

AWS State, Local, and Education Learning Days

Philadelphia



BREAKOUT

Generative AI for Public Sector

Satinder Sidhu

Sr. Solution Architect,
AWS WWPS Higher Education

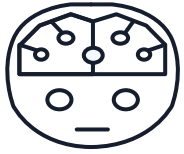
ssidhu@amazon.com

The background is a deep blue gradient. Overlaid on this is a complex, glowing network of white and light blue lines and dots, resembling a neural network or a data mesh. In the upper half, there are faint, semi-transparent binary digits (0s and 1s) scattered across the frame.

AI/Machine learning (ML) is at an inflection point

Key drivers: Compute capacity increase | Data growth | Model sophistication

AI, ML, deep learning?



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 corpuses of data and commonly referred to as foundation models (FMs)

Challenges we are hearing from public sector customers



Demand for government services is rising while resources and capacity to deliver them **aren't keeping pace**



Citizens increasingly expect the government to **provide modern digital experiences** for conducting online transactions



Aging infrastructure for data capture, storage, and management **creates friction** for leveraging data for analytics and machine learning

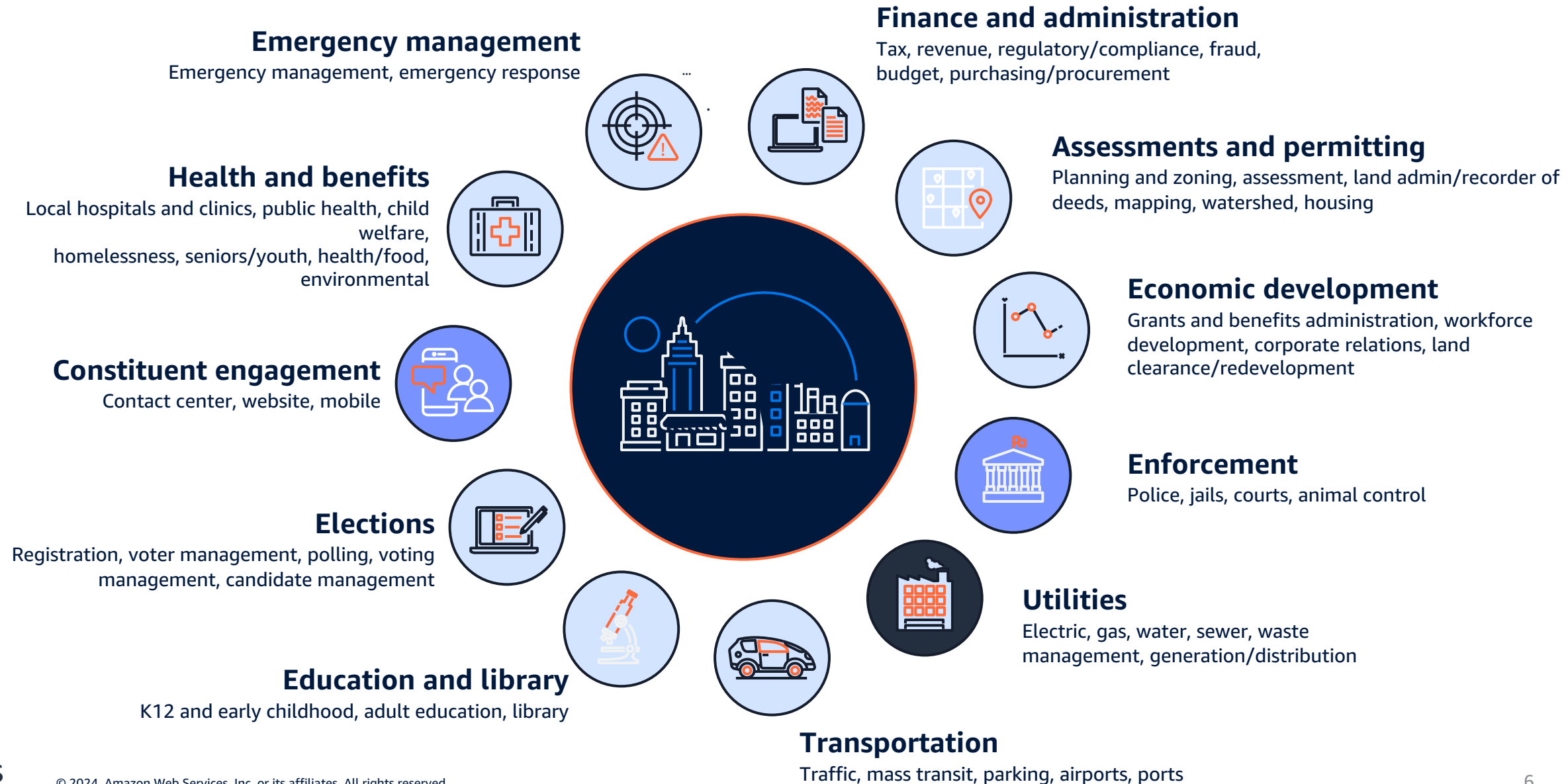


Complex security, privacy, and compliance requirements create barriers to change and block adoption of many SaaS solutions



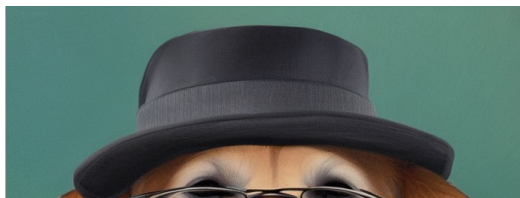
Risk averse culture and institutional inertia slow innovation

Machine learning is going mainstream in public sector



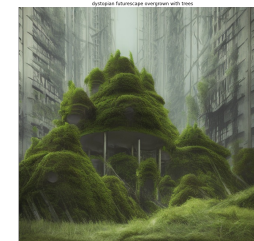
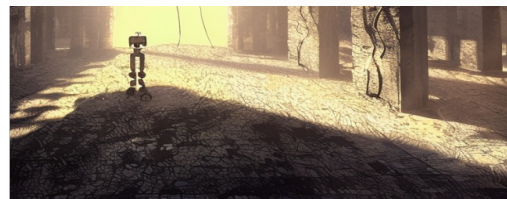
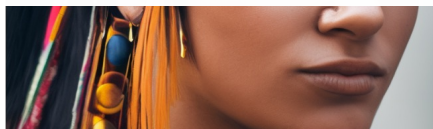
The background is a deep blue gradient. Overlaid on this is a complex, glowing network of white and light blue lines and dots, resembling a neural network or a data mesh. The lines are thin and connect various points, some of which are highlighted with small, bright white circles. In the upper portion of the image, there are faint, light blue binary digits (0s and 1s) scattered across the background, giving it a digital or data-centric feel.

Amazon Generative AI



Question: What is generative artificial intelligence (AI)?

- Creates new content and ideas, including conversations, stories, images, videos, and music
- Powered by large models that are pretrained on vast corpuses of data and commonly referred to as foundation models (FMs)



Generative AI core capabilities

Capabilities



Generation



Question
answering



Summarization



Translation



Correction



Classification

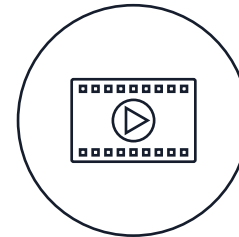
Input/Output



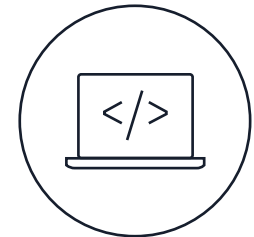
Text



Images

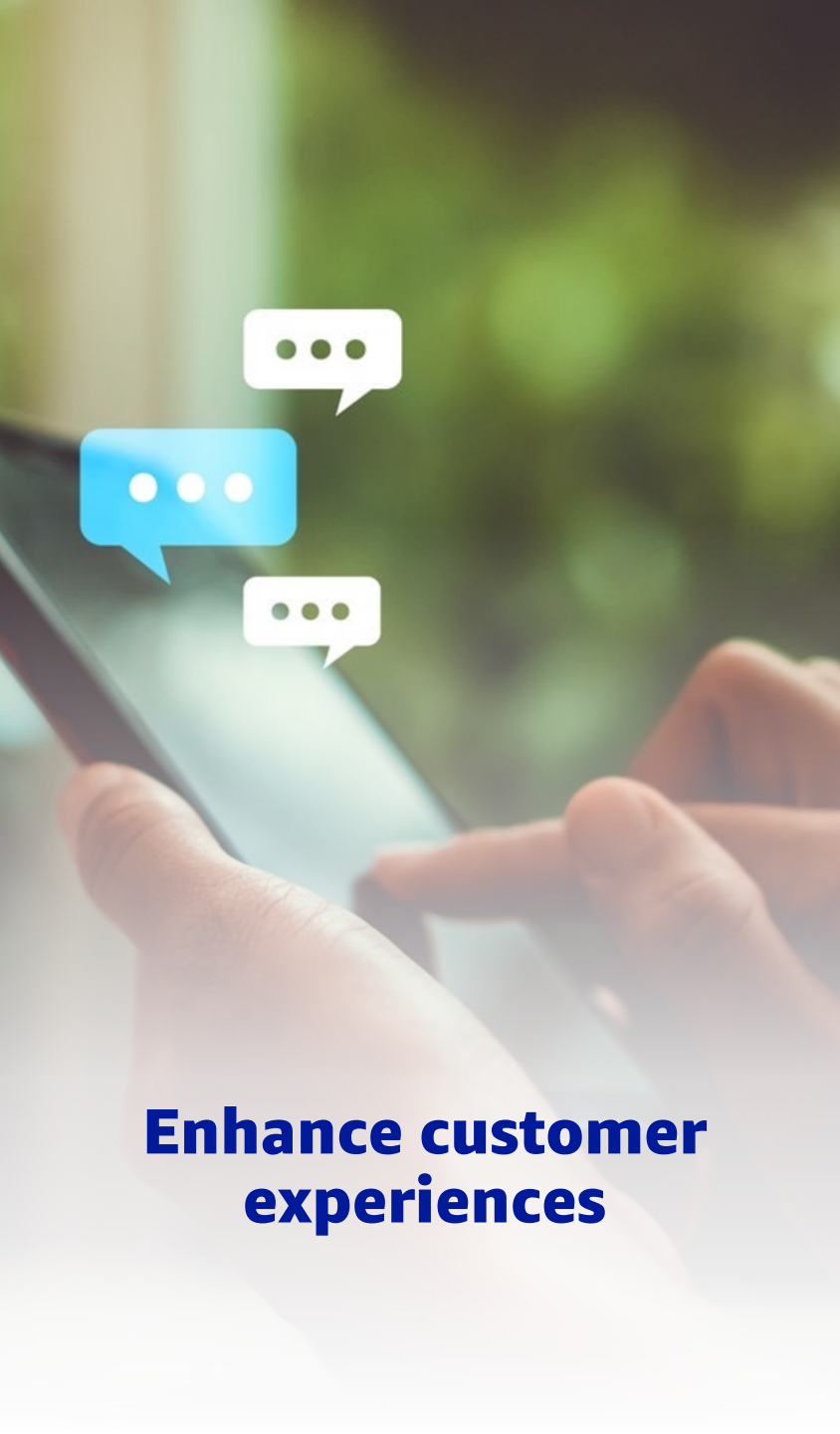


Video/Audio

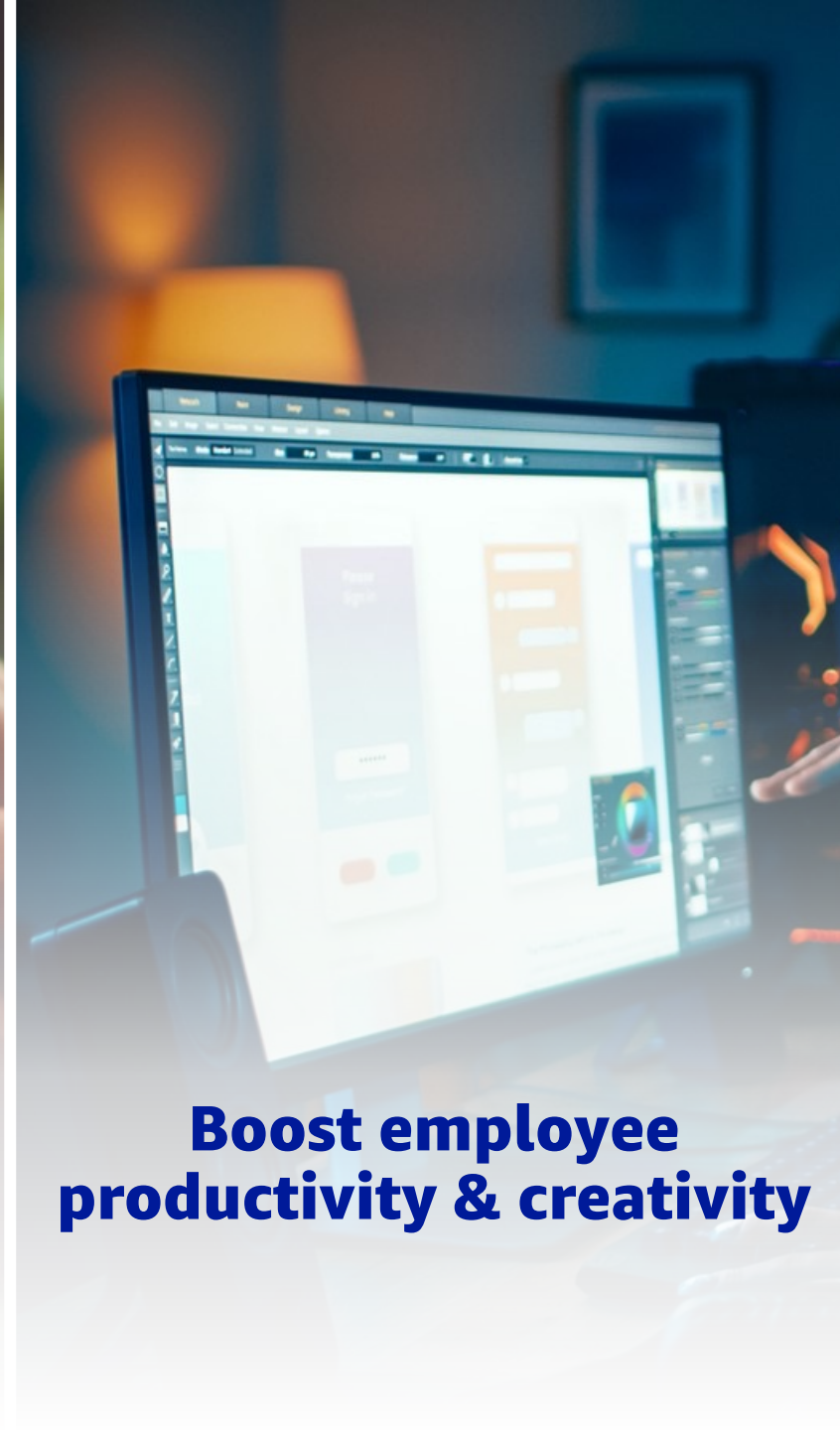


Code

Use Cases for Public Sector



**Enhance customer
experiences**



**Boost employee
productivity & creativity**



**Optimize business
processes**



Enhance customer experiences

Chatbots

Virtual assistants

Conversation analytics

Personalization



Boost employee productivity & creativity

Employee assistants

Code generation

Report generation



Optimize business processes

Document processing

Cybersecurity

Process optimization

Data augmentation

Enhance customer experiences

Chatbots

Virtual assistants

Conversation analytics

Personalization

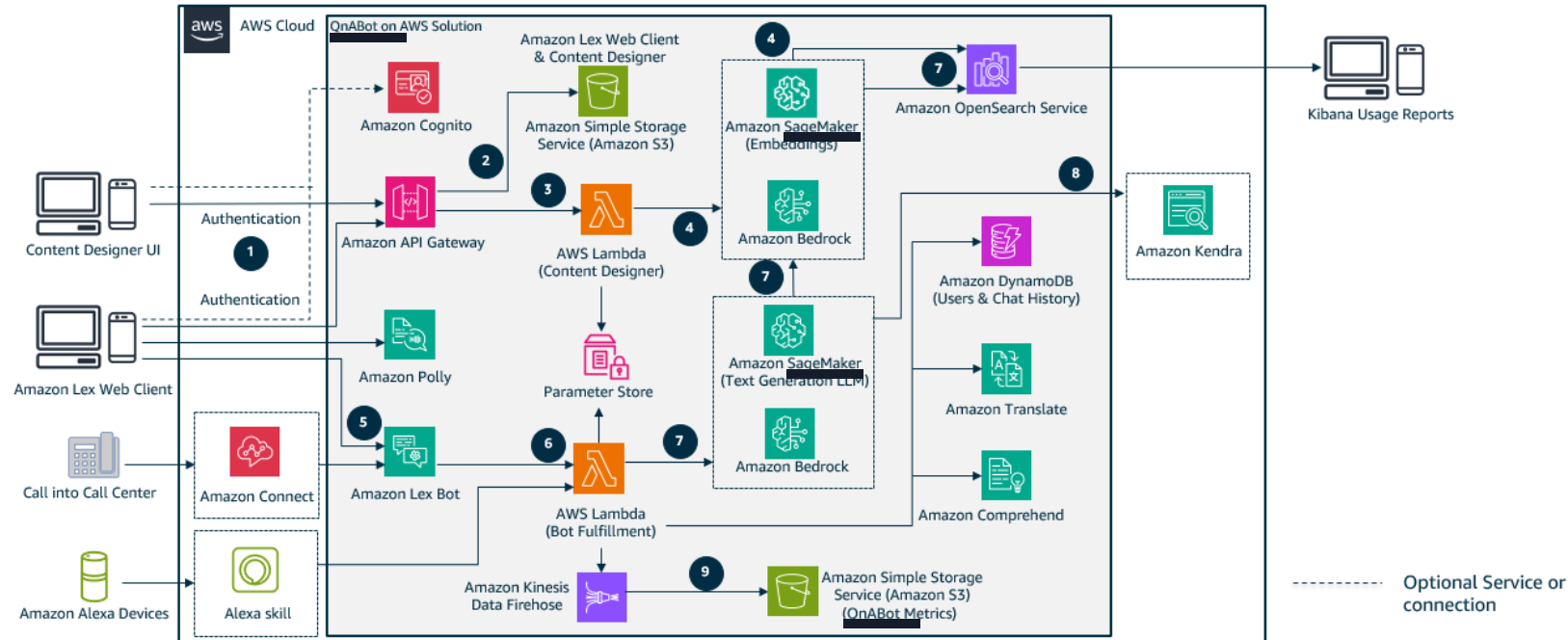
Generative AI-Enabled Citizen Engagement

Engage citizens and drive improvements in customer satisfaction

- Improve contact center agent effectiveness with real-time translation and decision support using Amazon Connect and Contact Center Intelligence
- Analyze call and text interactions with citizens to spot issues and trends and drive improvement
- Improve self service



Get started now: QnABot solution



Deliver impactful customer experiences using generative AI



Boost employee productivity & creativity

Employee assistants

Code generation

Report generation



Boost employee productivity & creativity



Content discovery

Provide a unified conversational search experience across all corporate information



Summarization and comparative analysis

Summarize data from various sources into actionable insights and perform comparative analysis on your enterprise content



Extract key insights

Unlock powerful insights from your data and enable every employee to make faster and better data-driven decisions



Content creation

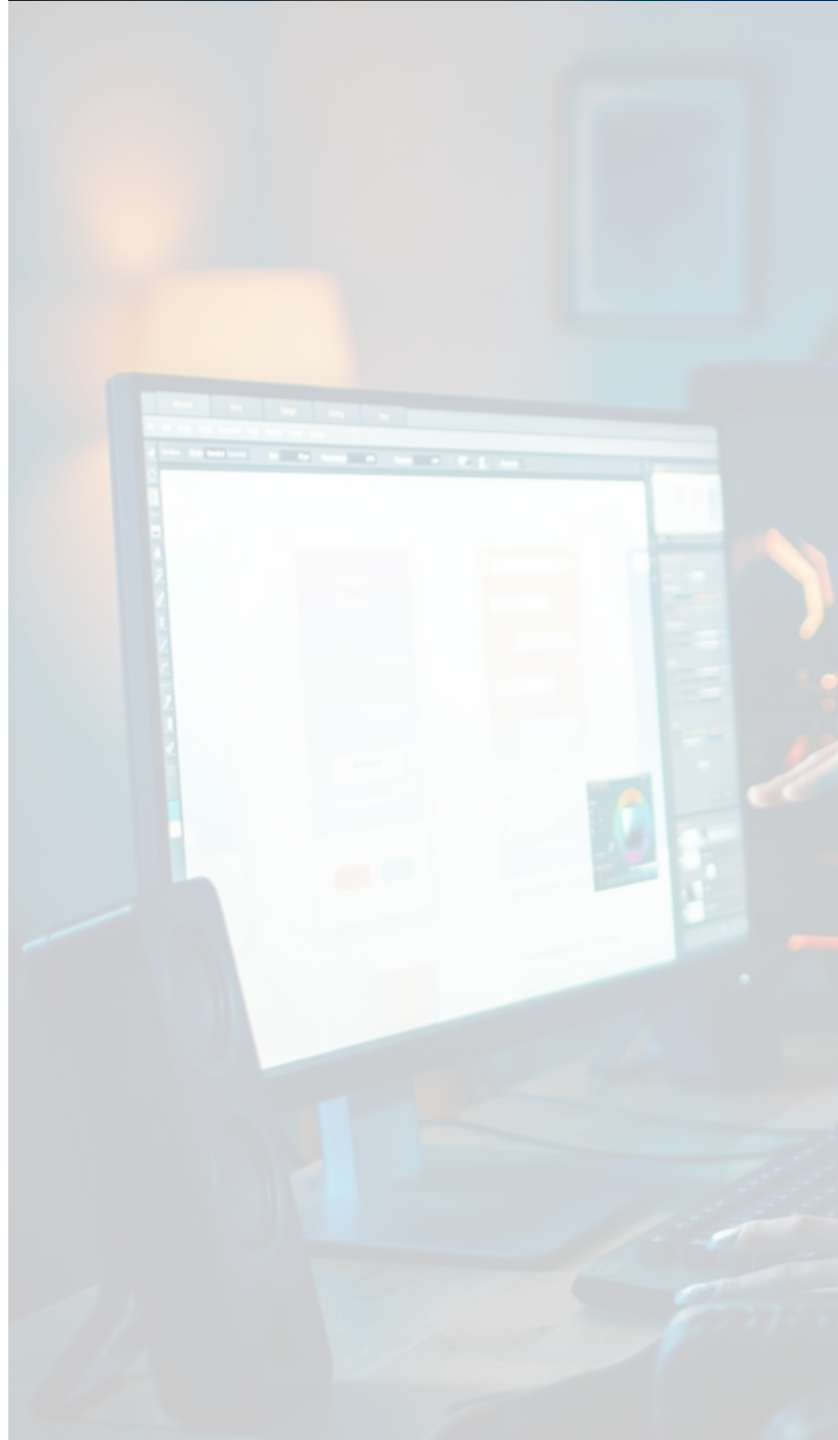
Empower employees to create content faster and smarter across departments—from marketing to sales to engineering

Using AI to improve agent efficiency

“During peak hours, previously you’re 45-50 minutes on hold, and now that’s has been reduced to about three and a half minutes. One of the other benefits we’ve gotten from Amazon Connect is sentiment analysis. On a call, we get real-time feedback on whether or not the customer was happy, frustrated, or angry...”

—Benny Chacko, Deputy General – LA County Internal Services Department





Optimize business processes

Document processing

Cybersecurity

Process optimization

Data augmentation



Optimize business process

Sales

Personalized outreach
Sales scripts
Competitive analysis
Personalized offers

Management

Decision support
Strategic planning
Team management
Project management

Marketing

Content creation
Personalized Marketing
Brand identity design
SEO optimization

Legal

Document review
Contract analysis
Legal research
Draft legal documents

Product Management

Product design
Feature prioritization
Market research
User persona creation

HR

Job descriptions/postings
Personalized learning paths
Generate surveys
Onboarding plans

Extract insights from unstructured content

Extract insights from unstructured documents and forms, like images, PDFs, and audio

- Analyze text with natural language processing (NLP) to identify topics, extract entities, understand sentiment, and classify documents with Amazon Textract, Amazon Rekognition, and Amazon Comprehend
- Translate content at scale with Amazon Translate

-disappeared in [redacted] TFSB0 positive [redacted] with a 5.0 cm L [redacted] with a [redacted]
-immunological tests have
-received amoxiclavast dose dense [redacted] and [redacted] [redacted] to [redacted] and then received [redacted] 64 [redacted] [redacted].
-allergic reactions does [redacted] with [redacted] [redacted] She had no [redacted] on the [redacted] and [redacted] with residual disease 2.5
cm in Nottingham grade 2 with [redacted] she had [redacted] sizes positive.
-treatment with concurrent [redacted] [redacted] and then went on [redacted] from [redacted] when she was switched to
~~[redacted]~~
-no evidence of progression until [redacted] when she was found to have [redacted] pain and swelling found to have a
[redacted] bladder from peritoneal implants and [redacted] Her biopsy showed [redacted] 28 mm x 17 mm x 12 mm negative HRP negative
-bladder [redacted] status [redacted]
-breasts and testicles [redacted] with stable disease although swelling about monthly [redacted]
-her [redacted] which showed a large C pleural effusion and [redacted] with multiple bilateral [redacted] [redacted] [redacted] [redacted] and
variable renal function.
-patient held from 11/20-11/1 to go on a study with [redacted] [redacted]. While off all therapy for week
off and while waiting for [redacted] [redacted] she had worsening symptoms, going to requiring [redacted] [redacted] weekly
tube had one on [redacted] and a second on [redacted] with 1.7 l alone with a [redacted] [redacted] She was found to be AN positive
-metformin removed 1600 mg found to be malignant [redacted] [redacted] [redacted] paraneoplastic removed 1500
at fluid found to be malignant [redacted] [redacted] because of [redacted]
-before [redacted] moved to [redacted] days on and 7 days off on [redacted] because of [redacted]
-[redacted] [redacted] on the week of [redacted] removed 2L fluid found to be [redacted] cancer as has [redacted] [redacted]
as [redacted]
-she had several stents replaced and found 2.5 cm velvety lesion on bladder wall biopsy at [redacted] [redacted] [redacted] [redacted] [redacted]
not consistent with breast cancer (but not consistent with her primary [redacted]) from [redacted]
-level 4 [redacted]
-vincristine [redacted] (admitted with worsening [redacted])
-now [redacted] added [redacted] on [redacted] but on [redacted] she got [redacted] so [redacted] held and on [redacted] got
uncontrolled only
-on admission had [redacted] allowed for [redacted] but [redacted] no response
-first assessment, she had [redacted] on [redacted] yo and [redacted] after [redacted] on [redacted] yo. She has [redacted] and had first [redacted] at
[redacted] yo, she took [redacted] from [redacted] yo until her [redacted]
-mother had multiple systems but no other history of [redacted], polycystic ovary or ovarian cancer
-patient states DNA positive in [redacted]
2. Post onset and treatment on first arriving, improved with course of [redacted]
3. After [redacted] patient came in at [redacted] some disease control gained weight to 91.5 kg, her [redacted] back down to 88.5 kg,
4. [redacted] she was admitted on [redacted] because [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]
She had her [redacted] [redacted] and [redacted] row down to [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] held and patient
admitted. E sent [redacted] and enterococcus faecalis growing in the urine and port.

REVIEW OF SYSTEMS:
has not had any [redacted] since day 2 of her admission but has severe [redacted]. She is very [redacted] and came in a
wheelchair. She has had rapid accumulation of her [redacted] and is very [redacted] therefore not eating well. She also is having
constant shortness of breath and has [redacted] making it difficult to get to the bathroom. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]
has slightly worse but mostly clear. No [redacted] in her lungs no rales crackles. No [redacted] or [redacted]. Complete review of
systems otherwise negative in detail. Patient is [redacted] performance status 2

Current Concerns:
Her [redacted] reports severe [redacted]. She reports [redacted] breathing but is still having [redacted] volume around the house and
to the bathroom. She did not get blood in the hospital and continue to drop. She also has [redacted] [redacted] from her
[redacted]. She is [redacted] because she was [redacted] until she had the [redacted] held. She is also worried about the [redacted]
making her feel [redacted] and not able to live her life. She does understand the [redacted] but wants to keep trying.

Allergies
No Known Medication Allergies

- Demographics
 - Geography
 - Gender
 - Age
 - Health history
 - Disease indicators
 - Anatomy
 - Symptoms
 - Diagnosis
 - Treatment
 - Test Name
 - Test Result
 - Rx
 - Medication Name
 - Strength
 - Route

King County Assessor's Office



CHALLENGE

Reduce data entry, eliminate data errors, and improve data time lines.

SOLUTION

Intelligent documents processing for documents and electronic files, streamlining and unlock data and information from paper documents and electronic files

RESULT

King County employees saved 60% of their time of loading the information into digital format, OCR, & made the data searchable digitally to their internal employees improving 50% faster assessment completions.

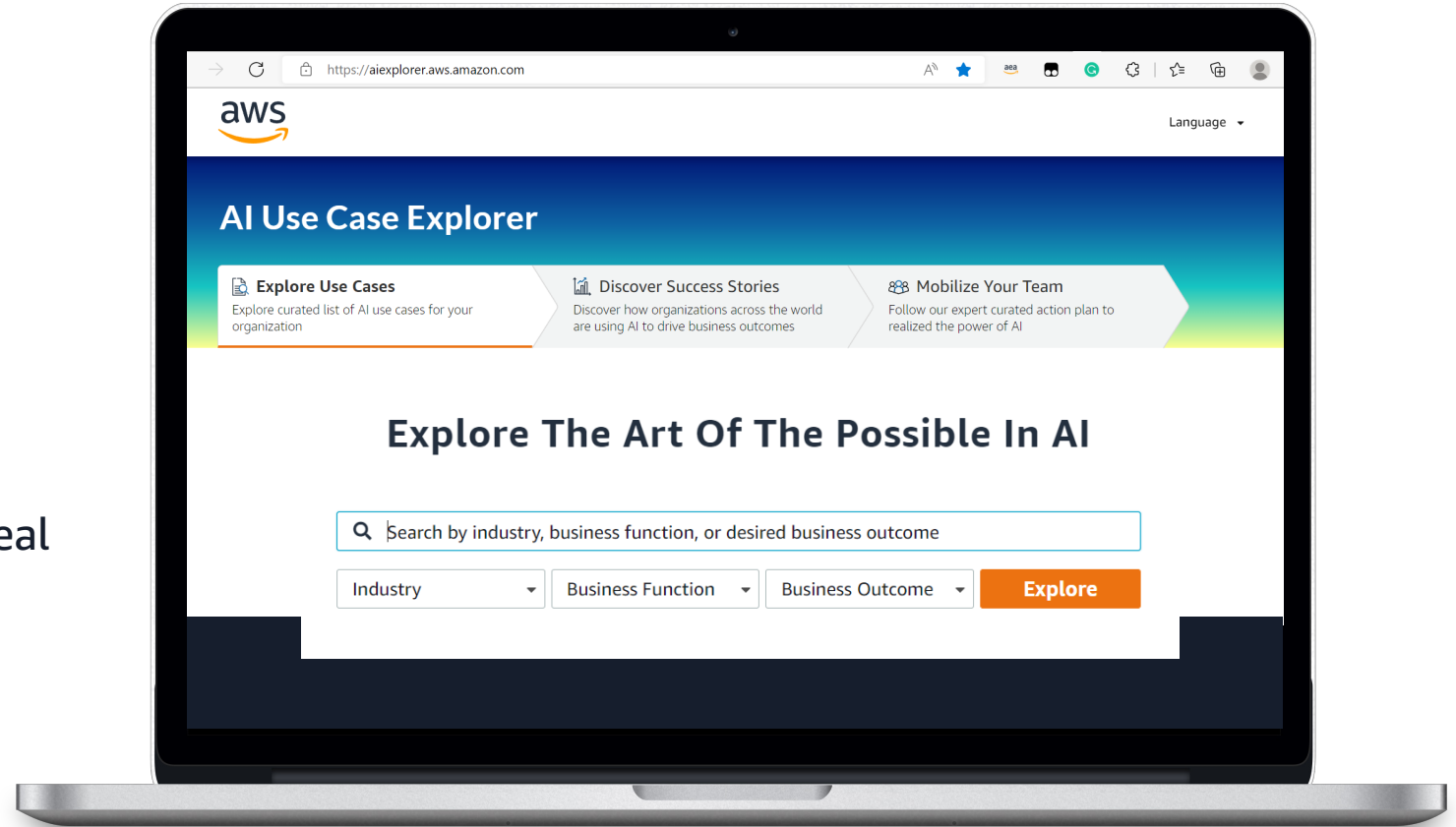


This Photo by Unknown Author is licensed under CC BY



AI Use Case Explorer

Easily find the most relevant AI use cases with related content and guidance to make them real



aiexplorer.aws.amazon.com





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



Make generative AI work with **your data**



RETRIEVAL-AUGMENTED GENERATION (RAG)

Specialized knowledge through prompt augmentation

Enterprise knowledge corpus

No change to the foundation model



FINE-TUNING

Specialized knowledge for specific tasks

Small number of labeled examples

Change a copy of the foundation model



CONTINUED PRE-TRAINING

Generalized and specialized knowledge for your domain

Unlabeled, unstructured enterprise data

Change a copy of the foundation model

Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND FMs



Amazon Q



AWS App Studio

TOOLS TO BUILD WITH LLMs AND OTHER FMs



Amazon Bedrock

Guardrails

Agents

Studio

Customization

Custom Model Import

Amazon Models

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE



SageMaker



Trainium



Inferentia



GPUs



UltraClusters



EFA



EC2 Capacity Blocks



Nitro



Neuron

Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND FMs

TOOLS TO BUILD WITH LLMs AND OTHER FMs

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE



SageMaker



Trainium



Inferentia



GPUs



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

Innovating at the silicon level

AWS Trainium



AWS Inferentia

Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND FMs






TOOLS TO BUILD WITH LLMs AND OTHER FMs

Amazon Bedrock

Guardrails | Agents | Studio | Customization | Custom Model Import | Amazon Models

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE

 SageMaker  Trainium  Inferentia  GPUs

 UltraClusters  EFA  EC2 Capacity Blocks  Nitro  Neuron

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

ANTHROPIC

 **cohere**

 **Meta**

stability.ai



Jurassic

Claude

Command + Embed

Llama 2

Stable Diffusion

Amazon Titan

Contextual answers,
summarization,
paraphrasing

Summarization, complex
reasoning, writing, coding

Text generation,
search, classification

Q&A and reading
comprehension

High-quality images
and art

Text summarization,
generation, Q&A,
search

Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND FMs



Amazon Q



AWS App Studio

TOOLS TO BUILD WITH LLMs AND OTHER FMs



Amazon Bedrock

Guardrails | Agents | Studio | Customization | Custom Model Import | Amazon Models

INFRASTRUCTURE FOR FM TRAINING AND INFERENCE



GPUs



Trainium



Inferentia



SageMaker



UltraClusters



EFA



EC2 Capacity Blocks



Nitro



Neuron

Amazon Q

Reinvent work with AWS'
generative AI-powered assistant

Unites your data sources, repositories, and
enterprise systems

Available where you work

Delivers industry-leading assistance

Reinvent work with **Amazon Q**

BUSINESS



Amazon Q Business

KNOWLEDGE SEARCH

SUMMARIZATION

CONTENT CREATION

EXTRACT INSIGHTS

RESEARCH & ANALYSIS



Amazon Q in QuickSight

UNDERSTAND DATA

BUILD & REFINE VISUALS

BUILD CALCULATIONS

EXECUTIVE SUMMARIES

CREATE DATA STORIES

DEVELOPERS



Amazon Q Developer

PLAN APPLICATION

CODE GENERATION

UNIT TESTING

SECURITY SCANNING

CODE REMEDIATION

CODE MIGRATION

TROUBLESHOOTING

DEVELOPER KNOWLEDGE

SPECIALIZED USERS



Amazon Q in Connect

AGENT ASSIST



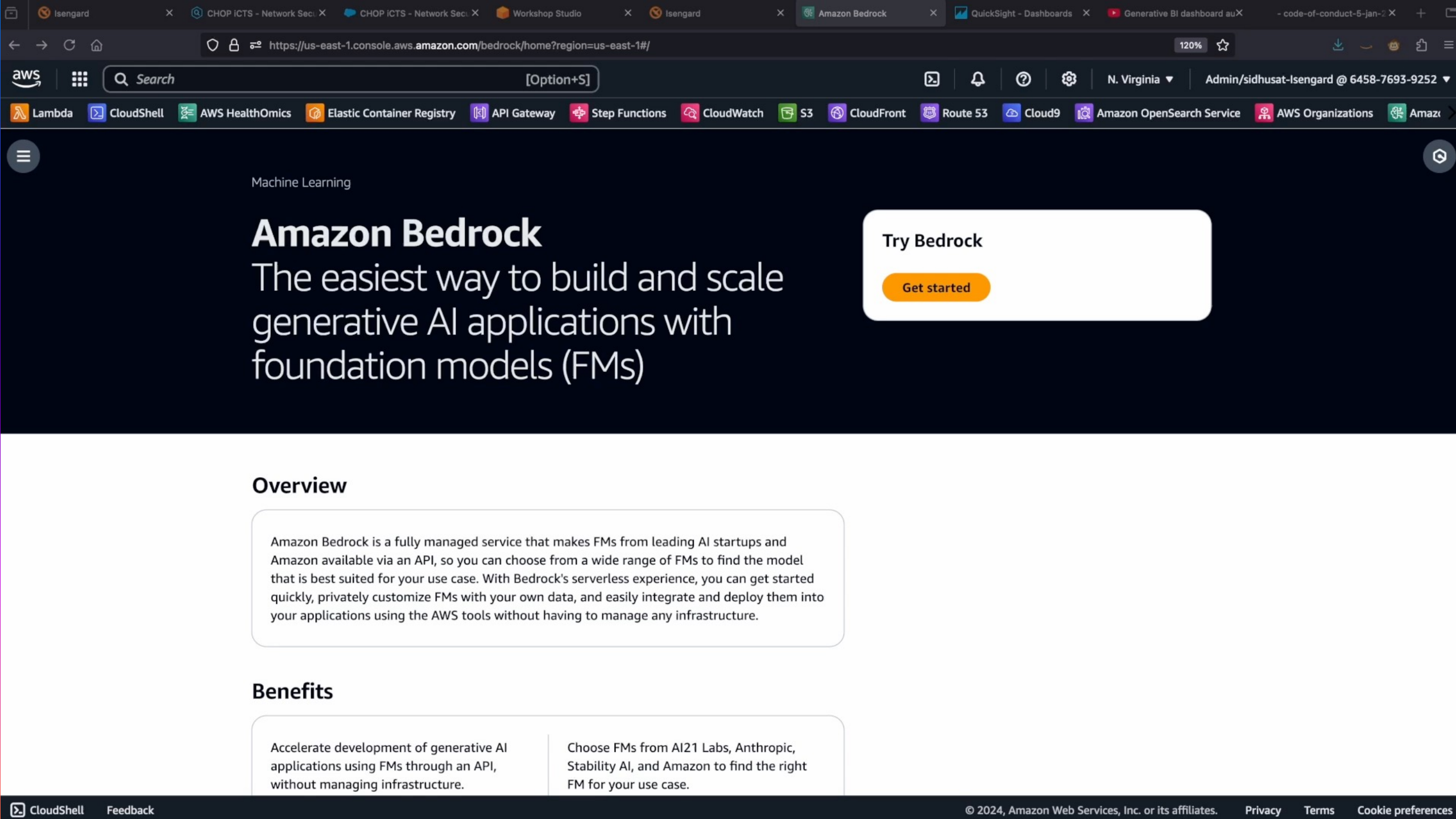
Amazon Q in AWS Supply Chain

SUPPLY CHAIN

Demo(s)

Amazon Bedrock *Content Generation*





Machine Learning

Amazon Bedrock

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

Try Bedrock

Get started

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Accelerate development of generative AI applications using FMs through an API, without managing infrastructure.

Choose FMs from AI21 Labs, Anthropic, Stability AI, and Amazon to find the right FM for your use case.

Amazon Bedrock

Text Summarization



Amazon Bedrock

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

Try Bedrock

Get started

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Accelerate development of generative AI applications using FMs through an API, without managing infrastructure.

Choose FMs from AI21 Labs, Anthropic, Stability AI, and Amazon to find the right FM for your use case.

Amazon Bedrock

Text to Image



Machine Learning

Amazon Bedrock

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

Try Bedrock

Get started

Overview

Amazon Bedrock is a fully managed service that makes FMs from leading AI startups and Amazon available via an API, so you can choose from a wide range of FMs to find the model that is best suited for your use case. With Bedrock's serverless experience, you can get started quickly, privately customize FMs with your own data, and easily integrate and deploy them into your applications using the AWS tools without having to manage any infrastructure.

Benefits

Accelerate development of generative AI applications using FMs through an API, without managing infrastructure.

Choose FMs from AI21 Labs, Anthropic, Stability AI, and Amazon to find the right FM for your use case.

Amazon Q Console



Reset to default layout + Add widgets

⋮ Applications (0) Info

Create application

- us-east-1 (Current Region) Find applications

< 1

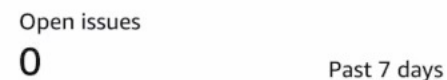
No applications

Get started by creating an application.

Create application

[Go to myApplications](#)

⋮ AWS Health [Info](#)



Scheduled changes

0 Upcoming and past 7 days

Other notifications

Current month costs

↑ 37% compared to last month for same period

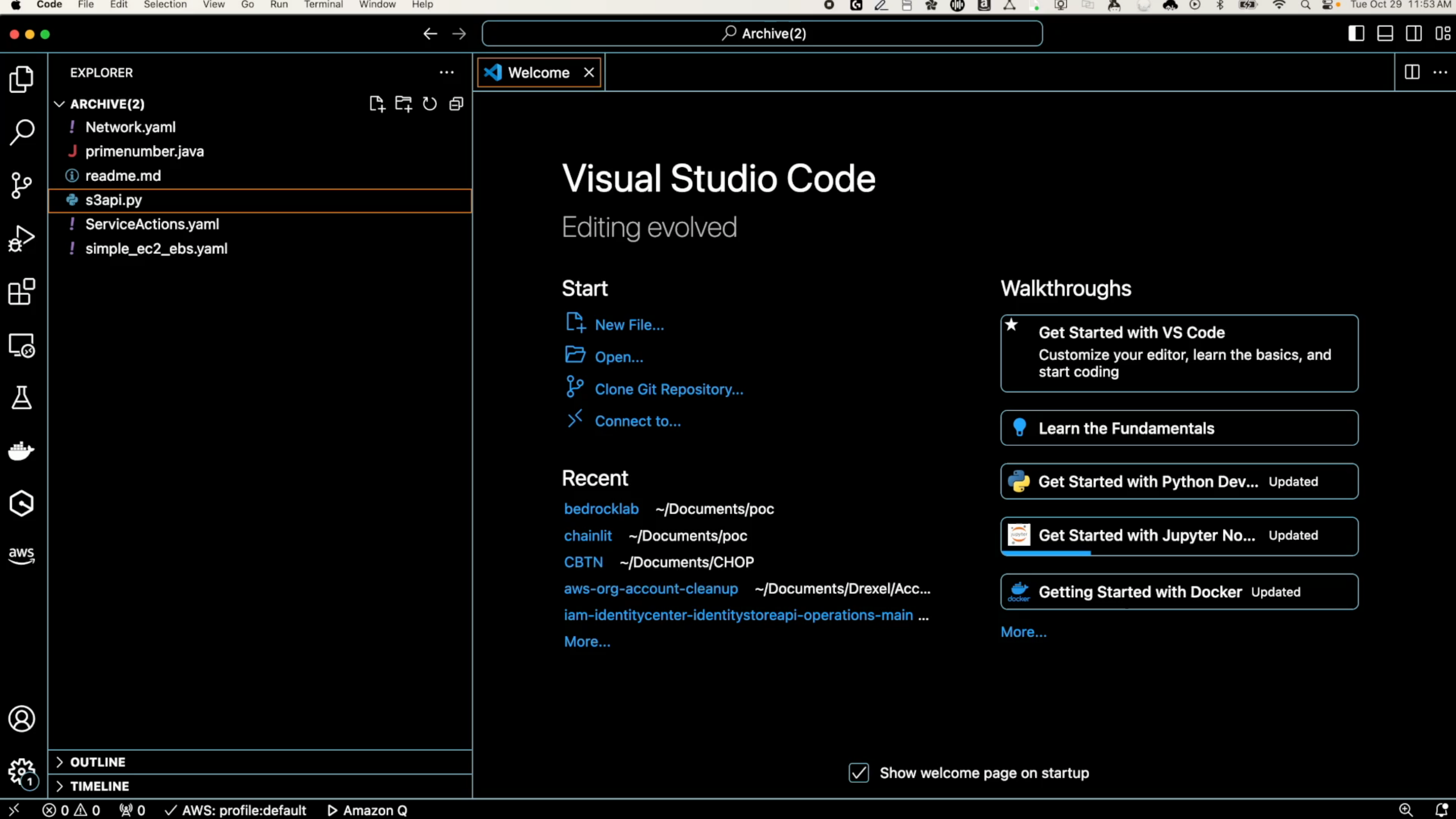
Forecasted month end costs

Cost (\$)



Amazon Q Developer





Amazon Q in QuickSight





Sheet 1

100%

Sales Pipeline

lds

RELATED FIELD

tem

sted Monthly Revenue

ed

t

ame

Community Stage

erson

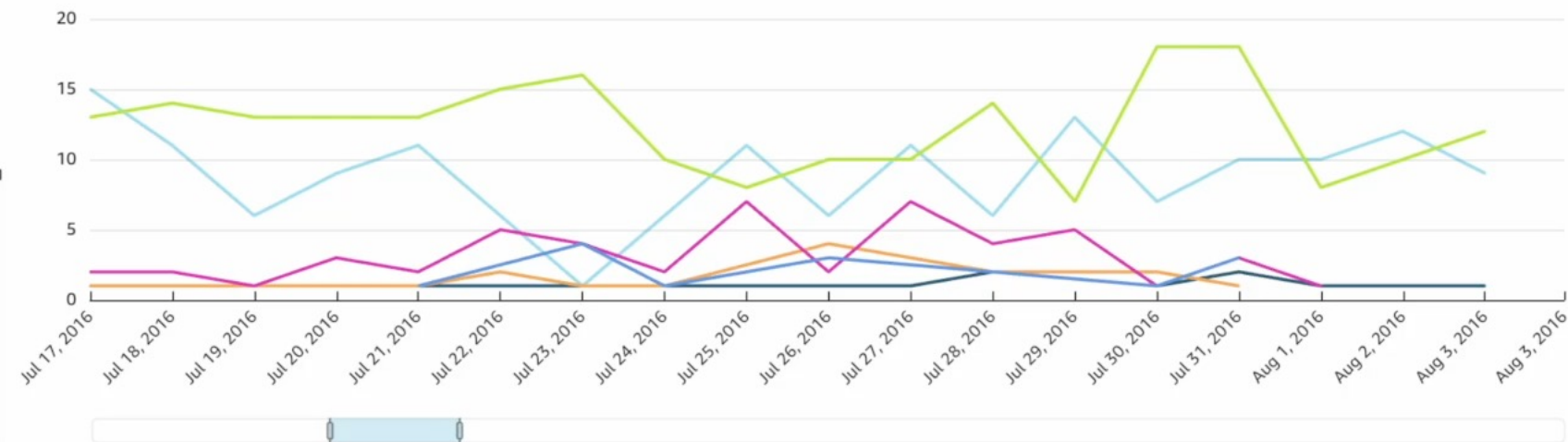
nt

Close

ed Revenue

Historical opportunity pipeline

SHOWING TOP 200 IN DATE AND BOTTOM 6 IN OPPORTUNITY STAGE



Properties

Visual Interacti

> Display Settings

> Data series

> X-axis

> Y-axis

> Group/Color

> Reference lines

> Legend

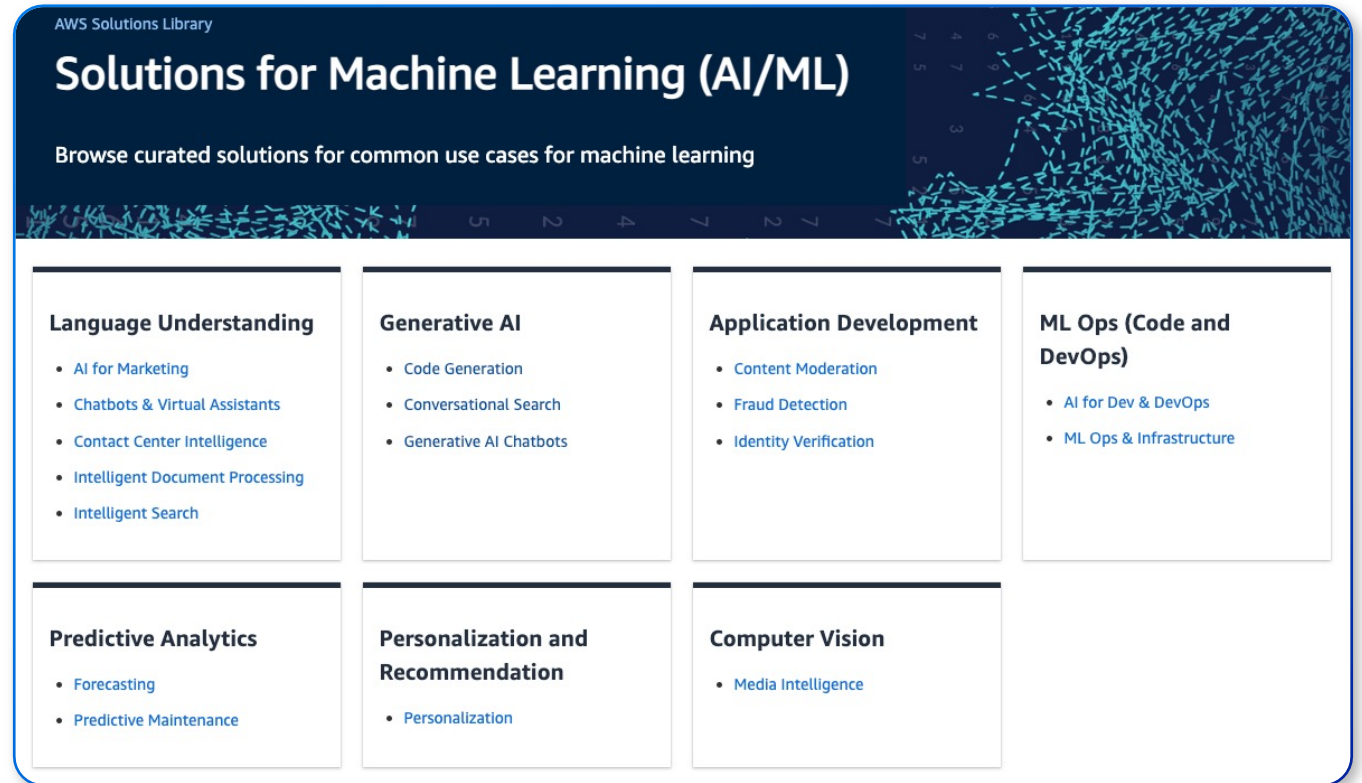
> Data labels

Generative AI solutions library

AWS Solutions

FOR MACHINE LEARNING

The optimal way to
meet your AI/ML needs



www.aws.amazon.com/solutions/ai-ml



Security considerations for generative AI

COMPLIANCE & GOVERNANCE

The policies, procedures, and reporting needed to empower the business while minimizing risk

Create generative AI usage guidelines

Establish process for output validation

Develop monitoring & reporting processes

LEGAL & PRIVACY

The specific regulatory, legal, and privacy requirements for using or creating generative AI solutions.

Retain control of your data

Encrypt data in transit and at rest

Support regulatory standards

CONTROLS

The implementation of security controls that are used to mitigate risk.

Human-in-the-loop

Explainability & auditability

Testing strategy

Identity and access management

RISK MANAGEMENT

Identification of potential threats to generative AI solutions and recommended mitigations.

Threat modeling

Third-party risk assessments

Ownership of data, including prompts and responses

RESILIENCE

How to architect generative AI solutions to maintain availability and meet business SLAs.

Data management strategy

Availability

High Availability and Disaster Recovery strategy



Responsible AI

Design, build & operate AI Systems responsibly.



Responsible AI Dimensions

FAIRNESS

Considering impacts on different groups of stakeholders

EXPLAINABILITY

Understanding and evaluating system outputs

CONTROLLABILITY

Having mechanisms to monitor and steer AI system behavior

SAFETY

Preventing harmful system output and misuse

PRIVACY & SECURITY

Appropriately obtaining, using and protecting data and models

GOVERNANCE

Incorporating best practices into the AI supply chain, including providers and deployers

TRANSPARENCY

Enabling stakeholders to make informed choices about their engagement with an AI system

VERACITY & ROBUSTNESS

Achieving correct system outputs, even with unexpected or adversarial inputs

Responsible AI: Best practices



Put your people first



Assess risk on a (use) case-by-case basis



Iterate across the AI lifecycle



Test, test again, and then test again



Thank you!

skillbuilder.aws
aws.amazon.com/solutions/ai-ml
aiexplorer.aws.amazon.com

Satinder Sidhu

Sr. Solution Architect,
WWPS Higher Education
ssidhu@amazon.com
Linekdin: SatinderSidhu

Please complete the survey
for this session



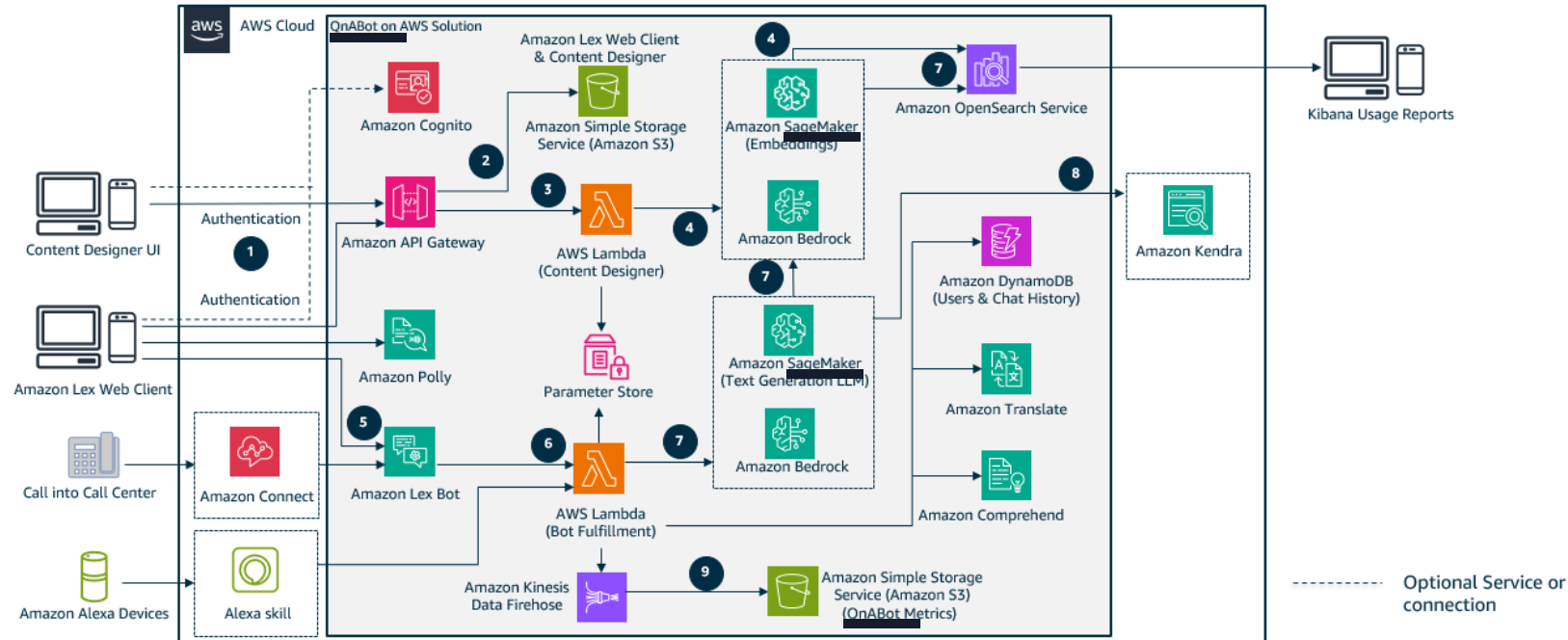
Artificial Intelligence & Machine Learning

Generative AI for Public Sector & Higher Education



Annexure

Get started now: QnABot solution

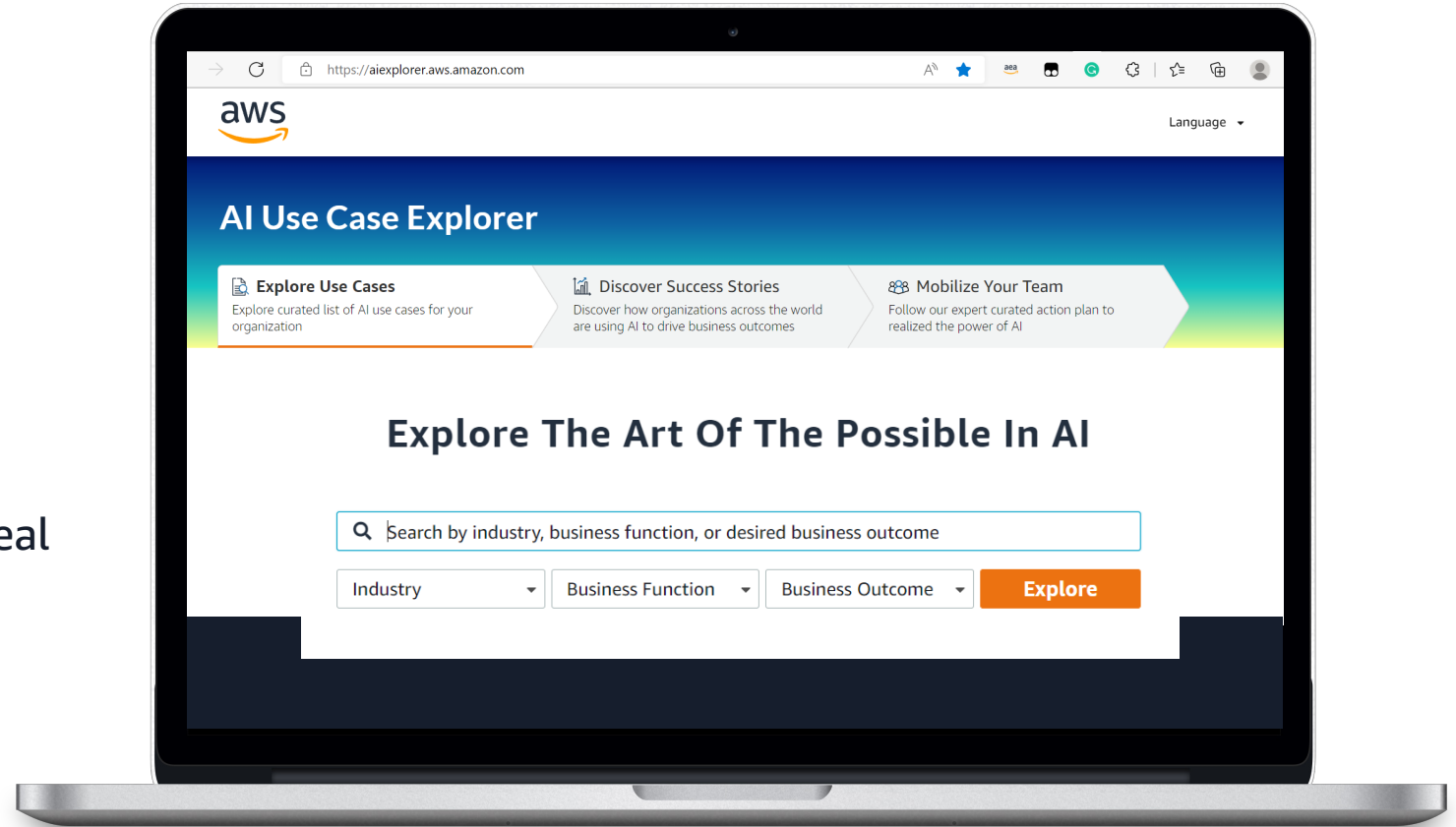


Deliver impactful customer experiences using generative AI



AI Use Case Explorer

Easily find the most relevant AI use cases with related content and guidance to make them real



aiexplorer.aws.amazon.com

