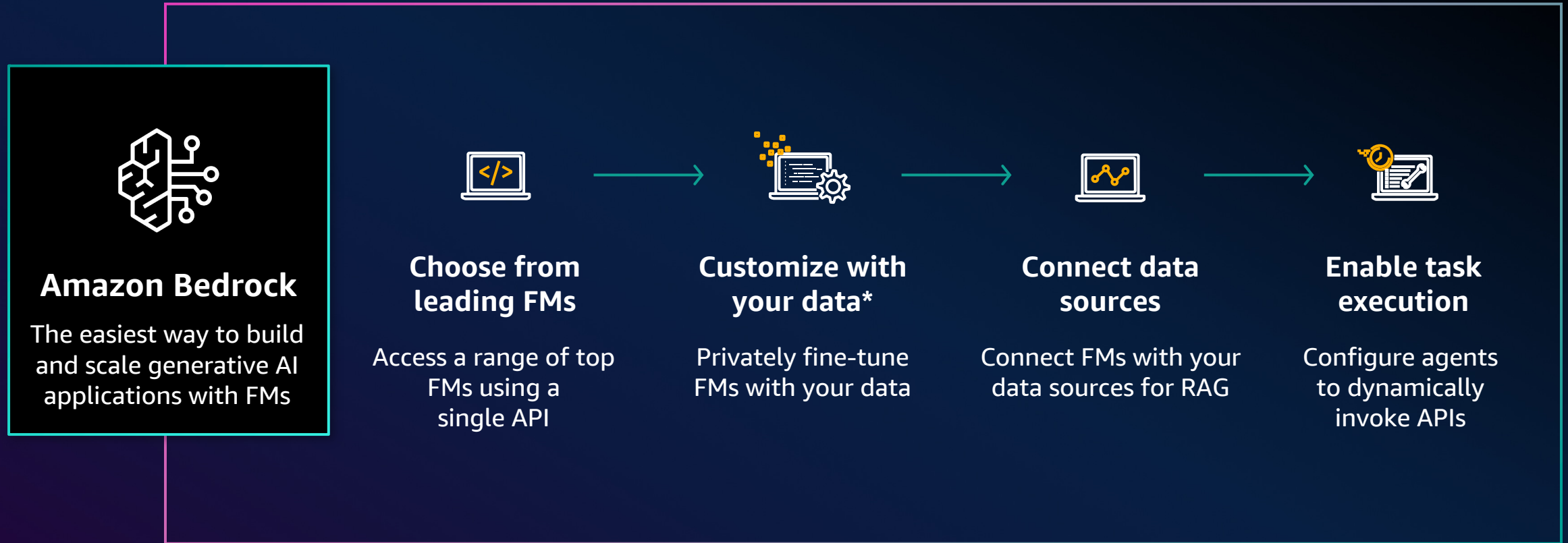


Extend the power of FMs with your data using Retrieval Augmented Generation (RAG)



Enable data security and compliance

How does Amazon Bedrock work?



*your content is not used to improve the base models and is not shared with third-party model providers.

Amazon Bedrock supports leading foundation models

AI21labs

Jurassic-2

Contextual answers, summarization, paraphrasing

ANTHROPIC

Claude 3, Claude 2.1 & Claude Instant

Multimodal reasoning, vision capabilities, translation, summarization, writing, coding

cohere

Command & Embed

Text generation, search, classification

Mistral AI

Mistral Large, 7B & Mixtral 8X7B

Text summarization, Q&A, Text classification, Text completion, code generation

Meta

Llama 2

Dialogue use cases and language tasks

stability.ai

Stable Diffusion XL 1.0

High-quality images and art

Amazon Titan

Summarization, image and text generation and search, Q&A

Amazon Bedrock core API: InvokeModel

```
bedrock.invoke_model (  
    modelId = model_id,  
    contentType = "...",  
    accept = "...",  
    body = body  
)
```



**Use foundation
models**

Amazon Titan models
Third-party models
Fine-tuned models

Model Evaluation on Amazon Bedrock

EVALUATE FMS TO SELECT THE BEST ONE FOR YOUR USE CASE

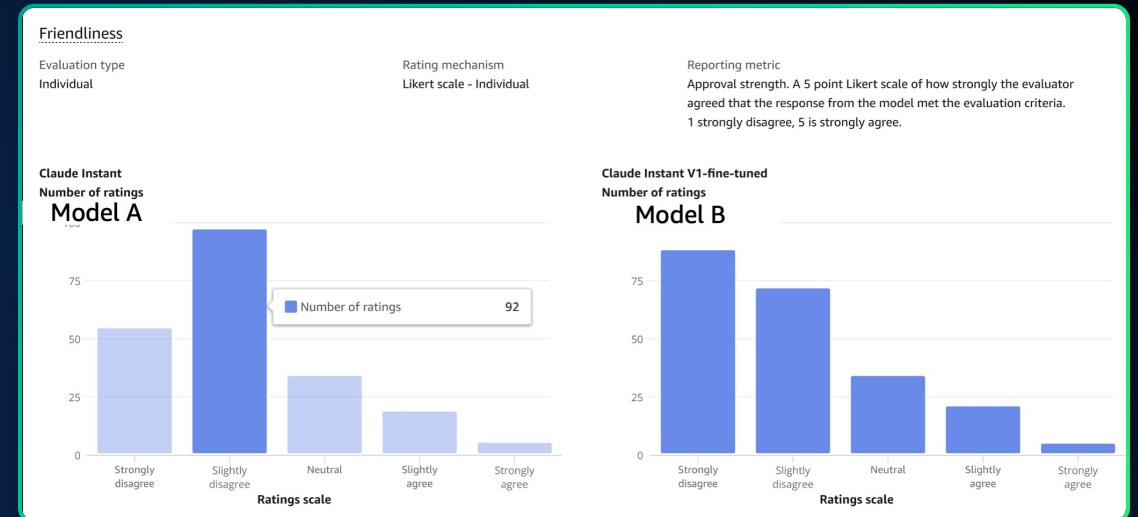
Preview

Choose automatic or human evaluation method

Curated datasets or bring your own

Pre-defined and custom metrics

Human evaluation report



Automatic evaluation report

Text summarization evaluation summary (3)

The results for text summarization consist of accuracy, toxicity, and robustness, which indicate the quality of the summaries generated by the model. [Learn more.](#)

Accuracy	
Dataset	Value
CNN/DailyMail	.6
S3 URI 3	.4

Toxicity	
Dataset	Value
S3 URI	.5

Robustness	
Dataset	Value
CNN/DailyMail	.4
S3 URI 2	.6

**Your data is your
differentiator**



Customize your foundation models



**Customize to
specific business
needs**

Example: Healthcare –
Understand medical terminology
and provide accurate responses
related to patients' health



**Adapt to
domain-specific
language**

Example: Finance –
Teach financial and accounting
terms to provide good analyses
for earnings reports



**Enhance
performance
for specific tasks**

Example: Customer service –
Improve ability to understand and
respond to customers' inquiries
and complaints



**Improve
context-awareness
in responses**

Example: Legal services –
Better understand case facts
and law to provide useful
insights for attorneys

Common approaches for customizing FMs



Privately customize models with your data

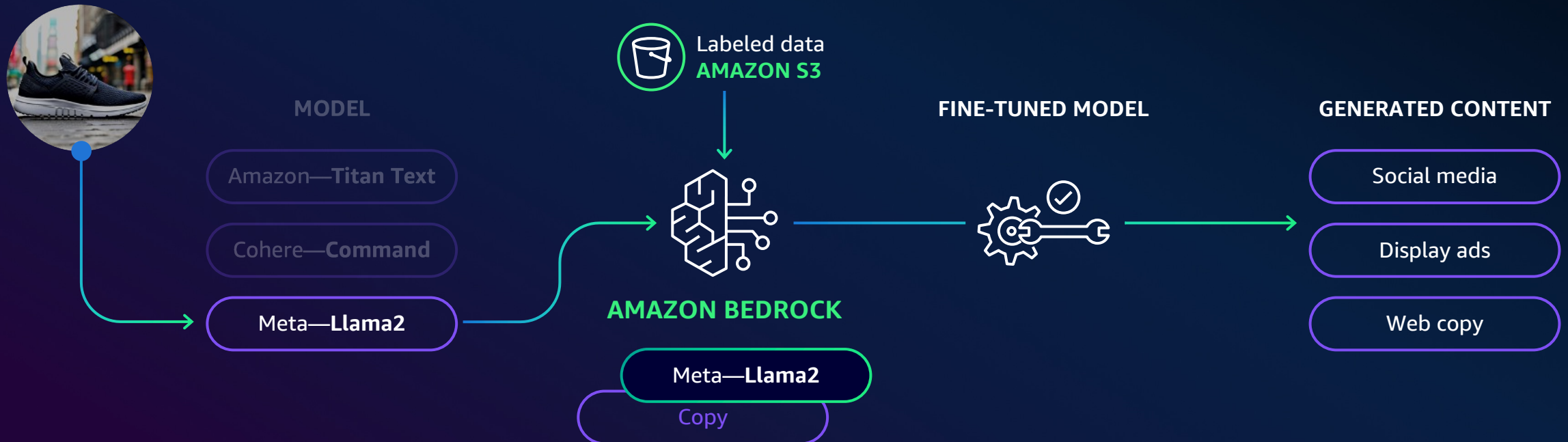
FINE-TUNING AND CONTINUED PRE-TRAINING

Deliver tailored, differentiated tail user experiences with customized FMs

Fine-tune Llama 2, Command, and Titan FMs for specific tasks with labeled data

Use continued pre-training to adapt Titan Text FMs to your domain with unlabeled data

None of your inputs to or outputs from Amazon Bedrock will be used to train the original base models



Knowledge bases for Amazon Bedrock

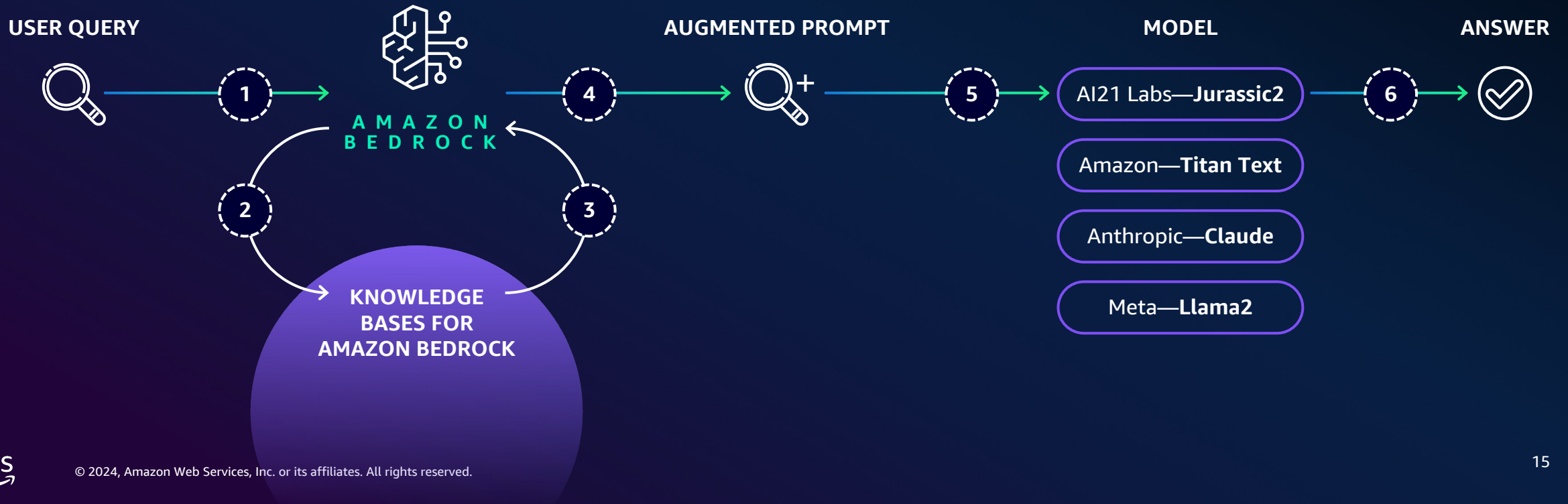
NATIVE SUPPORT FOR RETRIEVAL AUGMENTED GENERATION (RAG)

Securely connect FMs to data sources for RAG to deliver more relevant responses

Fully managed RAG workflow including ingestion, retrieval, and augmentation

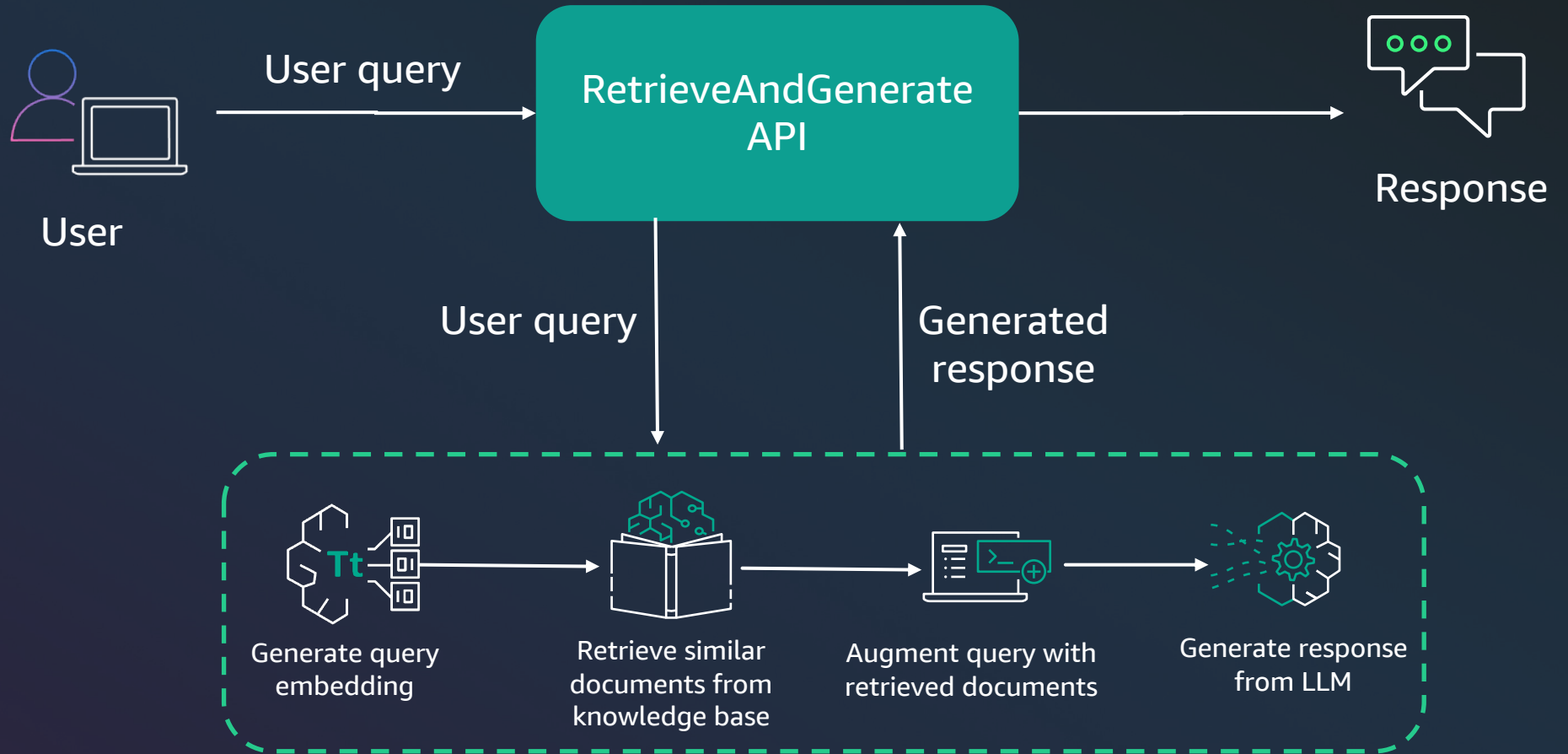
Built-in session context management for multi-turn conversations

Automatic citations with retrievals to improve transparency



Simple RetrieveAndGenerate API

Fully
managed
RAG



**Foundation models alone
cannot execute tasks**

Agents for Amazon Bedrock

ENABLE GENERATIVE AI APPLICATIONS TO COMPLETE TASKS IN JUST A FEW QUICK STEPS



1

**Select your
foundation model**



2

**Provide basic
instructions**



3

**Select relevant
data sources**



4

**Specify
available actions**

| Breaks down and orchestrates tasks |

| Securely accesses and retrieves company data |

| Takes action by invoking API calls on your behalf |

| Provides fully managed infrastructure |

Protecting your generative AI application

Security and privacy

Private connectivity between Amazon Bedrock and your Amazon Virtual Private Cloud (Amazon VPC)

Your data is always encrypted in transit and at rest, and you can also use your keys to encrypt the data

Support for standards, including GDPR compliance and HIPAA eligibility

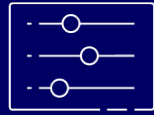
Your content is not used to improve the base models and is not shared with third-party model providers



PREVIEW

Guardrails for Amazon Bedrock

Implement safeguards customized to your application requirements and responsible AI policies



Apply guardrails to multiple foundation models and Agents for Amazon Bedrock



Configure harmful content filtering based on your responsible AI policies



Define and disallow denied topics with short natural language descriptions



COMING SOON

Redact sensitive PII information in FM responses

Demo



Summary

THE EASIEST WAY TO BUILD AND SCALE GENERATIVE AI APPLICATION WITH FMS



Access a range of leading FMs via a single API



Privately customize FMs with your own data



Enable data security and compliance



Build agents that execute complex business tasks by dynamically invoking APIs



Extend the power of FMs with your data using retrieval augmented generation (RAG)



Get the best price performance without managing infrastructure

Getting started with Amazon Bedrock



**Get started with
Amazon Bedrock**



**Discover Amazon Bedrock
features with
a step-by-step tutorial**



**Dive deep with
Amazon Bedrock
hands-on workshop**

Start your generative AI journey today

1

Select the right use case

2

Empower your teams
through a variety of training
opportunities

3

Get started on a PoC for
your top use cases

skillbuilder.aws 

Build beyond

Redeem your free 7-day
trial of AWS Skill Builder

Thank you!

Sahil Verma

sahver@amazon.com



Please complete the session
survey in the mobile app