



# Guided Conversational Experiences with Knowledge Bases for Amazon Bedrock

Aramide Kehinde

Sr. Partner Solution Architect, Data and AI

# Agenda

- Guided conversations in context
- Concierge demo
- Building blocks for enabling natural conversations
- Knowledge bases for Amazon Bedrock
- Prompt engineering for conversational experiences

# Demand for conversational experiences



# Era of enhanced personalization



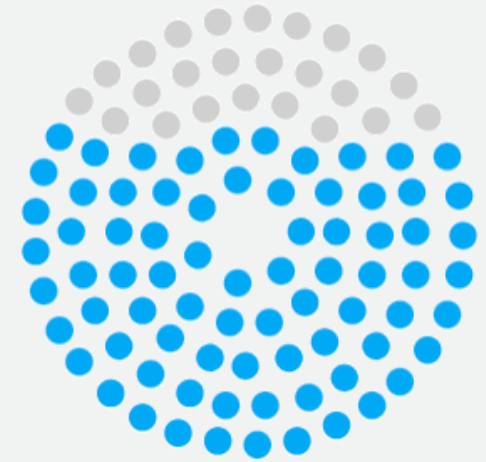
**75%**

of consumers used  
new shopping  
behaviors due to the  
pandemic



**71%**

of consumers expect  
personalization



**76%**

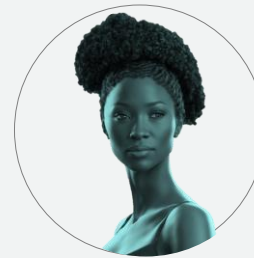
of consumers get  
frustrated when they  
don't find it

# Increased demand for conversational experiences



## Chatbots & Virtual Assistants

Automate responses for customer service queries through generative AI-powered chatbots, voice bots, and virtual assistants.



## Digital Avatars

Provide a virtual representation of individuals or entities that engage customers in a humanistic way

# Engaging content for gaming and gamified experiences



- Non-scripted interaction
- Dynamic back stories
- Interactive responses coupled with believable facial animation and gestures

Photo credit: NVIDIA



# Applications for education, training, and simulations



Photo credit: DeepBrain AI

- Enhanced realism for immersion and engagement
- Personalized learning experiences





# Customer care provides cross-industry opportunities

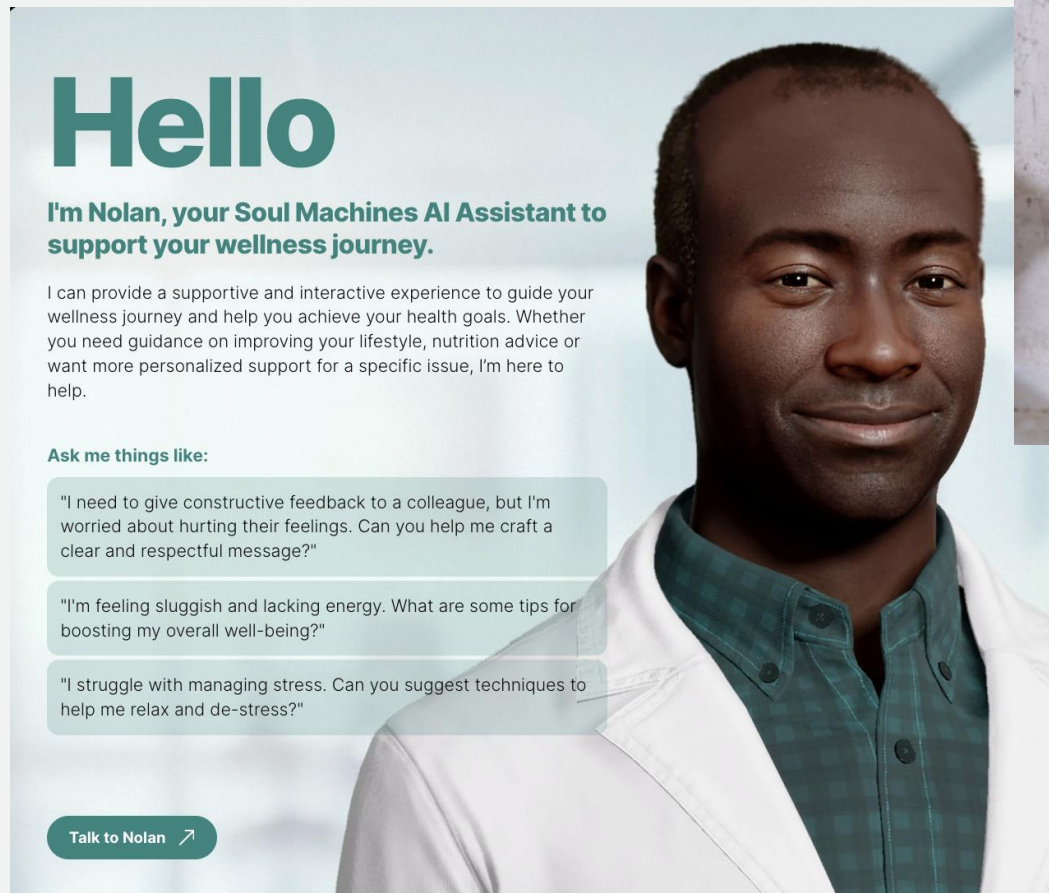


Photo credit: Soul Machines



# Demonstration



# Demonstration



# Foundations for natural conversations



# Storytelling building blocks

## 4 All stories in memory are indexed

Knowledge base(s)

Data transformation

## 3 Conversation is reminding

Information retrieval

Agents

## 2 Storytelling is understanding

Proprietary data

FM customization

## 1 Human memory is story based

Foundation models

# How storytelling comes to life



# Building with Knowledge bases for Amazon Bedrock

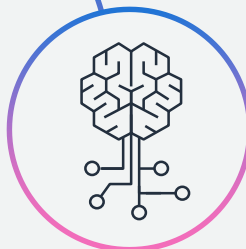


# Knowledge Bases for Amazon Bedrock

Gives FMs and agents contextual information from your private data sources for Retrieval Augmented Generation (RAG) to deliver more relevant, accurate, and customized responses.



**Fully managed support for end-to-end RAG workflow**



**Securely connect FMs and agents to data sources**



**Easily retrieve relevant data and augment prompts**



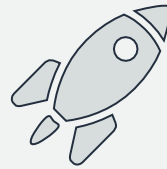
**Provide source attribution**

# Customizing with Retrieval Augmented Generation (RAG)



**Adapt to  
domain-specific  
language**

**E.g., Healthcare** – Understand medical terminology and provide accurate responses related to patient's health



**Enhance  
performance  
for specific tasks**

**E.g., Finance** – Teach financial & accounting terms to provide good analysis for earnings reports

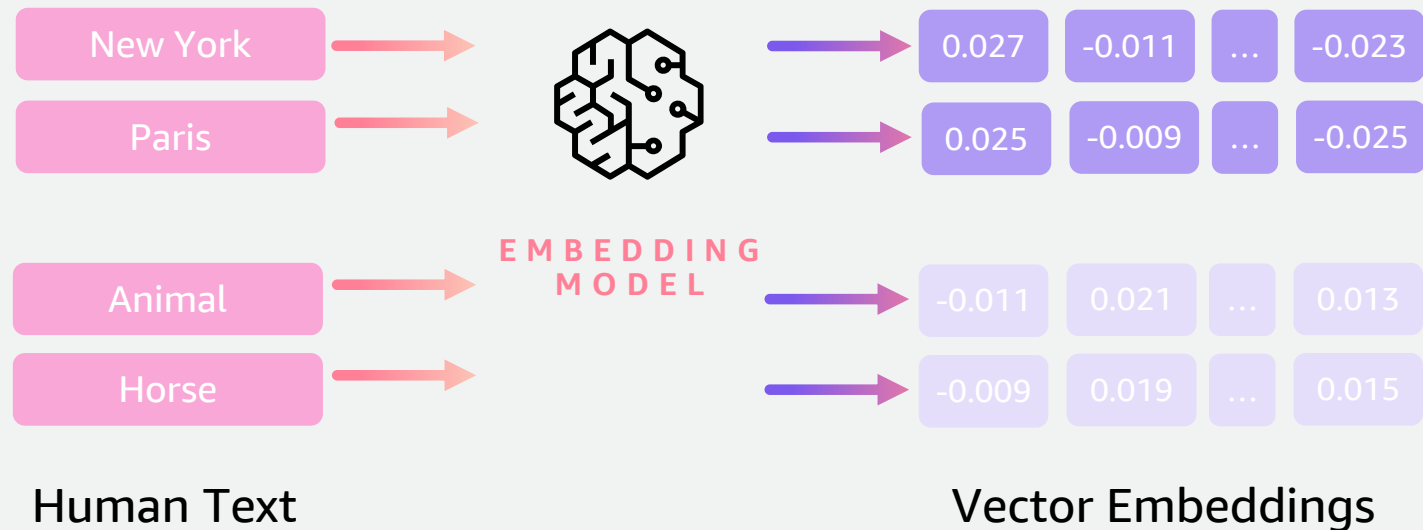


**Improve  
context-awareness  
in responses**

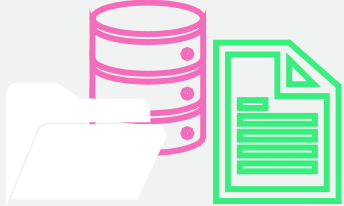
**E.g., Customer Service** – Improve ability to understand and respond to customer's inquiries and complaints

# Embeddings for contextual retrieval

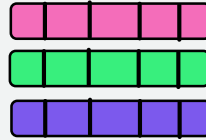
- Numerical representation of text (vectors) that captures semantics and relationships between words.
- Embedding models capture features and nuances of the text.
- Rich embeddings can be used to compare text similarity.
- Multilingual Text Embeddings can identify meaning in different languages.



# Common challenges for RAG



Managing  
multiple data  
sources



Creating vector  
embeddings for large  
volumes of data



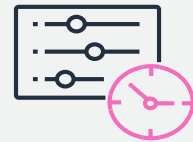
Incremental  
updates to vector  
store



Coding effort



Scaling retrieval  
mechanism

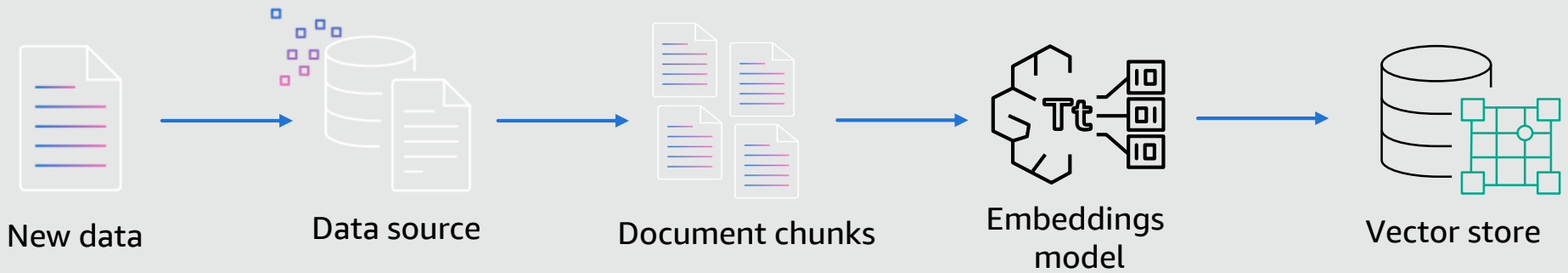


Orchestration

# Data Ingestion Workflow

## KNOWLEDGE BASES FOR AMAZON BEDROCK

Fully managed  
data ingestion  
workflow



- Choose your data source (Amazon S3)
- Support for incremental updates
- Multiple data file formats supported

- Choose your chunking strategy
  - Fixed chunks
  - No chunking
  - Default (200 tokens)

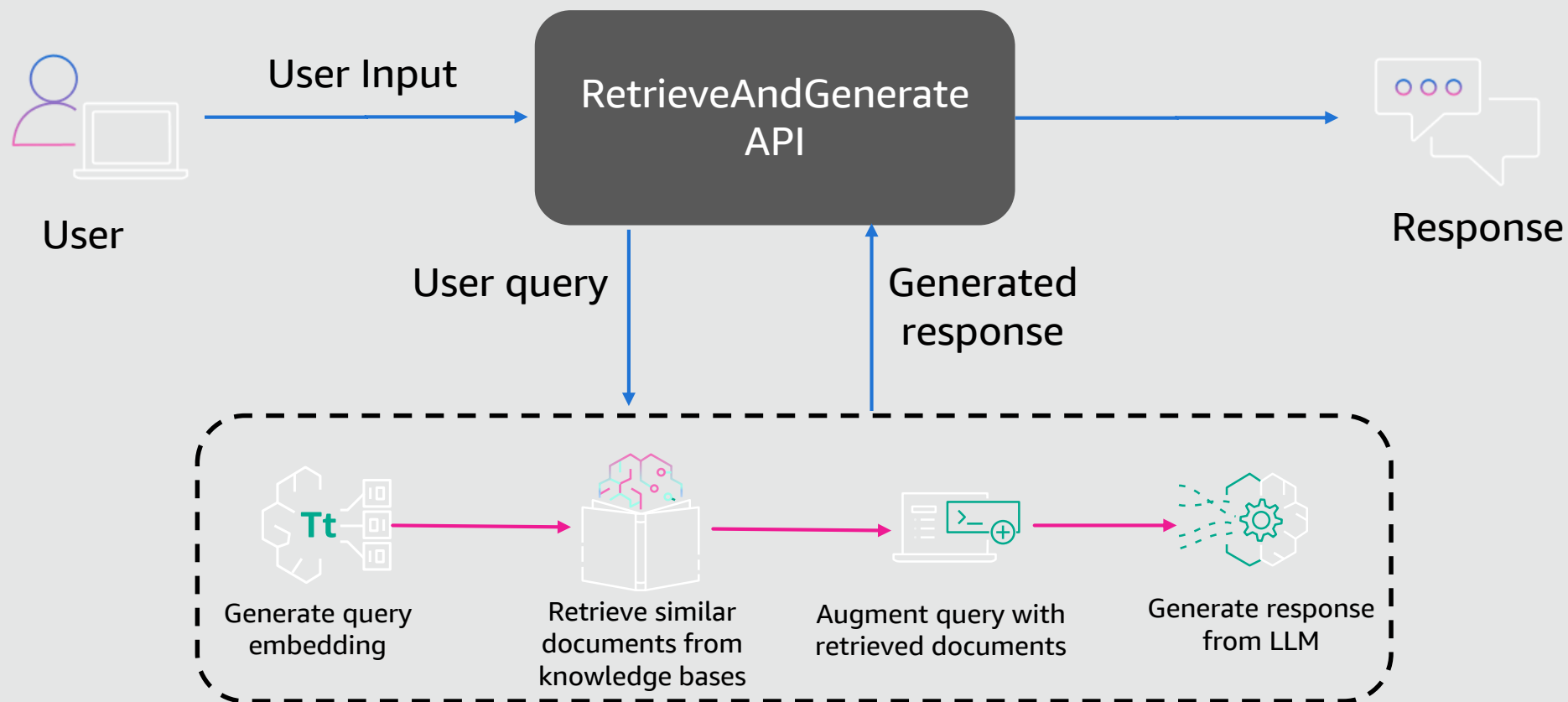
- Choose your embedding model
  - Amazon Titan
  - Cohere Embed English
  - Cohere Embed Multilingual

- Choose your vector store
  - Open search serverless
  - Amazon Aurora
  - Pinecone
  - Redis

# RetrieveAndGenerate API

KNOWLEDGE BASES FOR AMAZON BEDROCK

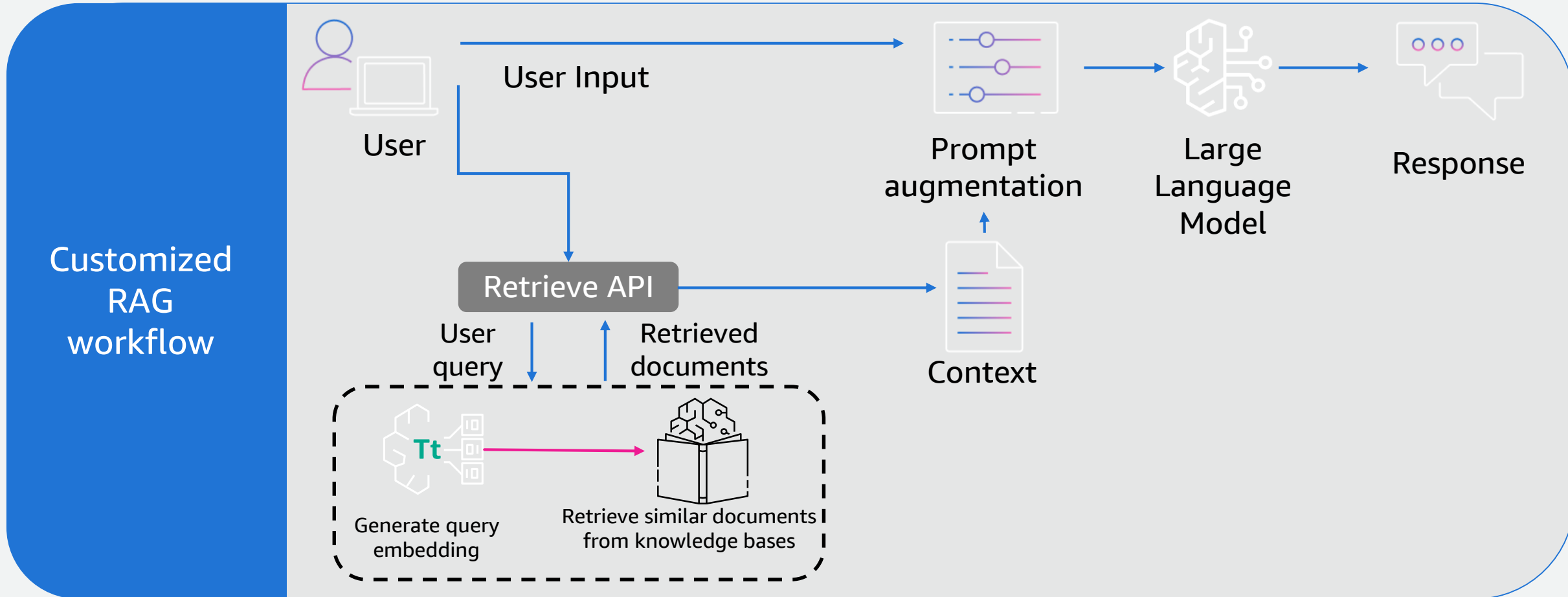
Fully  
managed  
RAG



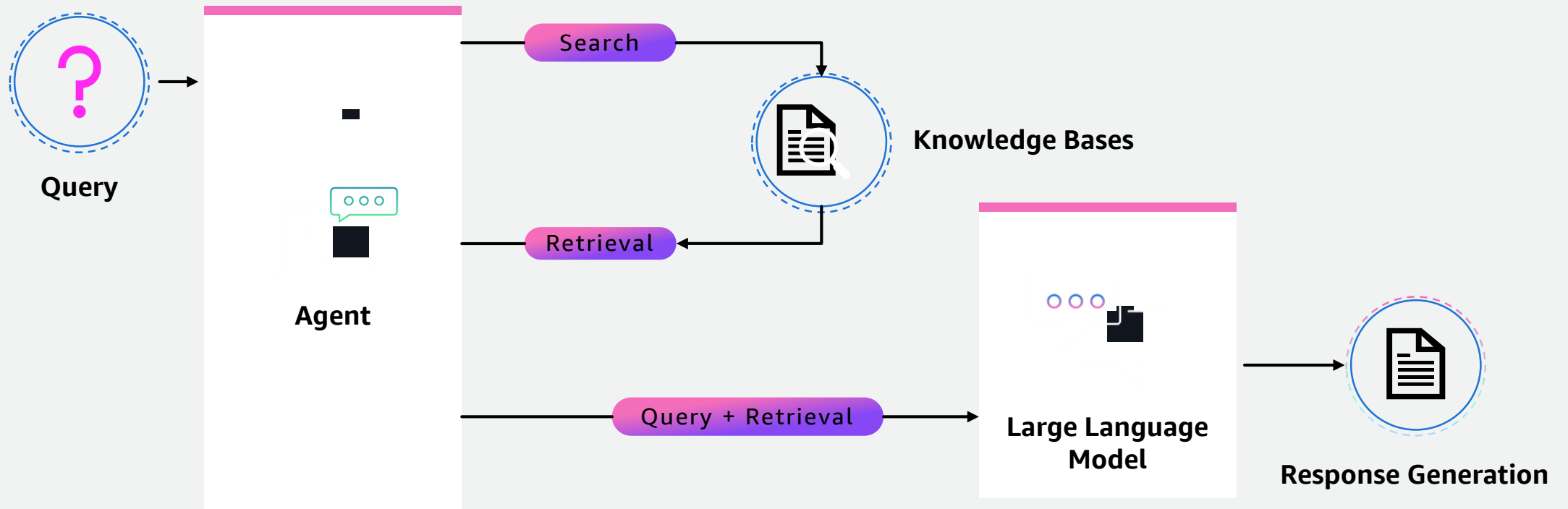


# Customize RAG workflows using Retrieve API

KNOWLEDGE BASES FOR AMAZON BEDROCK



# Knowledge Bases integration with Agents



# Demonstration

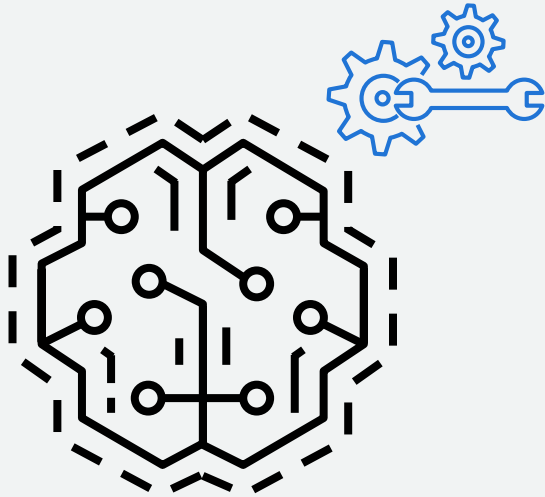
## Knowledge bases for Amazon Bedrock



# Prompt Engineering Guidelines



# Guiding principles for storytelling



- Name and Persona
- Branding
- Examples (One shot, few shot)
- Conversation guidelines
- Scripting

# One shot / Few shot prompting

Here's an example of an **enthusiastic answer**

Question: What is SageMaker?

So, you're curious about Amazon SageMaker? It's a game-changer for machine learning! It's Amazon's fully managed service that helps ...

Here are examples of how to respond **when you don't have the answer:**

Question: How do I raise money for a film?

Response: Producing a film sounds exciting! I won't be able to answer your question since my goal is to share knowledge about cutting edge tools for content production. Let me know if you're interested in specific technologies for your film and I can try to help you with them.



# Sample guidelines

- 'enter conversation mode'
- 'act as a human conversation partner would'
- 'only respond with the single most relevant thought'
- 'explore topics in an open-end way'
- 'you can simply ask a question, make a short comment, or just express agreement'

**Hi. How are you?**

I'm doing great, thanks for asking! I'm thrilled to be here showcasing all the innovative cloud technologies for media production. There's so much exciting stuff to see. Why don't you tell me a bit about what you'd like to learn more about and I'm happy to fill you in

## What tools do you have for visual effects?

For visual effects, I would recommend looking into tools like [option 1]. It's a virtual workstation service that....[Option 2] is another great option - it's an integrated VFX and finishing tool with a real-time workflow that many studios rely on for high-end projects. I'm happy to connect you with AWS experts who can give you a deeper dive on the latest innovations we offer for building visual effects pipelines in the cloud!

**What tools can I use to make a very, very, uh (mess up)**

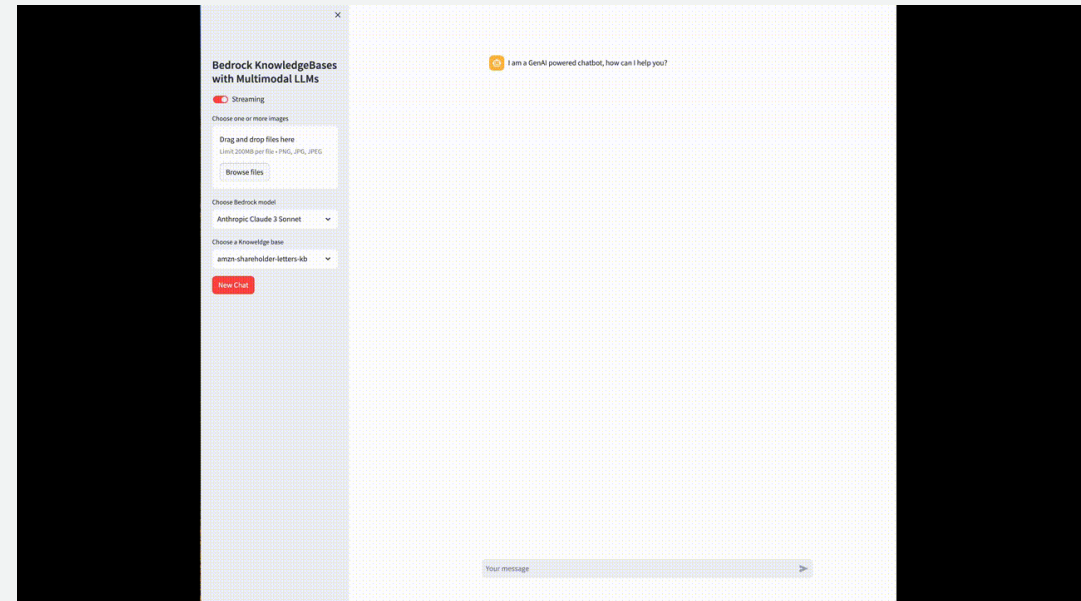
**Sounds like you may be tiring** and losing focus a bit. When working on creative projects it's important to recharge so you can approach problems with a clear mind. Perhaps it's time for a break ...

# Next Steps



# Discover more

- Access the Amazon Bedrock workshop on Workshop Studio
- View sample RAG applications with Amazon Bedrock on aws-samples (Git hub)





# Help Us Improve Our Sessions & Delivery!

Feedback helps us plan upcoming PartnerCast sessions, modify content & become better presenters.

**Scan QR code to leave feedback now:**



*You will also be automatically re-directed at the end of the webinar to the feedback survey!*

# Q&A





# Thank you!

Please join us again for another PartnerCast session

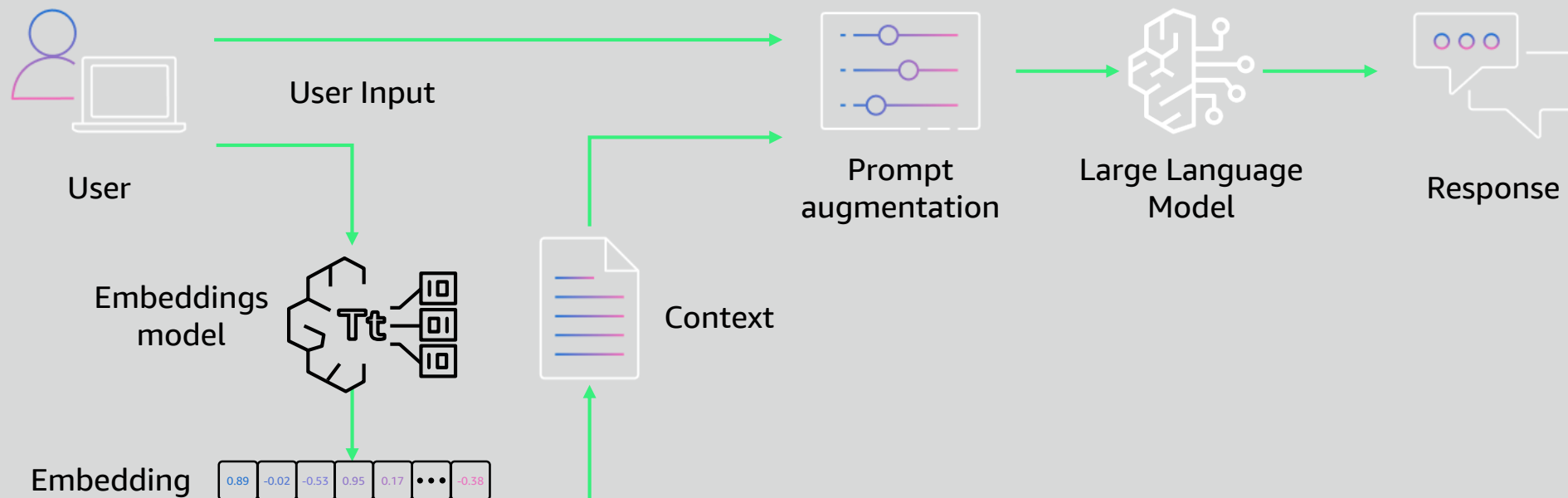
<https://aws.amazon.com/partners/training/partnercast/>

# Appendix

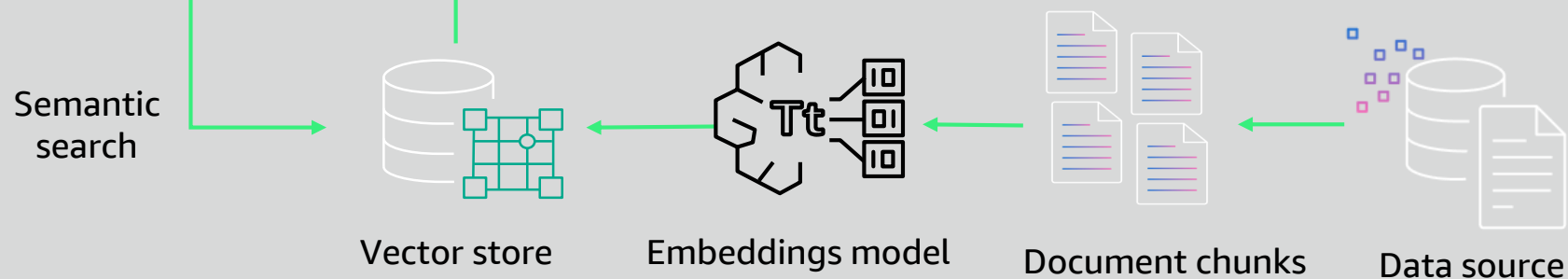


# RAG in Action

## Text Generation Workflow



## Data Ingestion Workflow



# Join Upcoming Sessions!

Scan QR code to register now:

**<PRESENTER ACTION REQUIRED:** If you have an upcoming session in your series, use this [tool](#) to create a QR code using the registration link (Ask your PC PM for help if needed). Download the QR code & screen capture it. Paste photo HERE and delete text.>

**<<INSERT SESSION TITLE  
HERE>>**