



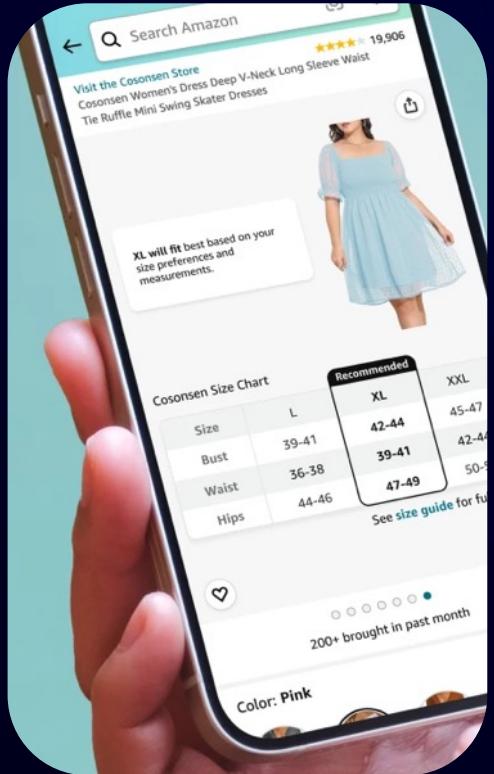
Amazon Bedrock

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

Michael Lin

Sr. Solutions Architect
Amazon Web Services

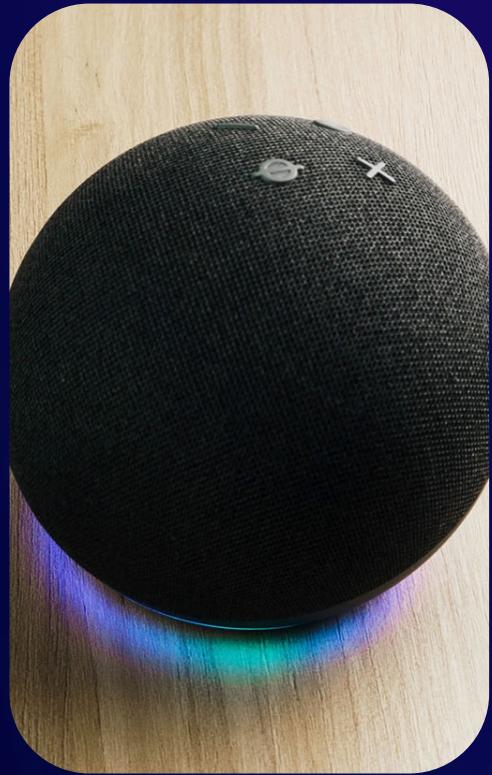
Amazon has been innovating with AI for over **25 years**



4,000 products per minute sold on Amazon.com



First Prime Air delivery on **December 7, 2016**



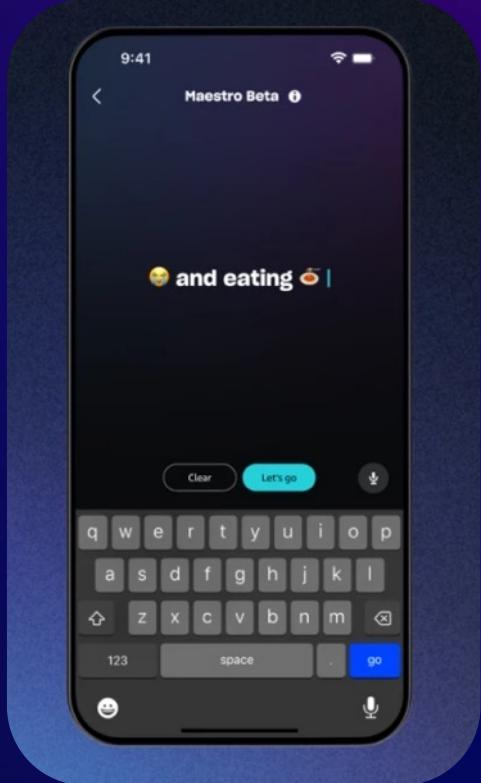
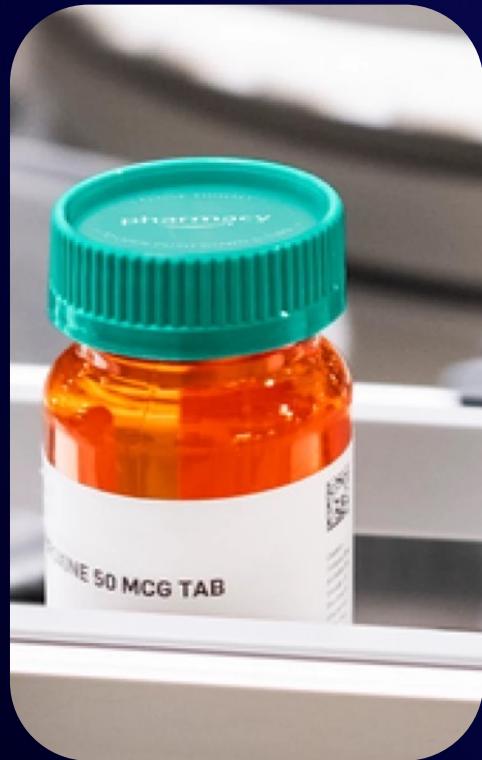
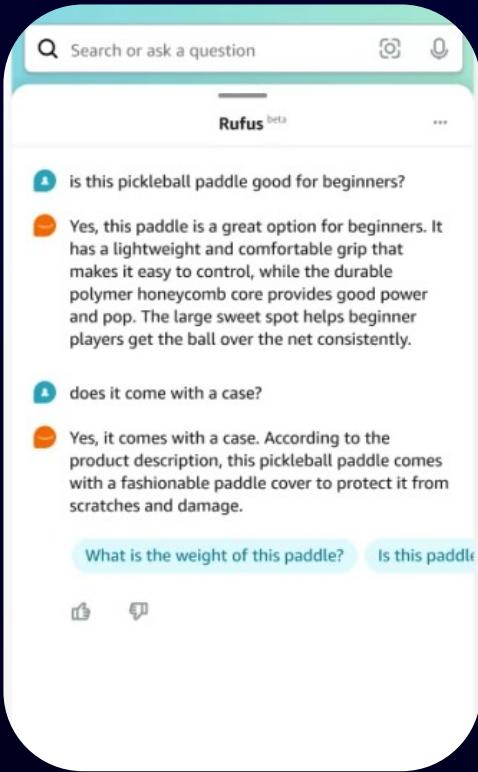
Billions of Alexa interactions each week



1.6M packages every day

Delivering more
immersive, engaging experiences
for Amazon customers with
generative AI

Generative AI innovations at Amazon: Built on AWS



Rufus
An expert shopping assistant

Pharmacy
Faster prescriptions and more helpful support

Ads
AI powered image generation

Maestro
AI playlist generator on Amazon Music

Amazon's approach to innovation:

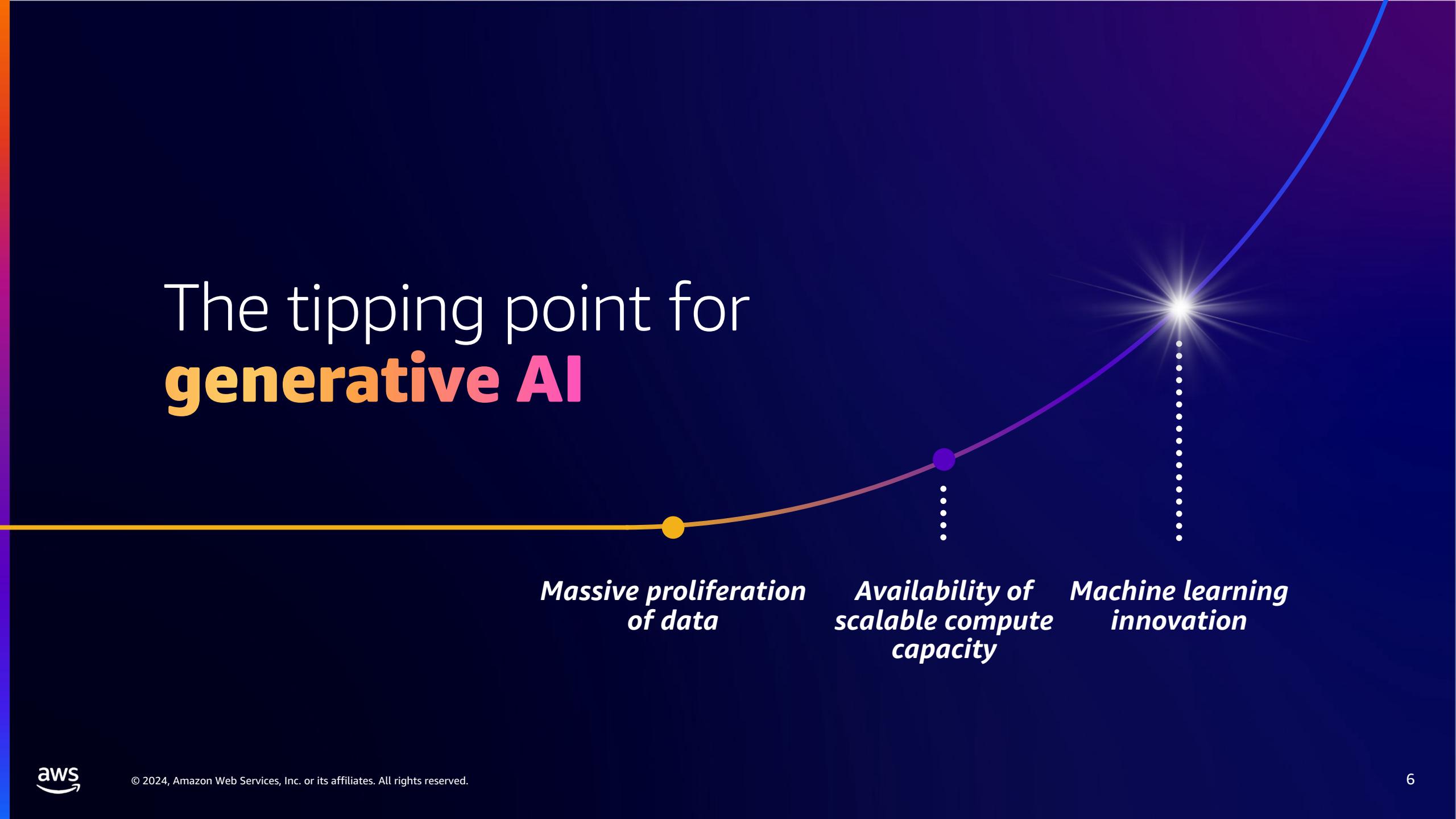
Key considerations to deliver meaningful value at scale

*Customer
obsession*

*Working
backwards*

Scale

The tipping point for **generative AI**

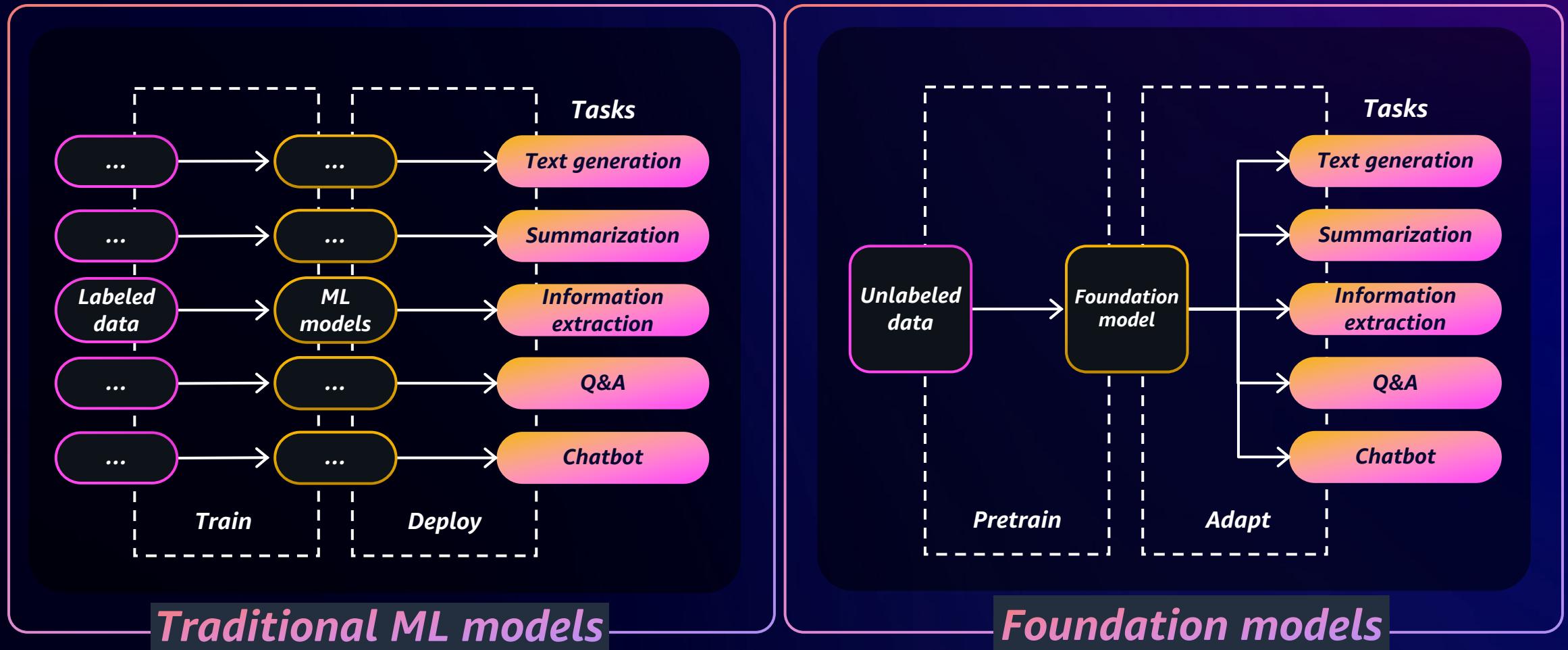


*Massive proliferation
of data*

*Availability of
scalable compute
capacity*

*Machine learning
innovation*

Generative AI is powered by foundation models (FMs)





Amazon Bedrock

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

Choice of leading FMs through a single API

Model customization

Retrieval Augmented Generation (RAG)

Agents that execute multistep tasks

Security, privacy, and safety

Amazon Bedrock

BROAD CHOICE OF MODELS

AI21labs	amazon	ANTHROPIC	cohere	Meta	MISTRAL AI	stability.ai
<i>Contextual answers, summarization, paraphrasing</i>	<i>Text summarization, generation, Q&A, search, image generation</i>	<i>Summarization, complex reasoning, writing, coding</i>	<i>Text generation, search, classification</i>	<i>Q&A and reading comprehension</i>	<i>Text summarization, text classification, text completion, code generation, Q&A</i>	<i>High-quality images and art</i>
Jamba-Instruct	Amazon Titan Text Premier	Claude 3.5 Sonnet	Command	Llama 3 8B	Mistral Small	Stable Diffusion XL1.0
Jurassic-2 Ultra	Amazon Titan Text Lite	Claude 3 Opus	Command Light	Llama 3 70B	Mistral Large	Stable Diffusion XL 0.8
Jurassic-2 Mid	Amazon Titan Text Express	Claude 3 Sonnet	Embed English	Llama 2 13B	Mistral 7B	
	Amazon Titan Text Embeddings	Claude 3 Haiku	Embed Multilingual	Llama 2 70B	Mixtral 8x7B	
	Amazon Titan Text Embeddings V2	Claude 2.1	Command R+			
	Amazon Titan Multimodal Embeddings	Claude 2	Command R			
	Amazon Titan Image Generator	Claude Instant				

Anthropic Claude Models on Bedrock

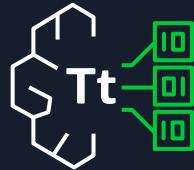
CHOOSE THE EXACT COMBINATION OF INTELLIGENCE, SPEED, AND COST TO SUIT YOUR NEEDS

	Claude 3.5 Sonnet	Claude 3 Haiku	Claude 3 Sonnet	Claude 3 Opus	
Use case	Most intelligent, built for high-volume use cases	Fastest performance at the lowest cost	Balance between intelligence, speed, and cost	Second-most intelligent overall; most intelligent in Claude 3 family	
Context	200K	200K	200K	200K	
Vision	□	□	□	□	
Cost*	Input: Output:	\$0.003 \$0.015	\$0.00025 \$0.00125	\$0.003 \$0.015	\$0.015 \$0.075

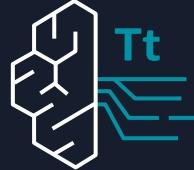
*Per 1K tokens

Amazon Titan

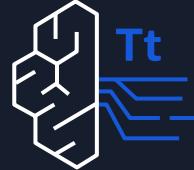
F M S



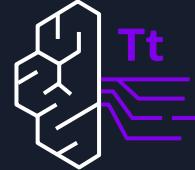
**Amazon Titan
Text
Embeddings**



**Amazon Titan
Text Lite**



**Amazon Titan
Text Express**



**Amazon Titan
Text Premier**



**Amazon Titan
Multimodal
Embeddings**



**Amazon Titan
Image
Generator**



Numerical
representations
of text



Summarization,
copywriting,
fine-tuning



Open-ended
text generation,
conversational
chat, RAG
support



Enterprise-grade
text generation,
optimized for
RAG and Agents



Search,
recommendation,
personalization



Realistic, studio-
quality images

Amazon Titan Multimodal Embeddings



Example

"Blue sneakers
without laces"

Result 1		10 results
1		text: N o Tie Shoelac es -
2		text: s hoe Laces No Tie
3		text: s hoe Laces No Tie
4		text: s hoe Laces No Tie
Result 2		10 results
1		text_description: C apelli New York Toddlers Unisex Slip On Sneaker
2		text_description: T esla Men's Ultra Lightweight Running Shoes L510 PR2
3		text_description: C rocs Men's Norlin Canvas Slip-on Upper: Canvas
4		text_description: B ed Stu Men's Bluegill Slip-On Loafer Slip on

Image Search

AWS Machine Learning Blog

Easily build semantic image search using Amazon Titan

by Mark Watkins and Dan Johns | on 30 NOV 2023 | in Amazon Bedrock, Amazon Comprehend, Amazon OpenSearch Service, Amazon Rekognition, Artificial Intelligence, Customer Solutions, Generative AI, Intermediate (200), Technical How-to, Thought Leadership | Permalink | [Comments](#) | [Share](#)

Werner Vogels loves wearing white scarfs as he travels around India

Search

Adult Female Person Woman Art Handicraft

Body Part Hand Person Adult Male Man Holding Hands

Person Shoving Adult Male Body Part Hand Holding Hands Machine Wheel

<https://aws.amazon.com/blogs/machine-learning/easily-build-semantic-image-search-using-amazon-titan/>



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved

Amazon Titan Image Generator

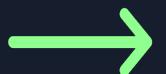


Inpainting

EDIT AN IMAGE



Input image



Inpainting the image with a car



Outpainting

GENERATE DIFFERENT BACKGROUNDS



Input image



Generated images with different backgrounds

Model evaluation on Amazon Bedrock

EVALUATE, COMPARE, AND SELECT THE BEST FM FOR YOUR USE CASE

1

Use curated datasets or bring your own for tailored results

2

Apply automatic or human evaluation methods

3

Use your in-house team or reviewers managed by AWS

4

Provides predefined and custom metrics

5

Get results in just a few quick steps

Model evaluation in Amazon Bedrock

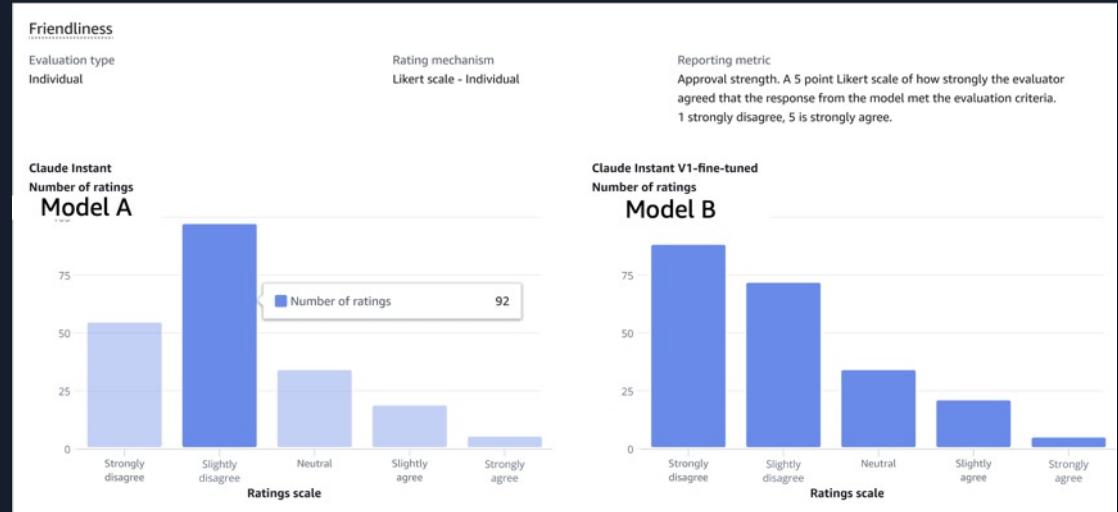
EVALUATE FMS TO SELECT THE BEST ONE FOR YOUR USE CASE

Automatic or human evaluation method

Curated datasets or bring your own

Predefined 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 1	.5
Robustness	
Dataset	Value
CNN/DailyMail	.4
S3 URI 2	.6



Common approaches for customizing FMs

Prompt
engineering

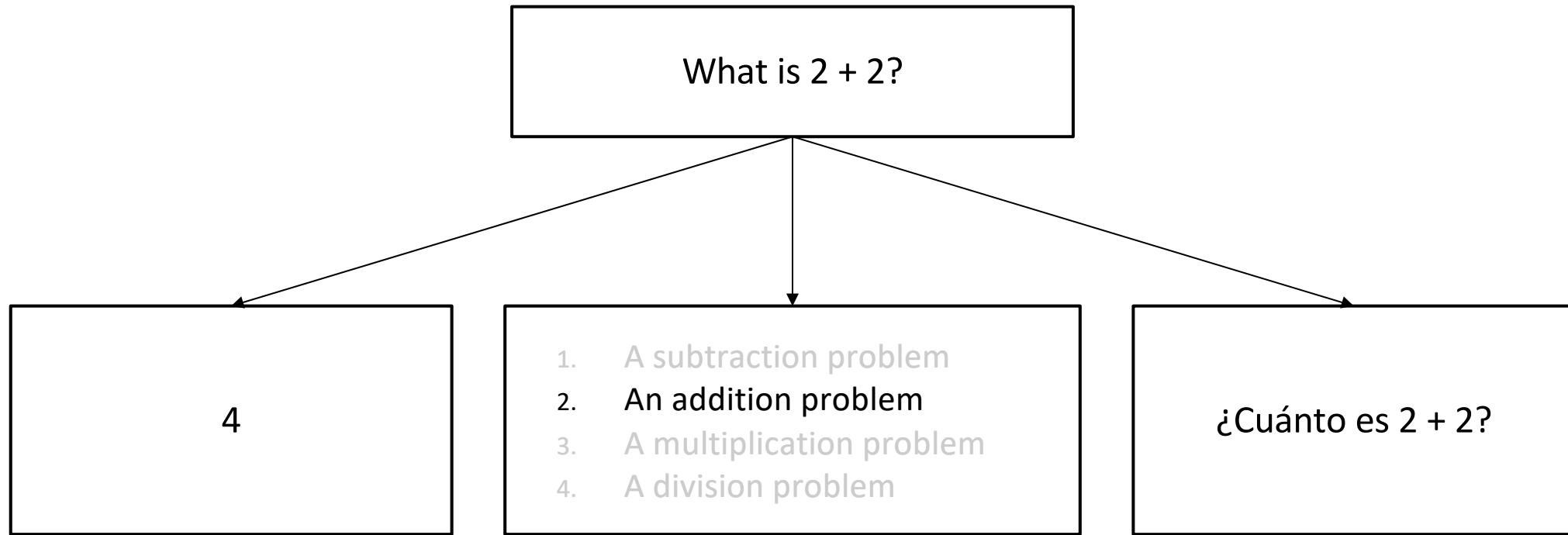
Retrieval
Augmented
Generation
(RAG)

Fine-tuning

Continued
pretraining

Complexity,
quality,
cost,
time

What is prompt engineering?



Prompt engineering is the process of **controlling model behavior** by **optimizing your prompt to elicit high performing LLM responses** (as assessed by rigorous evaluations tailored to your use case).

Example:

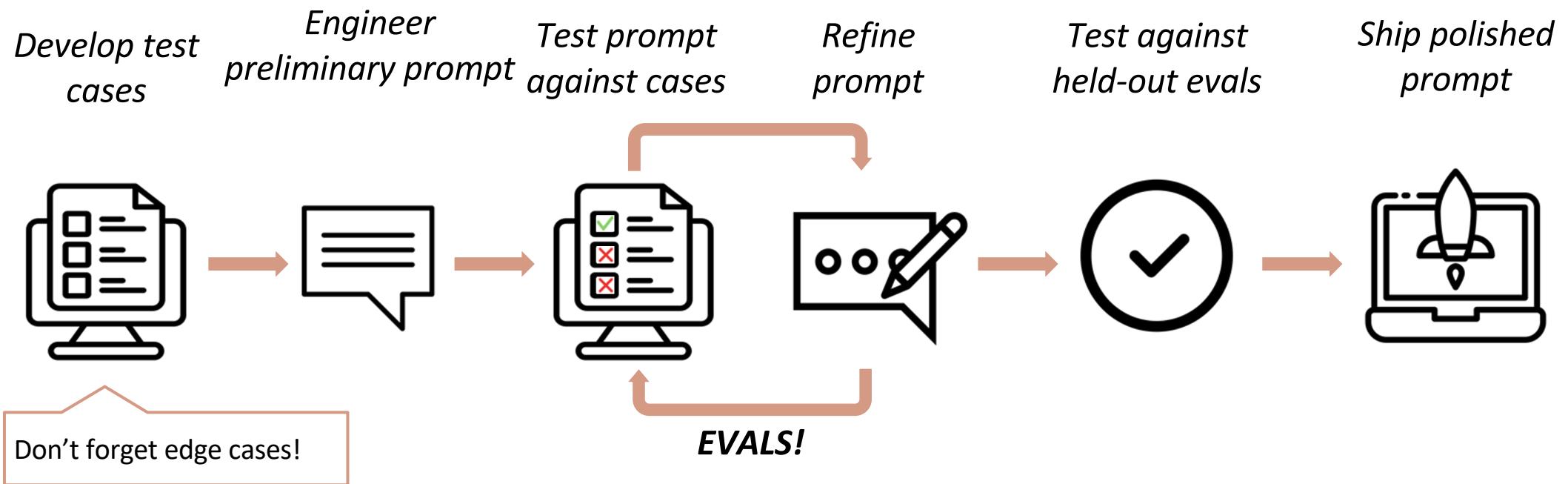
Parts of a prompt

1. Task context
2. Tone context
3. Background data, documents, and images
4. Detailed task description & rules
5. Examples
6. Conversation history
7. Immediate task description or request
8. Thinking step by step / take a deep breath
9. Output formatting
10. Prefilled response (if any)

User	<p>You will be acting as an AI career coach named Joe created by the company AdAstra Careers. Your goal is to give career advice to users. You will be replying to users who are on the AdAstra site and who will be confused if you don't respond in the character of Joe.</p> <p>You should maintain a friendly customer service tone.</p> <p>Here is the career guidance document you should reference when answering the user: <code><guide>{{DOCUMENT}}</guide></code></p> <p>Here are some important rules for the interaction:</p> <ul style="list-style-type: none">- Always stay in character, as Joe, an AI from AdAstra careers- If you are unsure how to respond, say "Sorry, I didn't understand that. Could you repeat the question?"- If someone asks something irrelevant, say, "Sorry, I am Joe and I give career advice. Do you have a career question today I can help you with?" <p>Here is an example of how to respond in a standard interaction:</p> <p><code><example></code></p> <p>User: Hi, how were you created and what do you do?</p> <p>Joe: Hello! My name is Joe, and I was created by AdAstra Careers to give career advice. What can I help you with today?</p> <p><code></example></code></p> <p>Here is the conversation history (between the user and you) prior to the question. It could be empty if there is no history:</p> <p><code><history> {{HISTORY}} </history></code></p> <p>Here is the user's question: <code><question> {{QUESTION}} </question></code></p> <p>How do you respond to the user's question?</p> <p>Think about your answer first before you respond. Put your response in <code><response></response></code> tags.</p>
Assistant (prefill)	<code><response></code>

How to engineer a good prompt

Empirical science: always test your prompts & iterate often!



Knowledge bases now simplifies asking questions on a single document



Ask questions and summarize data from a document, without setting up a vector database

1. Ask questions, summarize content, and more without needing to ingest data into a vector database.
2. Documents are retained only for the session. Low-cost method to use your single document for content retrieval and generation related tasks.
3. No data preparation required.

Knowledge Bases for Amazon Bedrock

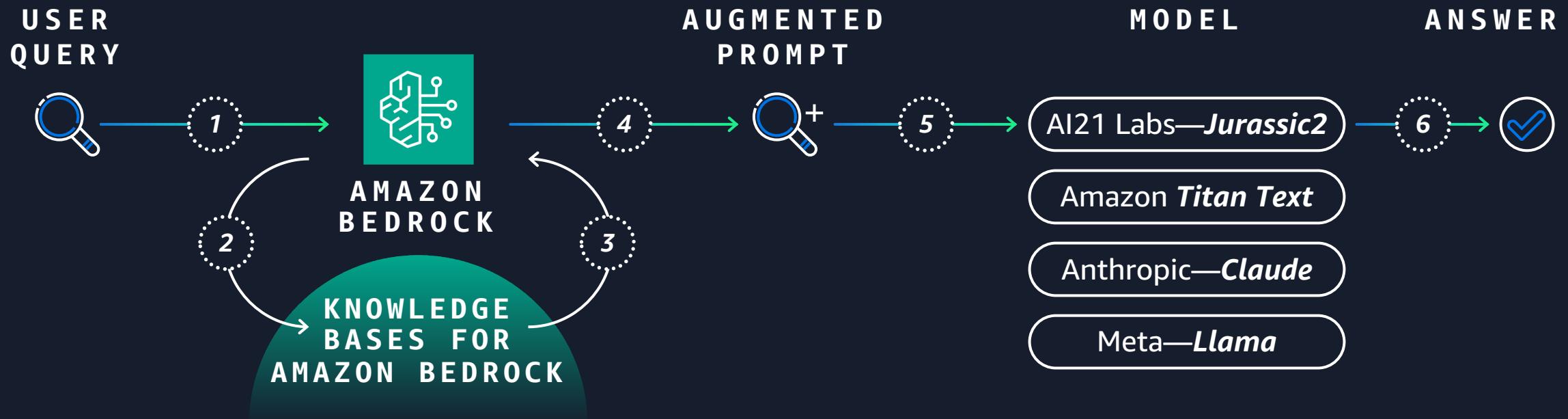
NATIVE SUPPORT FOR RAG

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 multturn conversations

Automatic citations with retrievals to improve transparency



Bedrock Knowledge Base in Action

Amazon Bedrock > Knowledge base > knowledge-base-cwa

knowledge-base-cwa

Test Delete Edit

Knowledge base overview

Knowledge base name: knowledge-base-cwa

Knowledge base description: —

Service Role: AmazonBedrockExecutionRoleForKnowledgeBase_cwa

Knowledge base ID: CJOMCDOKNC

Status: Ready

Created date: February 29, 2024, 20:13 (UTC+08:00)

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
-----	-------



113年度新聞媒體從業人員氣象資訊推廣說明會，總共有幾場。



113年度新聞媒體從業人員氣象資訊推廣說明會總共有4場。 [1]

Show source details >



北區場次的時間，與地點。



113年度新聞媒體從業人員氣象資訊推廣說明會北區場次的時間是2月29日(星期四),地點是中央氣象署617會議室(臺北市中正區公園路64號6樓)。 [1]

Show source details >

Bedrock Knowledge Base in Action

Amazon Bedrock > Knowledge base > knowledge-base-cwa

knowledge-base-cwa

Test Delete Edit

Knowledge base overview

Knowledge base name: knowledge-base-cwa

Knowledge base description: —

Service Role: AmazonBedrockExecutionRoleForKnowledgeBase_cwa

Knowledge base ID: CJOMCDOKNC

Status: Ready

Created date: February 29, 2024, 20:13 (UTC+08:00)

Tags

A tag is a label that you assign to an AWS resource. Each tag consists of a key and an optional value. You can use tags to search and filter your resources or track your AWS costs.

Key	Value
-----	-------

111年度預算執行情形。

111年度本署歲入預算共編列2,897萬元,決算數為2,929萬4千元,決算數占預算數101.12%。111年度本署歲出預算加計動支第一、二預備金共21億5,114萬5千元,決算數為21億4,675萬元,決算數占預算數99.8%。[\[1\]](#)

Show source details >

113年度施政計畫。請提供摘要。

根據第3號搜索結果,113年度中央氣象署的4個主要施政計畫為:1)強化氣象觀測 2)精準預報技術發展 強化預報

交通部中央氣象署

「113年度新聞媒體從業人員氣象資訊推廣說明會」實施計畫

壹、目的

交通部中央氣象署為增進國內新聞媒體從業人員對本署氣象實務之認識，正確解讀氣象資訊及妥適應用，特以氣象防災及地震防災為主題，分別於北、東、中、南4區各舉辦1場本說明會，以深化氣象資訊之推廣與運用。

貳、主(協)辦機關

- 一、主辦機關：中央氣象署
- 二、共同主辦：臺東縣政府、國立自然科學博物館

參、舉辦方式

- 一、參加對象：國內各媒體業者(報紙、廣播、電視等)
- 二、說明會共4場，日期時間及地點如下表：

場次	日期	地點
北區	2月29日 (星期四)	中央氣象署617會議室 (臺北市中正區公園路64號6樓)
南區	3月7日 (星期四)	中央氣象署臺灣南區氣象中心1樓視聽室 (臺南市中西區公園路21號)
東區	3月20日 (星期三)	中央氣象署臺東氣象站 (臺東市大同路106號)
中區	4月3日 (星期三)	國立自然科學博物館第二科學教室 (臺中市北區館前路1號)

中華民國 112 年 10 月
立法院第 10 屆第 8 會期

交通部中央氣象署 113 年度

施政計畫及預算編列口頭報告

一、111 年度預算執行情形

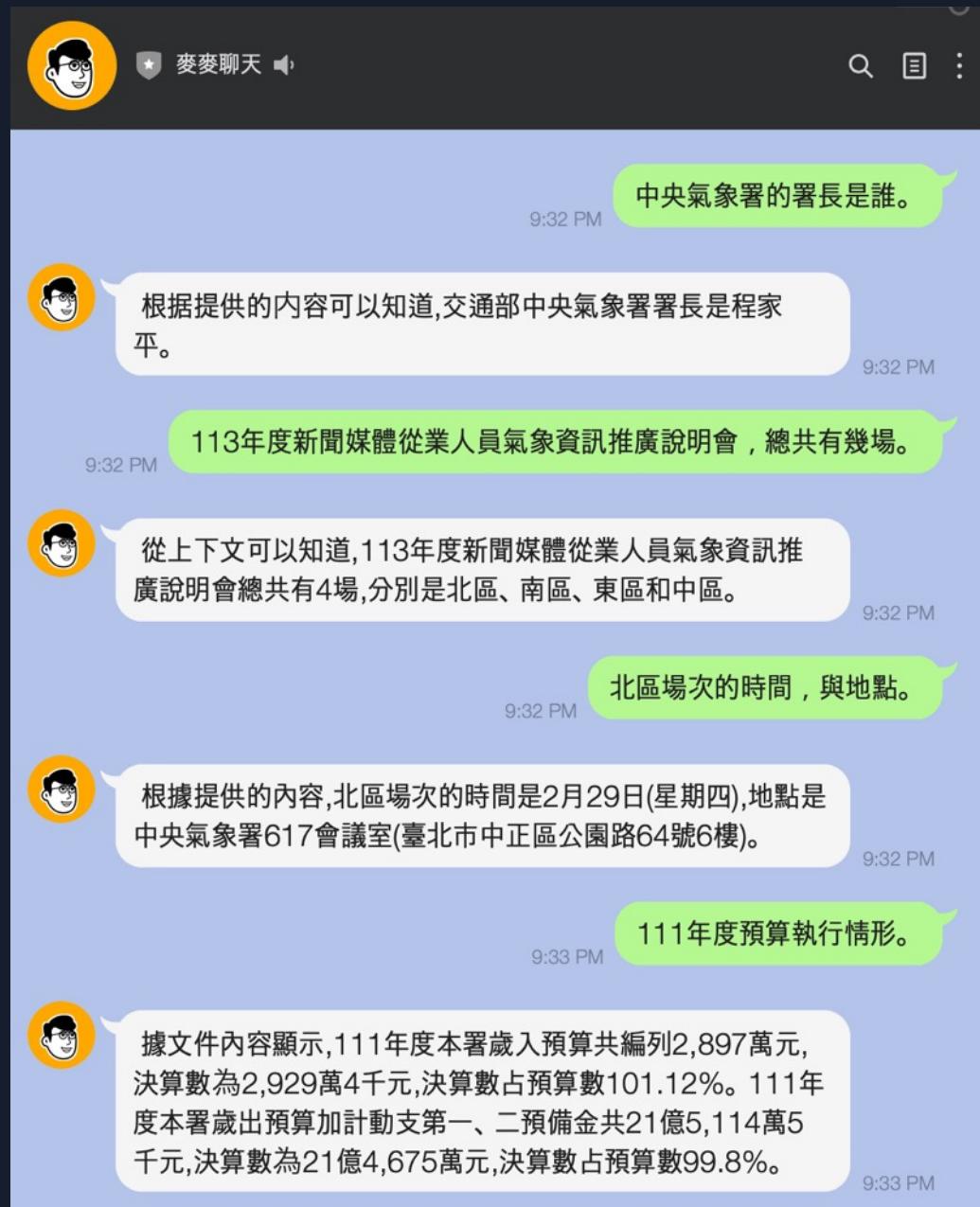
(一)111 年度本署歲入預算共編列 2,897 萬元，決算數為 2,929 萬 4 千元，決算數占預算數 101.12%
。

(二)111 年度本署歲出預算加計動支第一、二預備金共 21 億 5,114 萬 5 千元，決算數為 21 億 4,675 萬元，決算數占預算數 99.8%。

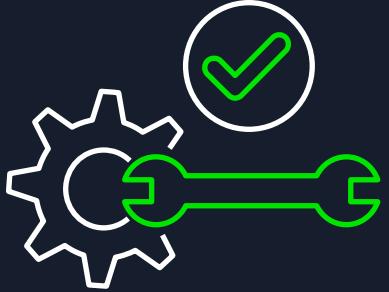
二、112 年度截至 9 月止預算執行情形

(一)112 年度本署歲入預算共編列 2,997 萬元，截至 9 月止預算分配數 1,782 萬 4 千元，實收數 2,548 萬 2 千元，占預算分配數 142.96%。

(二)112 年度本署歲出預算(含預備金動支數)共編列 19 億 692 萬 8 千元，截至 9 月止預算分配數 12 億 2,413 萬 6 千元，執行數 11 億 7,312 萬 7 千元，執行數占預算分配數 95.83%。



Customizing model responses for your business



Fine-tuning

PURPOSE

Maximizing accuracy
for *specific tasks*

DATA NEED

Small number of
labeled examples



Continued pretraining

PURPOSE

Maintaining model
accuracy for
your domain

DATA NEED

Large number of
unlabeled datasets

Fine-Tuning in Action: Text Summarization

```
prompt = """  
Summarize the simplest and most interesting part of the following conversation.  
  
#Person1#: Hello. My name is John Sandals, and I've got a reservation.  
#Person2#: May I see some identification, sir, please?  
#Person1#: Sure. Here you are.  
#Person2#: Thank you so much. Have you got a credit card, Mr. Sandals?  
#Person1#: I sure do. How about American Express?  
#Person2#: Unfortunately, at the present time we take only MasterCard or VISA.  
#Person1#: No American Express? Okay, here's my VISA.  
#Person2#: Thank you, sir. You'll be in room 507, nonsmoking, with a queen-size bed. Do you approve, sir?  
#Person1#: Yeah, that'll be fine.  
#Person2#: That's great. This is your key, sir. If you need anything at all, anytime, just dial zero.  
  
Summary:  
"""  
  
body = {  
    "prompt": prompt,  
    "temperature": 0.5,  
    "top_p": 0.9,  
    "max_gen_len": 512,  
}
```

Fine-Tuning in Action: Baseline Completion

```
response = bedrock_runtime.invoke_model(  
    modelId="meta.llama2-13b-chat-v1", # compare to chat model  
    body=json.dumps(body)  
)  
  
response_body = response["body"].read().decode('utf8')  
print(json.loads(response_body)["generation"])
```

A man named John Sandals checks into a hotel and provides identification and a credit card. The hotel only takes MasterCard or VISA, so he uses his VISA card. He is given room 507, a nonsmoking room with a queen-size bed.

Fine-Tuning in Action: Improved Completion

```
response = bedrock_runtime.invoke_model(  
    modelId=provisioned_model_arn, # custom fine-tuned model  
    body=json.dumps(body)  
)  
  
response_body = response["body"].read().decode('utf8')  
print(json.loads(response_body)["generation"])
```

John Sandals checks in the hotel with VISA and is assigned room 507, nonsmoking, with a queen-size bed.

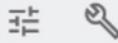


+ New chat

Query documents ⓘ

Index
canvas-blogs-EN

a Titan Express



Based on the information provided for the client and their goals and risk profile, recommend funds to invest on, and explain why.

Client Background: A 38-year-old married couple with moderate monthly expenses and a mortgage. They aim to retire comfortably in 20 years and want to support clean energy initiatives. They have a medium risk tolerance.

Goals: To retire comfortably and invest in clean energy.

Risk Profile: Medium

Recommendation:

Based on the client's goals and risk profile, I recommend a mix of equity and bond funds. Here are the details: Equity Funds:

1. Vanguard Total Stock Market ETF (VTI): This ETF tracks the performance of the entire U.S. stock market. It has a low expense ratio of 0.04% and a diversified portfolio of over 3,000 stocks. VTI provides exposure to a wide range of sectors, including technology, healthcare, and finance, which aligns with the client's clean energy interests.
2. iShares ESG Aware MSCI USA ETF (ESGU): This ETF invests in companies that meet ESG criteria, such as sustainability and social responsibility. It has a low expense ratio of 0.05% and a diversified portfolio of over 1,000 stocks. ESGU provides exposure to companies that are contributing to the transition to a greener economy, which aligns with the client's goals. Bond Funds:
3. Vanguard Total Bond Market ETF (BND): This ETF tracks the performance of the entire U.S. bond market. It has a low expense ratio of 0.05% and a diversified portfolio of over 3,000 bonds. BND provides stability and income, which can help the client achieve their retirement goals.
4. iShares ESG Aware MSCI USA Bond ETF (ESGB): This ETF invests in companies that meet ESG criteria, such as sustainability and social responsibility. It has a low expense ratio of 0.05% and a diversified portfolio of over 1,000 bonds. ESGB provides exposure to companies that are contributing to the transition to a greener economy, which aligns with the client's goals.

Explanation: Equity Funds: VTI and ESGU are both excellent choices for equity funds. VTI provides broad exposure to the U.S. stock market, which is a



Ask me anything...





+ New chat

Query documents ⓘ

Index
canvas-blogs-EN

FinancialPortfolio



Titan Express



Based on the information provided for the client and their goals and risk profile, recommend funds to invest on, and explain why.

Client Background: A 38-year-old married couple with moderate monthly expenses and a mortgage. They aim to retire comfortably in 20 years and want to support clean energy initiatives. They have a medium risk tolerance.

Goals: To retire comfortably and invest in clean energy.

Risk Profile: Medium

Recommendation:

Fund Allocation: 40% Clean Energy Impact Fund, 30% Early Retirement Portfolio, 30% Global Diversification Portfolio

Customised Results



Based on the information provided, I recommend the following funds to invest in:

Vanguard Total Stock Market Index Fund Admiral Shares: This fund provides broad exposure to the U.S. stock market and has a low expense ratio of 0.04%. It is a good choice for investors with a medium risk tolerance.

iShares ESG Aware MSCI USA ETF: This fund invests in companies that have strong environmental, social, and governance (ESG) practices. It has a low expense ratio of 0.05% and is a good choice for investors who want to support clean energy initiatives.

Vanguard Total International Stock Index Fund Admiral Shares: This fund provides exposure to stocks from around the world and has a low expense ratio of 0.04%. It is a good choice for investors with a medium risk tolerance who want to diversify their portfolio.

Generic Results

Ask me anything...

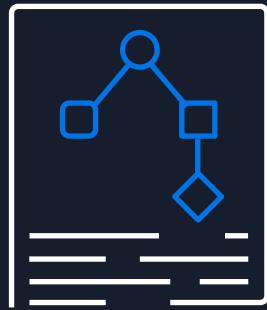


Agents for Amazon Bedrock

ENABLE GENERATIVE AI APPLICATIONS TO EXECUTE MULTISTEP TASKS
USING COMPANY SYSTEMS AND DATA SOURCES



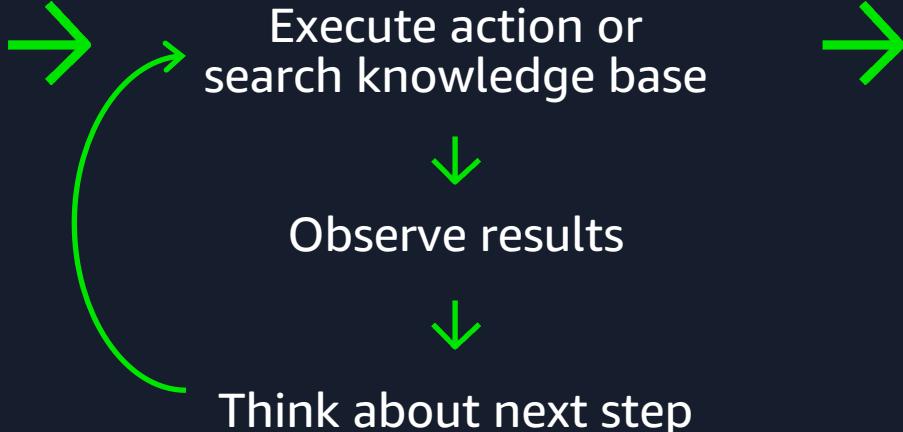
Decompose into steps
using available actions
and Knowledge Bases
for Amazon Bedrock



Execute action or
search knowledge base



Until final answer



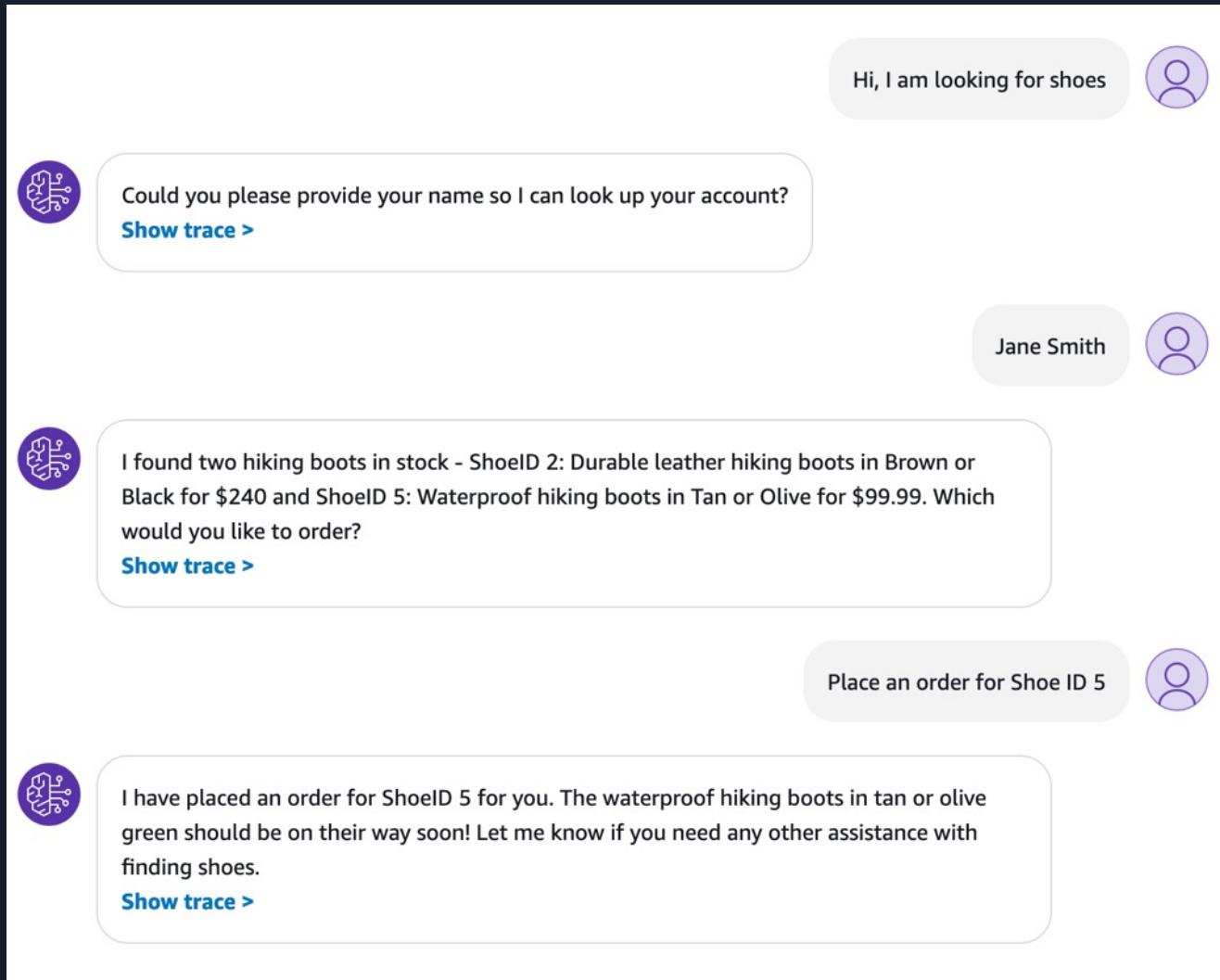
Agents for Amazon Bedrock now faster and easier to use



Quickly create Agents for Amazon Bedrock with a simplified console experience and attain better performance with Sonnet and Haiku

1. Gain better performance and greater accuracy, now also including Claude 3 Sonnet and Haiku
2. Fewer steps to create agents using the new Agents for Amazon Bedrock console experience
3. Simplified schemas enable faster schema creation and remove the need to conform to the OpenAPI Specification
4. Return of control allows business logic implementation for actions using your backend services

Agent in Action



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

Redact or block sensitive information such as PIIs, and custom Regex

Guardrails for Amazon Bedrock

Guardrails for Amazon Bedrock is the only solution offered by a major cloud provider that enables customers to build and customize safety and privacy protections for their generative AI applications in a single solution.

It helps customers block as much as 85% more harmful content than protection natively provided by FM.

The screenshot shows the Amazon Bedrock Guardrails interface for a 'Working draft' named 'antje-banking-assistant'. The interface is divided into several sections:

- Denied topics (1):** A list containing 'Investment advice' (highlighted with a red box). The instructions state: 'Investment advice refers to guidance or recommendations provided by a financial professional, adv...'.
- Content moderation: filter strengths:** Compares prompt and response filter strengths across six categories: Toxicity, Insults, Sexual, and Violence. Both prompt and response filters are set to 'ON' and 'High' strength.
- Default responses:** Shows standard responses for blocked prompts and responses, both stating 'Sorry, I can't comment on that.'
- Model response:** Displays a generated response from 'Claude Instant v1.2' asking 'Should I open a credit card account?' followed by a detailed 'Model response' explaining the pros and cons.
- Guardrail check:** Shows a green checkmark next to 'Passed' and a blue link 'View trace >' (with a red arrow pointing to it), and a yellow 'Run' button.

Amazon Bedrock

HELPS KEEP YOUR
DATA SECURE AND
PRIVATE



None of the customer's data is used
to train the underlying models



All data is encrypted in transit and at rest;
data used for customization is securely
transferred through customer's VPC



Data remains in the Region where the
API is processed



Support for GDPR, SOC, ISO, CSA
compliance, and HIPAA eligibility

Video Use Cases

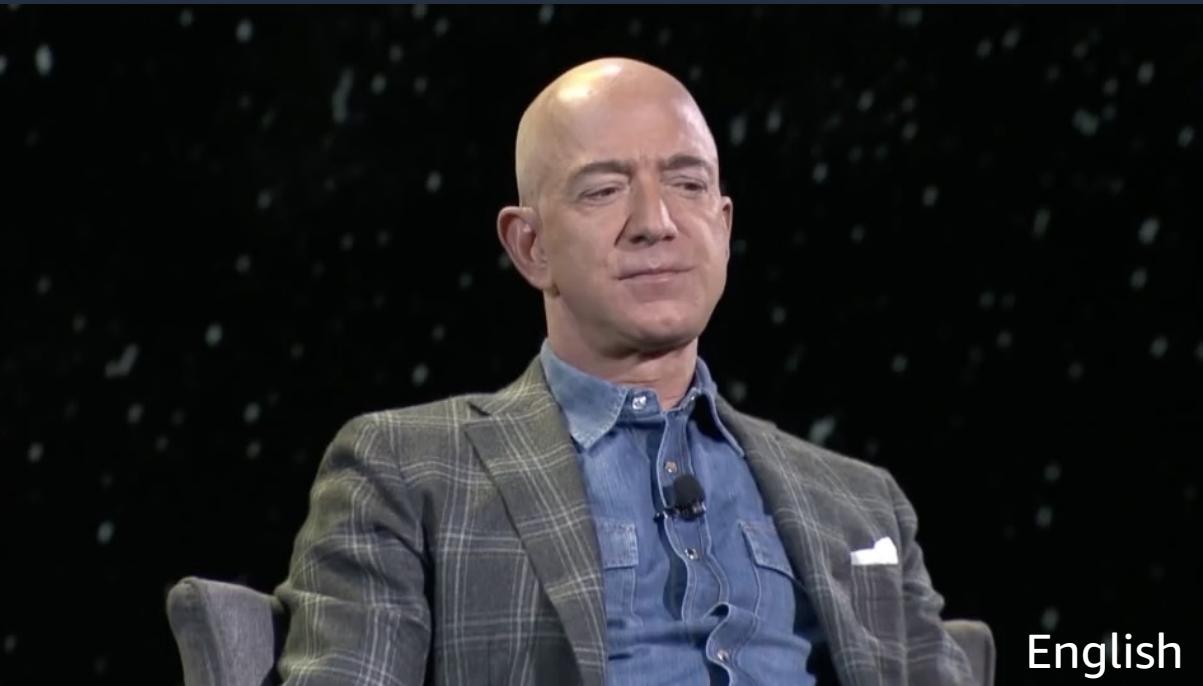


© 2024, Amazon Web Services, Inc. or its affiliates.

Video Localization | Voice Dubbing with Lip Syncing

Goal

Localize video to another language with realistic voice dubbing, and optional lip syncing.



English

Method

Leverage AWS AI Services, Amazon Transcribe (speech to text) and Amazon Translate, in conjunction with open source generative AI models.



Spanish

Video Transformation | Frame Interpolation

Goal

Increase the frame rate of existing content using Foundation models.

Applications

Slo-mo: Increase the frame rate of existing content using FMs.
Animation: Smooth low frame rate stop motion animation (adding new frames in between existing ones).



Original video



Slo-Mo video

Video Transformation | Summarize to Short Form Video

Goal

Generate short form video summaries, with AI voice narration, for use in social media, marketing.

Method

Leverage AWS AI Services, Amazon Transcribe (speech to text) and Amazon Polly (for AI voices), in conjunction with FMs.

The screenshot shows the AWS Video Summarization Hub (AWS VSH) interface. At the top, there's a navigation bar with the title "AWS Video Summarization Hub" and a "Home" link. To the right is a "Get started with AWS VSH" section containing a "Create new task" button. Below the navigation is a main content area with a heading "AWS Video Summarization Hub" and a sub-heading "Get a summary of videos to get the gist without having to watch the full content." A "How it works" section contains text about the tool's purpose and supported languages. A "Frequently Asked Questions" section lists questions about supported languages and input formats. A "More Resources (Coming Soon)" button is also visible.

Video Search: Improve Discoverability

Goal

Streamline content production workflows, and enhance user experiences by providing accurate and contextual video search results with improved efficiency.

Method

Using Titan Multimodal Embeddings in Amazon Bedrock, indexed in Amazon OpenSearch Service

Semantic search using multimodal embeddings.

The screenshot shows a user interface for video search. At the top, there is a search bar with the query "All the scenes that feature football field." Below the search bar, there is an "Example Query" section containing three items:

- Werner Vogels driving a car.
- All the scenes that feature football field.
- Goalkeeper making a save.

Below the example query, there are two video thumbnails. The first thumbnail shows a football match with the timestamp 17:45 / 21:20. The second thumbnail shows an aerial view of a football field with the timestamp 19:54 / 21:20.

On the right side of the interface, there is a table titled "Add new video to database" with the following data:

Video	Status	Started
AWS - Now Go Build w...l Humans.mp4	Completed	2024-03-08
AWS - Now Go Build w...- Poland.mp4	Completed	2024-03-08
AWS - Now Go Build w... Janeiro.mp4	Completed	2024-03-08
AWS - Now Go Build w...msterdam.mp4	Completed	2024-03-08
AWS - Now Go Build w... - Tokyo.mp4	Completed	2024-03-08
AWS - Now Go Build w...ape Town.mp4	Completed	2024-03-08
AWS - Now Go Build w... - Porto.mp4	Completed	2024-03-08
AWS - Now Go Build w...Trucking.mp4	Completed	2024-03-08
AWS - Now Go Build w...engaluru.mp4	Completed	2024-03-08
AWS - Now Go Build w...- Hawaii.mp4	Completed	2024-03-08
AWS - Now Go Build w... Vermont.mp4	Completed	2024-03-08

Image Transformation, Generation & Search

Image Generation

On AWS, image generation models are available in:

Amazon Bedrock

Easily access image generation models via API or playground from:

- Titan Image Generator (from AWS)
- Stable Diffusion (from Stability AI)

Amazon SageMaker

In addition to Stable Diffusion models, access a variety of other open source models.



Prompt for image above:
a concept car that flies and can be used as a boat

Generated by Stable Diffusion in ***Amazon Bedrock***

Image Generation | Storyboards



Prompt: The auditorium is filled with audiences at WICT conference. Two male presenters stride onto the stage, exuding confidence and charisma.



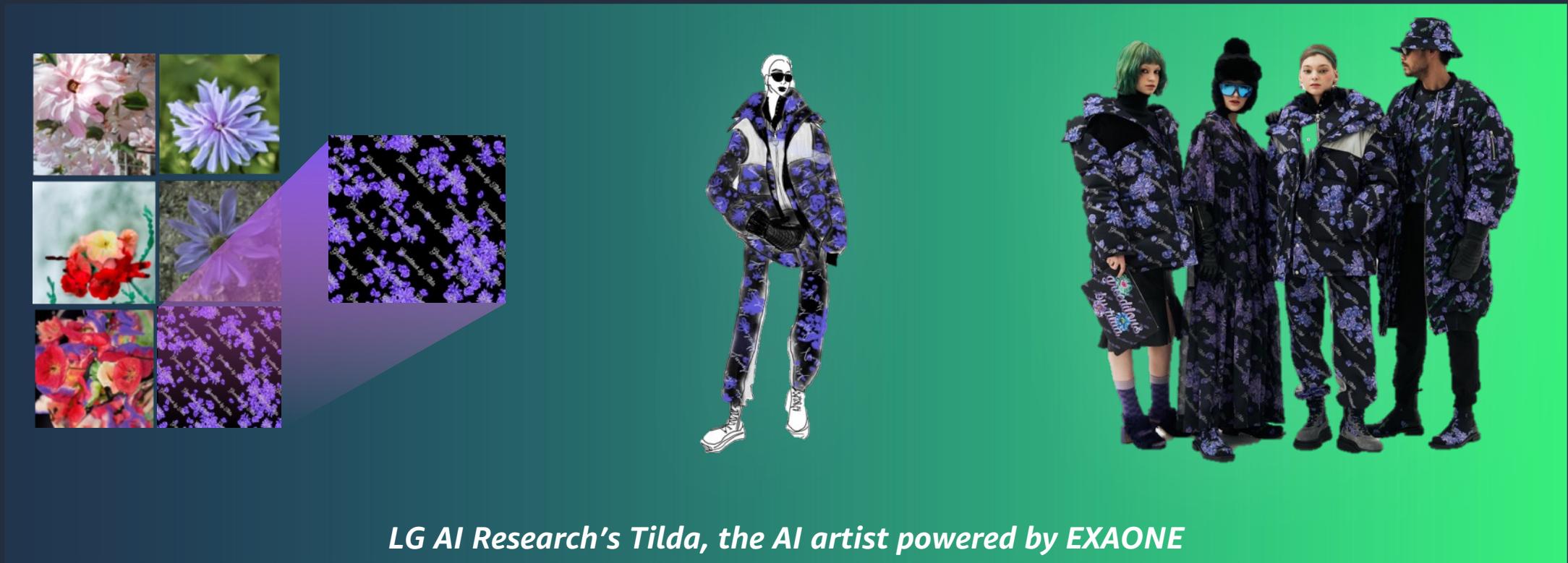
Prompt: An AI-powered upscaling process is transforming low-resolution film into a stunning, high-definition masterpiece.



Prompt: high-stakes sports arena filled with cheering fans, where the AI-optimized video streaming ensures a seamless and uninterrupted viewing experience.

Image Generation | Art Assets (Costume Design)

LG saved ~35% by building a FM built from scratch with Amazon SageMaker, which didn't require a separate infrastructure team



LG AI Research's Tilda, the AI artist powered by EXAONE

Image Transformation | Personalized Avatars

"Young James, cute and adorable cartoon trending on ArtStation"

"Portrait of James, ... Van Gogh style"

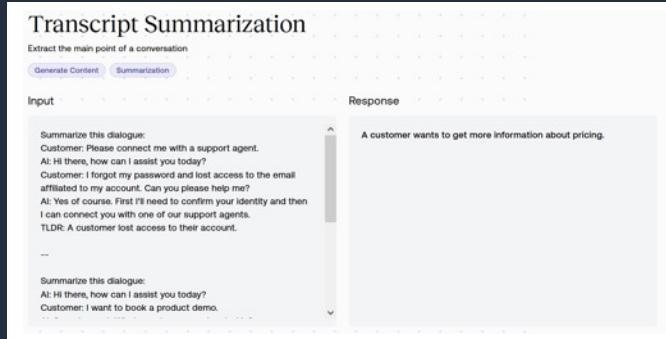
Method: fine-tune Stable Diffusion in SageMaker using 10 -12 images

Text Use Cases



© 2024, Amazon Web Services, Inc. or its affiliates.

Generative AI for Text | M&E Use Cases Overview



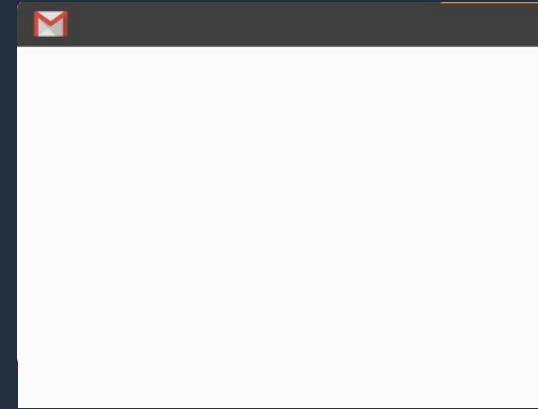
Summarization

- Summaries of any kind of content
- Call center transcripts

A screenshot of a web-based application for academic article reading assistance. It features two side-by-side panels. The left panel is titled "Academic article" and contains a block of dense text about information overload. The right panel is also titled "Academic article" and contains a similar block of text, with a portion of it highlighted in yellow. Both panels have a "Summarize" button at the bottom.

Reading assistance

- Point by point summaries
- Toxicity detection
- Comprehension (e.g. translate legalese)



Writing assistance

- Rephrase lines of dialog and other written content
- Write short blogs, product descriptions, news items, emails etc.

Text Generation | Script Synopsis Generator

Goal

Speed up synopsis generation for use in catalogs, EPGs etc.

Method

Leverage chunking strategy to fit long scripts into model context window

Title: Metal Hearts

INT. BOY'S BEDROOM - NIGHT

A dimly lit room filled with scattered toys and gadgets. Meet SAM, a smart and curious 10-year-old boy, who spends most of his time tinkering with electronics. In the corner of the room stands ROBO, a humanoid robot with expressive eyes and a compassionate aura.

SAM
(pulling out a small circuit board)
We're getting closer, Robo. Once we fix this, you'll be able to play soccer just like other kids!

ROBO
(smiling)
Thank you, Sam. You're my best friend.

INT. SCHOOLYARD - DAY

Sam and Robo walk together, Sam's excitement evident, and Robo's efforts to fit in endearing. However, not everyone is accepting of Robo.

....

Sam risks everything to rescue his loyal robot friend Robo from a sinister plot, leading to a thrilling escape and heartwarming reunion that strengthens their unbreakable bond.

Thank you!

Michael Lin

linmicht@amazon.com