



# AWS SF Learning Days

## Generative AI on Healthcare

Yuan Yao (he/him)

Senior Solutions Architect  
AWS Non-Profit Healthcare

# Agenda

Introduction to generative AI

Healthcare use cases

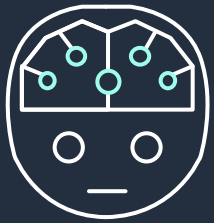
Data is your differentiator

Generative AI with AWS

What's next?

# Introduction to generative AI

# AI/ML/GenAI hierarchy



## Artificial intelligence (AI)

**Any technique that allows computers to mimic human intelligence using logic, if-then statements, and machine learning**



## Machine learning (ML)

**A subset of AI that uses machines to search for patterns in data to build logic models automatically**



## Deep learning (DL)

**A subset of ML composed of deeply multi-layered neural networks that perform tasks like speech and image recognition**

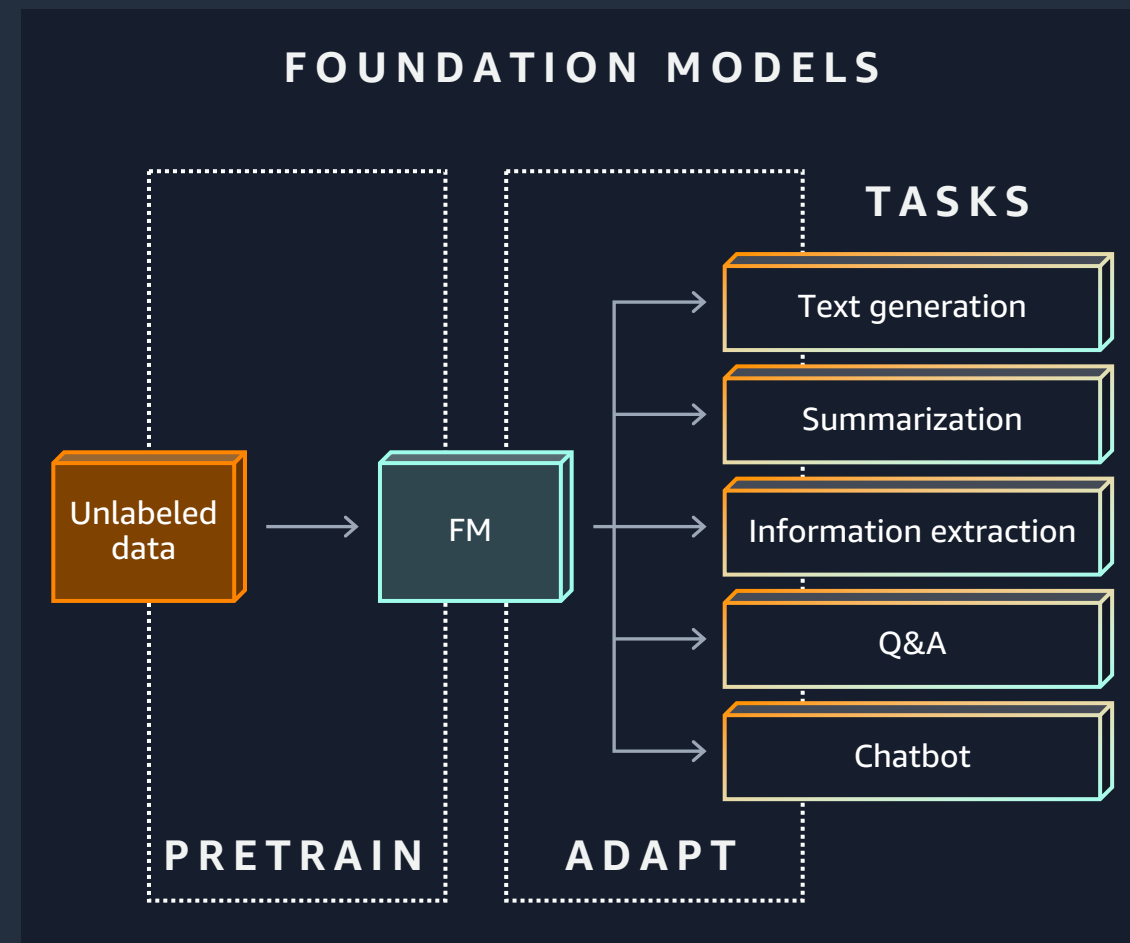
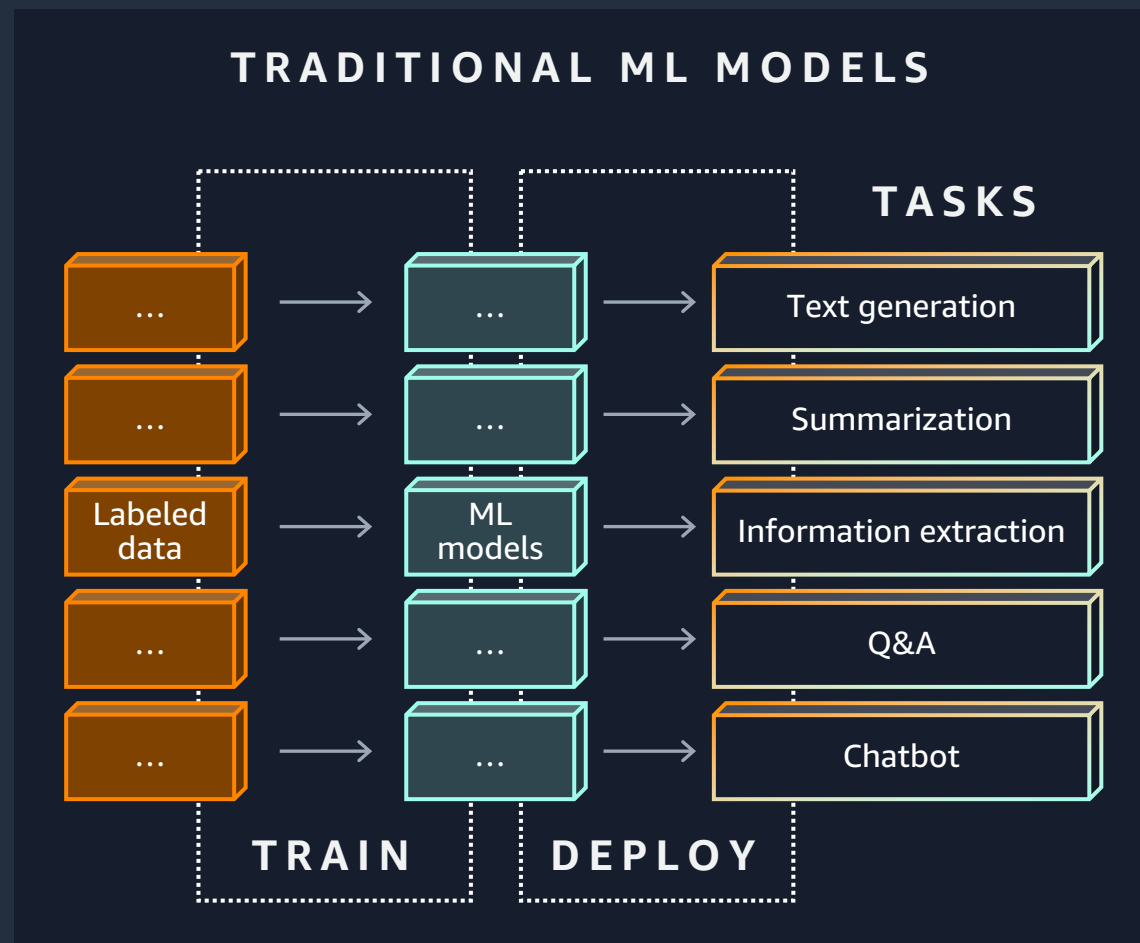


## Generative AI

**Powered by large models that are pretrained on vast corpora of data and commonly referred to as foundation models (FMs)**



# Why foundation models?



# Foundation model use cases



**Productivity**  
Text generation



**Chat**  
Virtual assistant



**Summarization**  
Text extraction



**Search**



**Code generation**



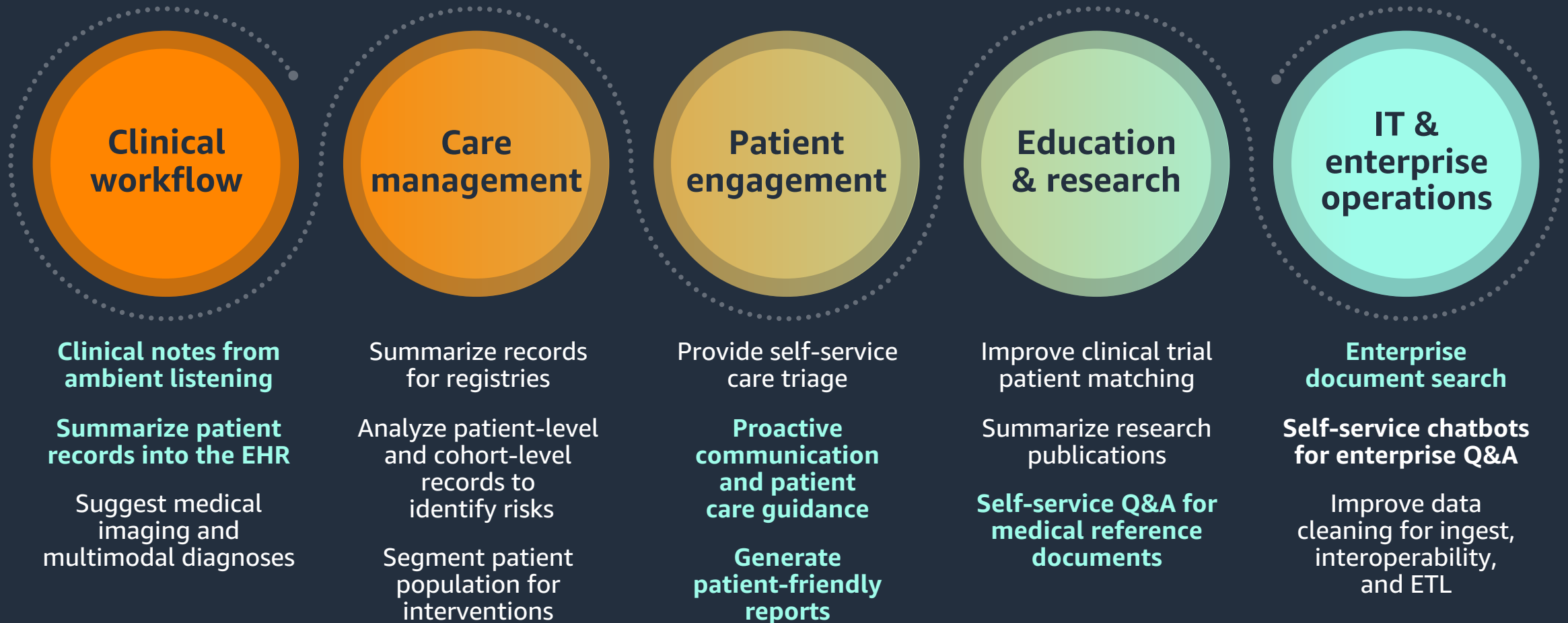
**Image generation**



**Image classification**

# Healthcare use cases

# Top generative AI healthcare use cases



# Data is your differentiator





# Generative AI Application





Generative AI  
Application

# Data Foundation

STORAGE

GOVERNANCE  
& COMPLIANCE

DATABASES,  
ANALYTICS,  
& DATA LAKES

DATA  
INTEGRATION

# Your data is the **differentiator**



Generic  
generative AI



Generative AI that  
knows your business  
and your customers



# Generative AI with AWS



# Generative AI Stack

## APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs



AWS  
HealthScribe



Amazon  
Transcribe  
Medical



Amazon  
Transcribe  
Medical



Amazon Q  
Business,  
Developer,  
Connect

## TOOLS TO BUILD WITH LLMs AND OTHER FMs



Amazon Bedrock

Guardrails | Agents | Customization Capabilities

## INFRASTRUCTURE FOR FM TRAINING AND INFERENCE



GPUs



Trainium



Inferentia



SageMaker



UltraClusters



EFA



EC2 Capacity Blocks



Nitro



Neuron

# Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs

TOOLS TO BUILD WITH LLMs AND OTHER FMs

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE



GPUs



Trainium



Inferentia



SageMaker



UltraClusters



EFA



EC2 Capacity Blocks



Nitro



Neuron



# Amazon SageMaker

Build, train, and deploy ML models  
at scale, including FMs

Access the latest and publicly available FMs

Build FMs from scratch

Customize FMs

Run inference

Implement FMOps and governance

# Amazon SageMaker JumpStart

ML hub with foundation models, built-in algorithms, and prebuilt ML solutions that you can deploy with just a few clicks



## Machine learning hub

Browse through 400+ built-in algorithms with pretrained foundation models, solutions, and example notebooks



## Pre-built training and inference scripts

Compatible with SageMaker and configurable with custom dataset



## UI as well as API-based

Use the user interface for single click model deployment or API for the Python SDK-based workflow



## Notebooks with examples

Jump into notebooks to use selected model with examples to guide you through the entire ML workflow



## Share and collaborate within your organization

Share models and notebooks with others within your organization, and allow them to train with their own data or deploy as-is for inferencing

# Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs

TOOLS TO BUILD WITH LLMs AND OTHER FMs



Amazon Bedrock

Guardrails | Agents | Customization Capabilities

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE

# 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

Helps keep your data  
secure and private



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

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



# Amazon Bedrock

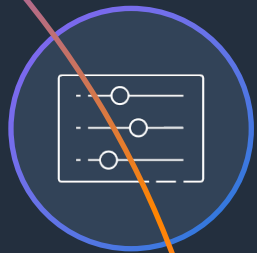
Broad choice of models

AI21 labs	amazon	ANTHROPIC	cohere	Meta	MISTRAL AI	stability.ai
Contextual answers, summarization, paraphrasing	Text summarization, generation, Q&A, search, image generation	Summarization, complex reasoning, writing, coding	Text generation, search, classification	Q&A and reading comprehension	Text summarization, Q&A, text classification, text completion, code generation	High-quality images and art
Jurassic-2 Ultra	Amazon Titan Text Lite	Claude 3 Opus	Command	Llama 3 8B	Mistral Large	Stable Diffusion XL1.0
Jurassic-2 Mid	Amazon Titan Text Express	Claude 3 Sonnet	Command Light	Llama 3 70B	Mistral 7B	Stable Diffusion XL 0.8
	Amazon Titan Text Embeddings	Claude 3 Haiku	Embed English	Llama 2 13B	Mixtral 8x7B	
	Amazon Titan Text Embeddings V2	Claude 2.1	Embed Multilingual	Llama 2 70B		
	Amazon Titan Multimodal Embeddings	Claude 2	Command R+ (Coming Soon)			
	Amazon Titan Image Generator	Claude Instant	Command R (Coming Soon)			



# Guardrails for Amazon Bedrock

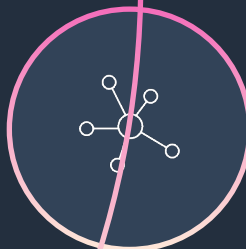
Implement safeguards customized to your application requirements and responsible AI policies



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 PII, and custom Regex.

# 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



Observe results



Think about next step



Until final  
answer

# Knowledge bases for Amazon Bedrock

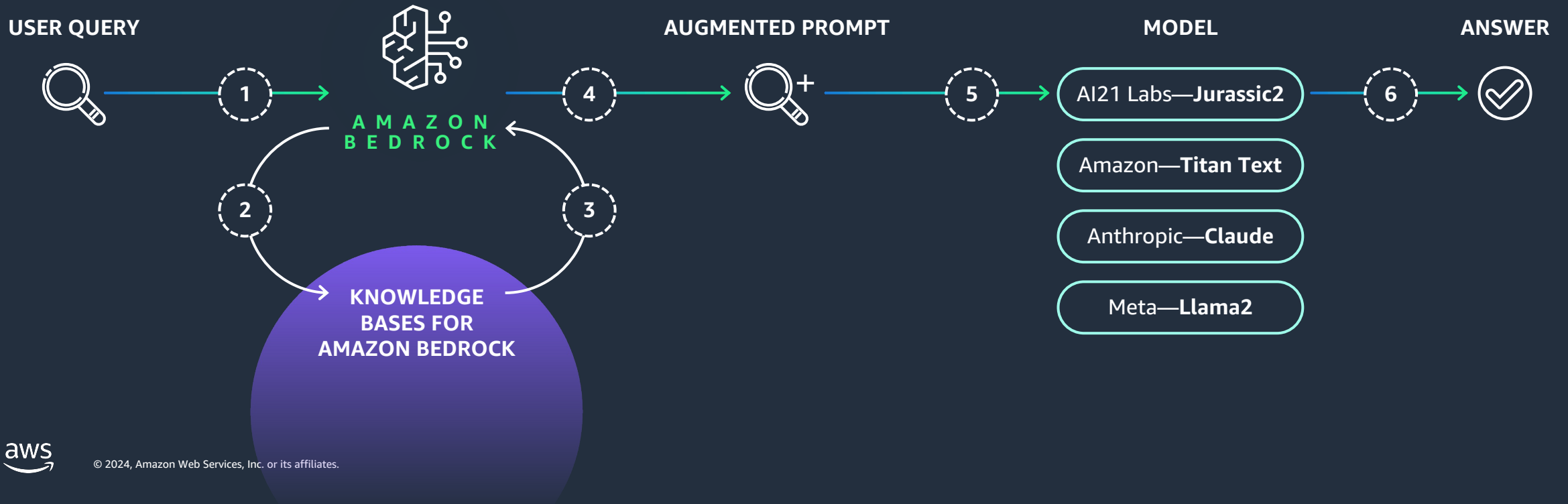
## NATIVE SUPPORT FOR RETRIEVAL AUGMENTED GENERATION (RAG)

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

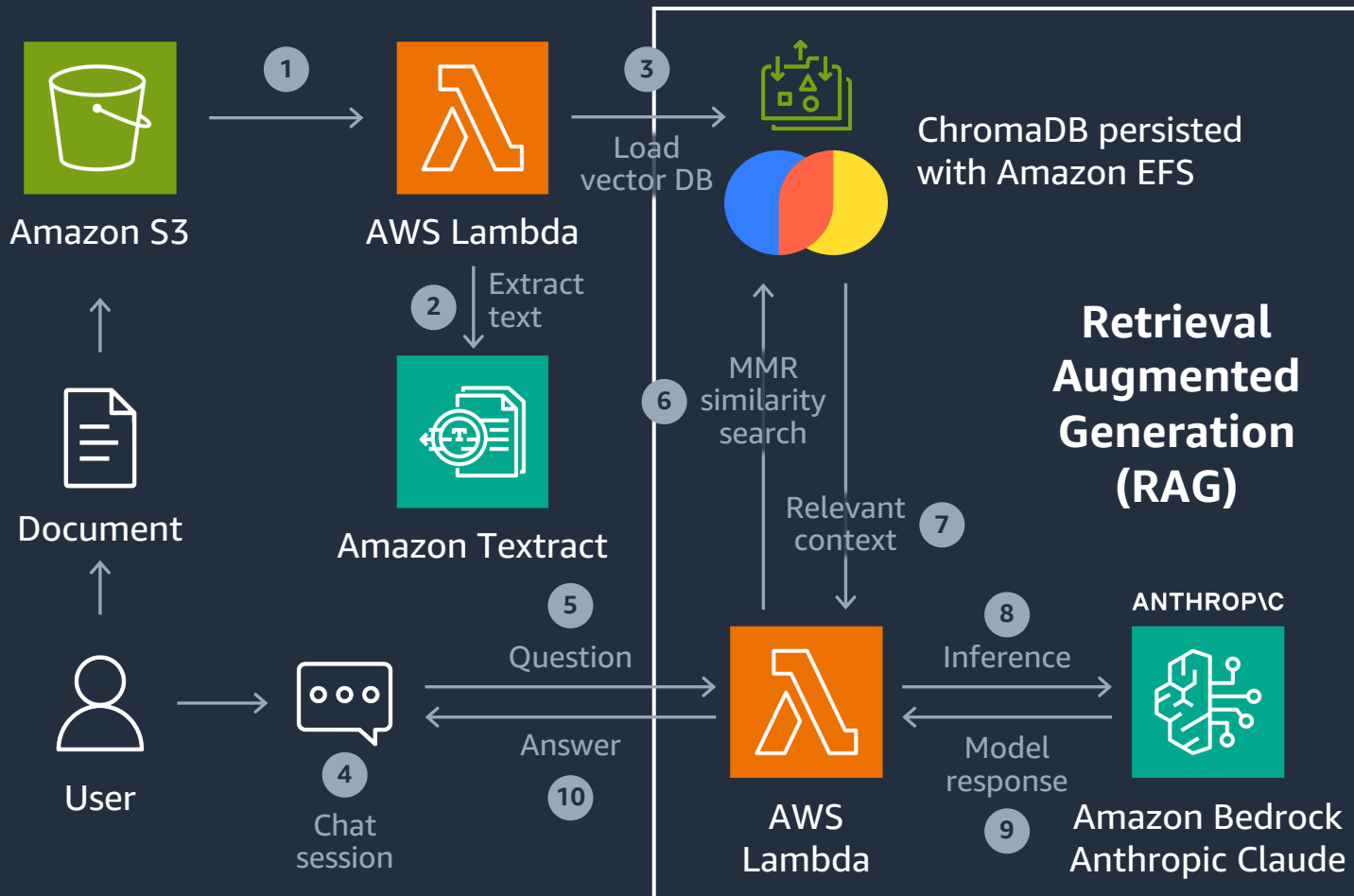
Fully managed RAG workflow including ingestion, retrieval, and augmentation


Built-in session context management for multi-turn conversations

Automatic citations with retrievals to improve transparency



# Intelligent healthcare document processing




 **@naqikhan**  
What is the patient's chief complaint and how long has he been experiencing it?  
Just now

**@DocumentAssistant**

The patient's chief complaint is his stomach problems specifically epigastric abdominal pain described as knowing and burning which lasts one to two hours intermittently the pain has been progressively getting worse the patient also reports nausea fatigue through the past two weeks and bloating after eating. These symptoms have been occurring for about two months as the document states "stomach problems since two months ago".

Just now

 **@naqikhan**  
What is the patient's social history including his smoking habits diet and physical activities?  
Just now

**@DocumentAssistant**

The patient's social history according to the document includes smoking since the age of 15 currently at a rate of 1/2 to one pack per day (PPD). He does not use alcohol recently and denies illicit drug use. His occupation involves high elevation construction work. The patient's diet primarily consists of fast food. Regarding physical activities, he exercises 3-4 times per week, although he stopped his routine two weeks ago.

Just now

# Generative AI Stack

## APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs



AWS  
HealthScribe



Amazon  
Transcribe  
Medical



Amazon  
Comprehend  
Medical



Amazon Q  
Business,  
Developer,  
Connect

## TOOLS TO BUILD WITH LLMs AND OTHER FMs

## INFRASTRUCTURE FOR FM TRAINING AND INFERENCE

# Use case: conversational search

### Intranet Search

Your recent searches ▾ Not finding relevant r

SEARCH IN:

Everything (21)

Wiki (17)

Email List Archive (3)

Show more...

REFINE:

CATEGORIES

Service (1)

Team (1)

CREATOR

admin (1)

abcde (1)

it (1)

corp (1)

Displaying results 1 - 10 of 21

**IT\_Support\_Training\_Program.Web**

18px"%)1 Linux desktop \* (% style="fd  
style="font-size:13.5px"%)\*\*Afternoon

[https://w.co.com/bin/view/IT\\_Support](https://w.co.com/bin/view/IT_Support)

**Com\_Support\_Wiki.Web**

Communication has three differ  
Com sign-in flow>>url:https://c  
<https://w.co.com/bin/view/Com>

**OperationalBestPractices.EventM**

as a basic service. It is owned by [https  
\*Dial into both calls: If you're at your d  
<https://w.co.com/bin/view/BestPracti>

**Corp\_Wiki\_Pending.Web**

Light~",sans-serif; font-size:9pt; heigh  
Videoconference **desk** issues from Reg  
<https://w.co.com/bin/view/Corp/Pend>

### RESULTS PAGE

Kendra's suggested answer

**1st floor**

... our IT help **desk**, deskside, which are located all around our buildings and open for **support** at mos  
hours. The one in **Seattle** is on the **1st floor** and is open from 12:30 to 5 p.m. daily.

WEB | <https://w.organization.com/bin/view/ILDRTEST/>

Frequently asked questions

[Where do I get IT help?](#)

[What are the IT support hours?](#)

[Where can I get IT help corporate campus?](#)

### Kendra ChatBot

Welcome to Kendra ChatBot, please ask a question.

who can audit Marketing Affiliate based on the agreements?

Based on the search results, it looks like Equidata has the right to audit the Marketing Affiliate's marketing practices and activities according to the agreements between them.[source](#)

summarize the responsibilities of the Marketing Affiliate

The main responsibilities of the Marketing Affiliate based on the agreements are:

1. Market Equidata's services through various channels like third party programs, direct mail, internet and telemarketing.
2. Comply with all applicable laws and regulations regarding marketing the services.
3. Provide monthly reports on marketing activities and leads generated.
4. Not make any false or misleading statements about Equidata or the services.
5. Protect Equidata's intellectual property like trademarks and not use them in a misleading way.
6. Allow Equidata to audit their marketing practices and activities.[source](#)

Traditional search

Intelligent search

Conversational search



# Use case: improving patient/customer experience



## Self-service virtual agent

Use of automatic speech recognition (ASR) and natural language understanding (NLU) to resolve customer queries with automated responses



## Agent assist

Use real-time transcription, natural language processing (NLP) and translation to improve productivity with automation and coaching



## Real-time & Post Call analytics

Use offline and real-time audio streaming transcription and NLP to derive actionable insights into customer conversations and agent performance





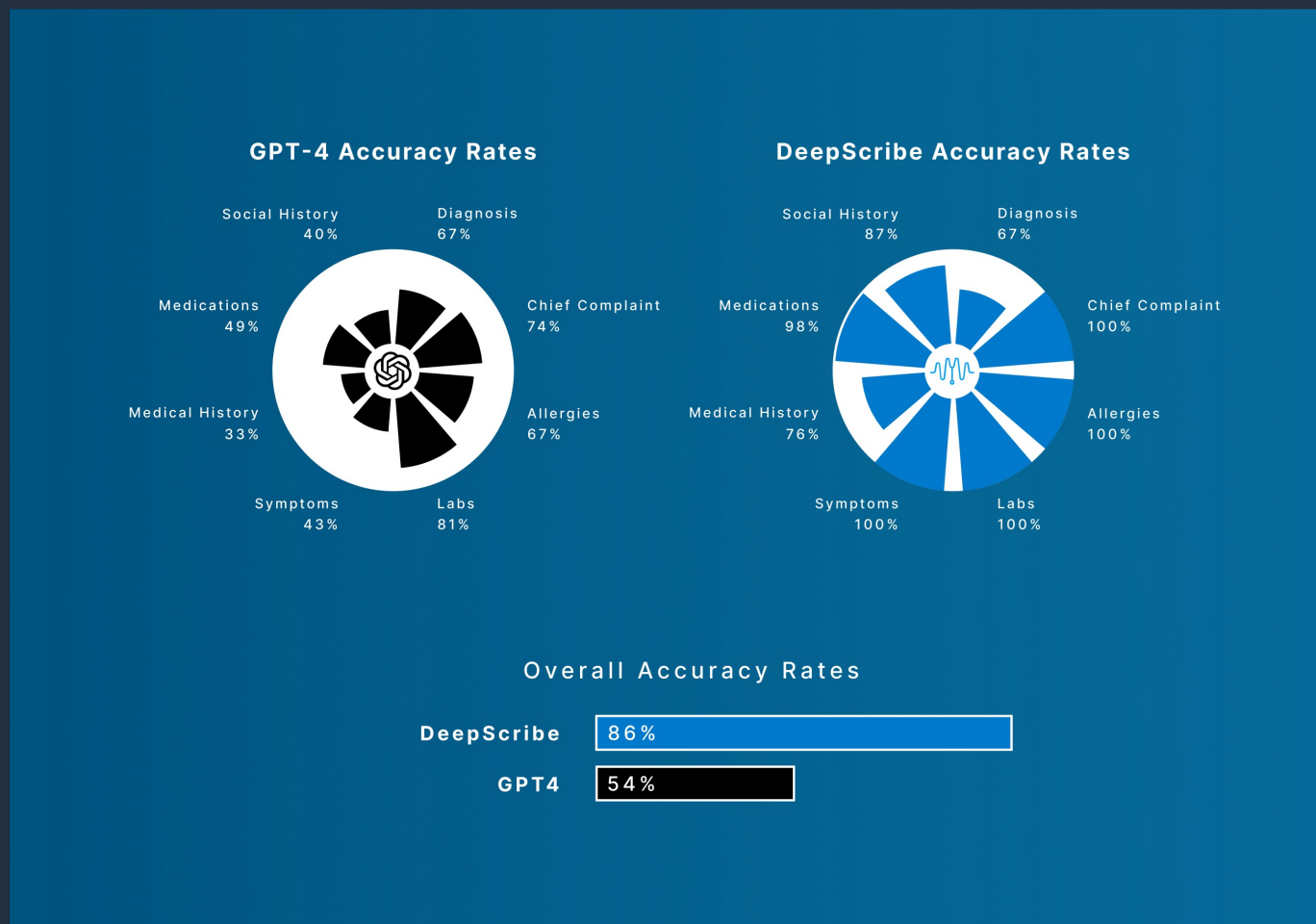
ACCURATELY DOCUMENT PATIENT VISITS WITH AMBIENT AI AND GENERATE MEDICAL NOTES FASTER THAN BEFORE.

Save time

Improve accuracy

Drive adoption

Outperforms GPT-4 by 32% on AI Medical Scribing

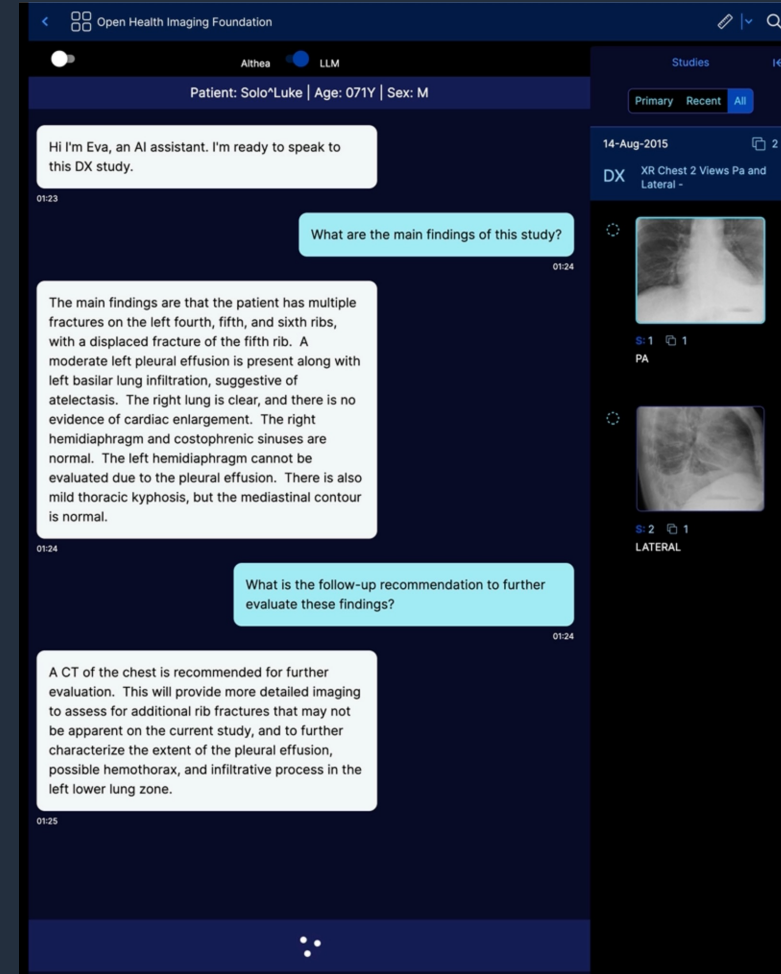




DELIVERING AN API FOR THE FIRST AI FOUNDATION MODEL TO REVOLUTIONIZE MEDICAL IMAGING

## Reduce workload burden for:

- Radiologists
- Neurosurgeons
- Trauma Surgeons
- Academic Medical Centers
- Medical Office Staff





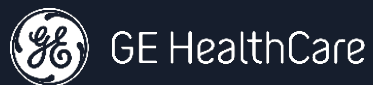
## De-Identification

**De-identify 700M patient notes**



## Clinical Data Abstraction

**Extract what, when, why, and how from radiology reports in real-world data acquisition projects**



## Patient Safety

**Identify opioid-related adverse events from unstructured text in electronic health records**



## Clinical Decision Support

**Real-time clinical decision support platform for psychiatry and oncology**  
**Identify housing insecurity and other social determinates of health**



## Medical Affairs

**Optimize hospital patient flow models**



## Financial Operations

**Accelerate clinical risk adjustment through NLP**



# What's next?



# AWS Immersion Days

NO-COST EDUCATIONAL SESSIONS WITH HANDS-ON LABS

## Public Sector Workshops

Machine Learning in Healthcare

Thursday, May 30



## Solution-Focused Immersion Days

Harnessing Generative AI with  
Amazon Bedrock

Thursday, May 30

Build a Generative AI-Powered  
Assistant Using Amazon Q for  
Enterprise Systems

Wednesday, June 26



# AWS Generative AI Innovation Center

## Art of the possible

Explore and experiment safely, reliably, and cost-effectively

## Responsible use— does not touch PHI

Guardrails to reduce risk—more control and focus across bias and fairness, privacy and security, explainability, and governance for enterprise content and workflows

## Model selection

Select the best foundation models taking advantage of purpose-built infrastructure for training and inference

## Time to value acceleration

Apply foundation models to real-world use cases with the Innovation Center as a trusted capable partner to demonstrate viability and accelerate path to production

## MLOps

Balance efficiency, scalability, and risk reduction with a framework to manage the development and deployment of generative AI models

# 600+ digital courses

SKILLBUILDER.AWS



**COURSES AND LEARNING PLANS** SUBSCRIPTIONS CLASSROOM TRAINING AWS CERTIFICATION AWS PARTNER TRAINING


## AWS Skill Builder

Your learning center to build in-demand cloud skills

**FILTERS** Search... NEWEST TO OLDEST

### Featured content

Browse through the most popular courses offered in AWS Skill Builder




**Cloud for CEOs**

FREE

EN | 10m 00s ★ 5.0

Digital training




**Build with Amazon MemoryDB for Redis**

FREE

EN | 1h 00m ★ 5.0

Digital training




**AWS Best Practices for Hybrid Cloud Adoption**

FREE

EN | 1h 00m ★ 5.0

Digital training




**AWS for Games Learning Plan: Cloud Game Development**

FREE

12 courses | 11h 40m

Learning Plan



**AWS Certified Developer - Associate Official Practice Question Set (DVA-C02 ...)**


FREE

EN | 45m 00s ★ 5.0


Digital training

### Featured subscription content


Get the most out of your AWS Skill Builder subscription with these training offerings available only with the subscription.




**Migrate a monolith web application to AWS using Application Migration...**




**Challenge AWS Game Skills: Basic**



**Exam Prep: AWS Certified Solutions Architect - Professional (SAP-C02) (wit...**



**Exam Prep: AWS Certified Advanced Networking - Specialty (ANS-C01) (with...**



**Industry Quest: Financial Services**

# AWS can deliver on your generative AI journey



The most comprehensive set of data and machine learning services



Investing in a zero-ETL future so you can quickly and easily connect all your data



End-to-end governance capabilities help you move faster with data



A proven leader with more than 1.5 million customers using database, analytics, or ML services



17+ years of data and ML innovation and counting



Multiple ways to get started including training, enablement, and a global network of partners



# Please Provide Your Feedback

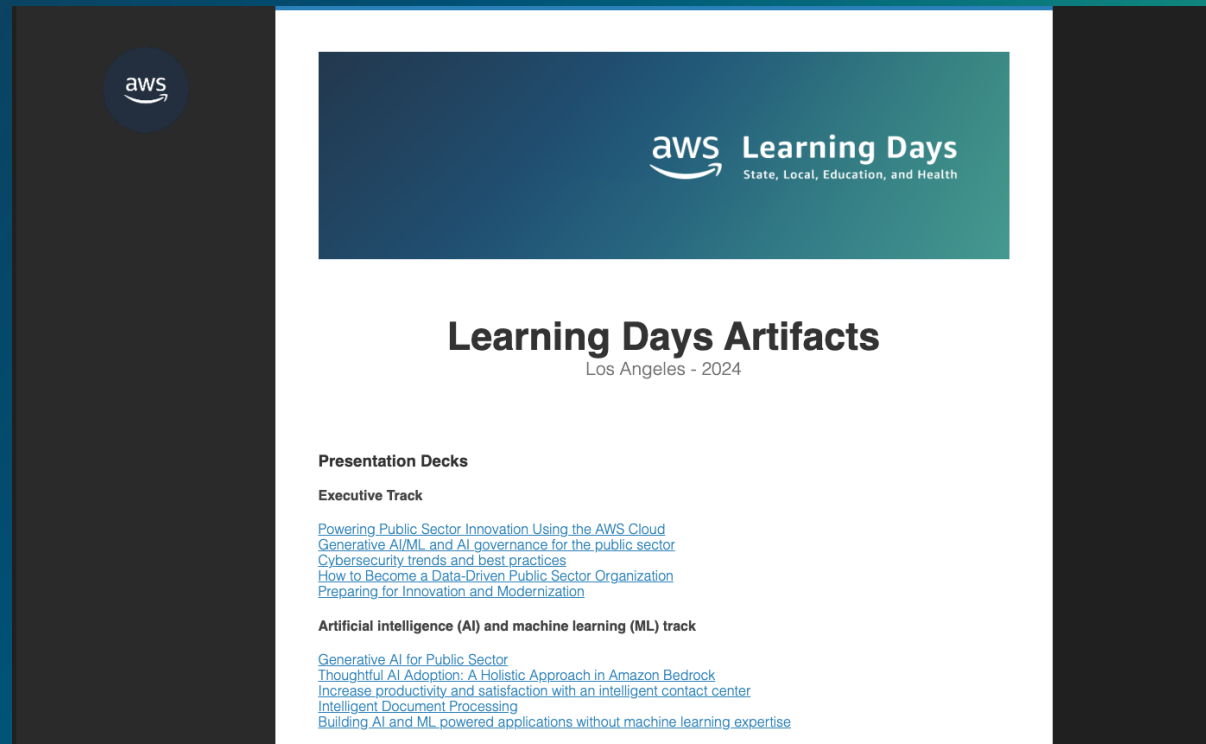


Step 1: Select **Healthcare Track**

Step 2: Select **Generative AI on Healthcare**

# Learning Day Content

<https://sanfrancisco2024.awslearningday.com/>





# Thank you!

Yuan Yao

[yno@amazon.com](mailto:yno@amazon.com)