

# AWS State, Local, and Education Learning Days

North Carolina



**BREAKOUT**

# **Generative AI for Public Sector**

**Anthony Harvey**

Sr. Security Specialist Solutions Architect

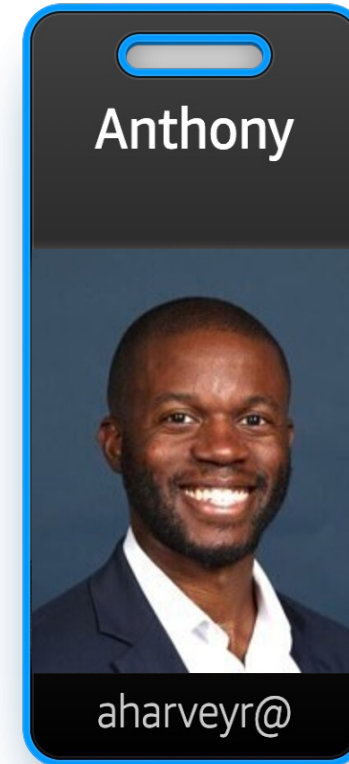
World Wide Public Sector – SLG/EDU

AWS


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

# Who am I?

- Anthony Harvey ([aharveyr@amazon.com](mailto:aharveyr@amazon.com))
  - Sr. Security Specialist Solutions Architect
- Former Chief Information Security Officer in local government
- Former Director of IT & Cybersecurity in higher education
- Career focused on the public sector



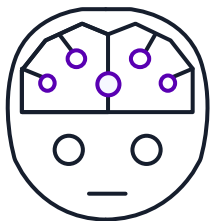
LinkedIn



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

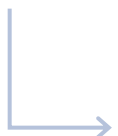
**Key drivers:** Compute capacity increase | Data growth | Model sophistication

# What is it?



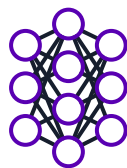
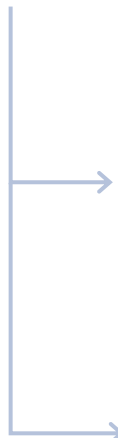
## Artificial intelligence (AI)

Any technique that enables 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



## Classification AI and Predictive AI

A subset of ML that recognizes patterns to identify something (Classification AI) or predicts future trends based on statistical patterns and historical data (Predictive AI)



## Generative AI

A subset of ML that can create new content and ideas powered by large, pretrained models called foundation models (FMs)

# Challenges we are hearing from public sector customers



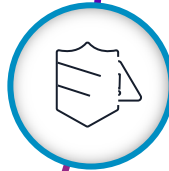
**Demand for government services is rising** while resources and capacity to deliver them **aren't keeping pace**



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



# Top AI/ML use cases for state and local government



**Speech  
and language**

SampleOutput.pdf (1 page)

Employment Application

This is a sample employment application form. and answer all questions.

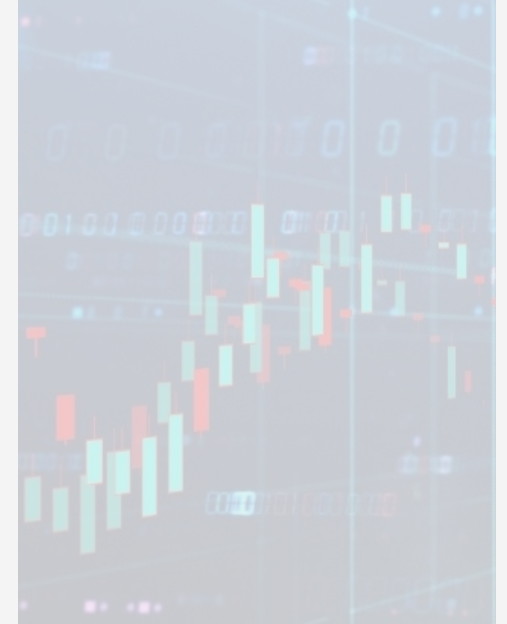
Personal Information	
Full Name:	Jane Doe
Phone Number:	555-0100
Home Address:	123 Any Street, Any Town, USA
Mailing Address:	Same as home address

Work History	
Current Company:	Any Company (2018-Current)
	Any Role
Company#1:	Previous Company # 1 (2014-2018)
	Previous Role # 1
Company#2:	Previous Company #2 (2010-2014)
	Previous Role # 2

**Intelligent document  
processing**



**Computer  
vision**



**Predictions  
and insights**

## 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



# 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



# Top generative AI use cases for state and local government



Speech  
and language

SampleOutput.pdf (1 page)

**Employment Application**

This is a sample employment application form. and answer all questions.

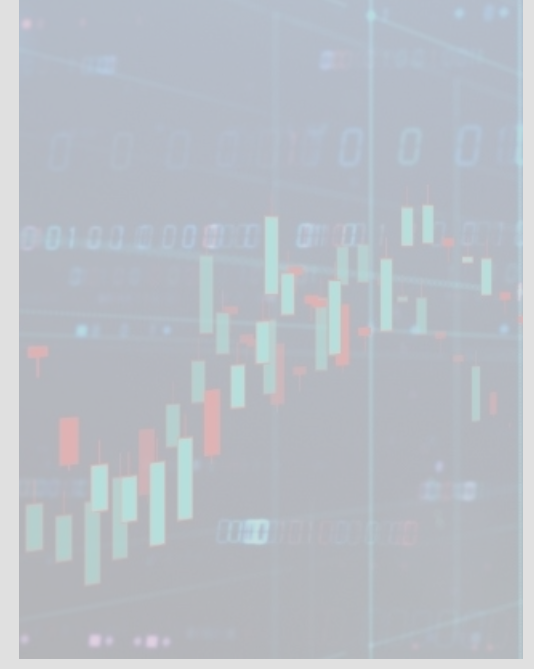
Personal Information	
Full Name:	Jane Doe
Phone Number:	555-0100
Home Address:	123 Any Street, Any Town, USA
Mailing Address:	Same as home address

Work History	
Current Company:	Any Company (2018-Current)
	Any Role
Company#1:	Previous Company # 1 (2014-2018)
	Previous Role # 1
Company#2:	Previous Company #2 (2010-2014)
	Previous Role # 2

Intelligent  
document  
processing



Computer  
vision



Predictions  
and insights

# 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

diagnosed in 1980 as having a 3.5 cm L. ... and then received ...  
-received neoadjuvant dose dense ... and ...  
- ... She had no ... on the ... and with residual disease 2.5 cm ... Nottingham grade 2 with ... she had ...  
- ... and then went on ... from ... when she was switched to ...  
- on evidence of progression until ... when she was found to have ... and ... found to have a ...  
- ... from peritoneal implants and ... Her ... showed ...  
- ... although tending about monthly ...  
- ... which showed a ... and ... with multiple ...  
- ... although tending about monthly ...  
- ... while off all ... for wash out and while waiting for ... she had worsening symptoms, going to requiring ... weekly (she had one on ... and a second on ... with a ... She was found to be ... positive ... removed 1600 mL found to be malignant ... and ... paracentesis removed 1500 mL fluid found to be malignant ...  
- ... moved to ... because of ...  
- ... removed 1L fluid found to be ...  
- ... and found 2.5 cm velvety ... on ...  
- ... (but not consistent with her primary ... from ...)  
- ... admitted with worsening ...  
- ... on ... she got ... held and on ... got ...  
- on admission had ... but ...  
- ... she had ... after ... She has ... and had first ... at ...  
- mother had ... but no other history of ... or ...  
- patient states ...  
2. ... improved with course of ...  
3. ... disease control gained weight to ... her ...  
4. ... because ...  
She had her ... and ...  
admitted. 5 ... and enterococcus faecalis growing in the urine and port.

REVIEW OF SYSTEMS:  
... has not had any ... since day 2 of her admission but has severe ... She is very ... and came in a wheelchair. She has had rapid accumulation of her ... and is very ... therefore not eating well. She also is having ... to the bathroom. ...  
... but mostly clear. No ... No ... Complete review of systems otherwise negative in detail. Patient is ...

Current Concerns:  
Ms. ... reports severe ... She reports ... but is still having ... around the house and to the bathroom. She did not get blood in the hospital and continue to drop. She also has ... from her ... She is ... until she had the ... held. She is also worried about the ... making her feel ... and not able to live her life. She does understand the ... 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

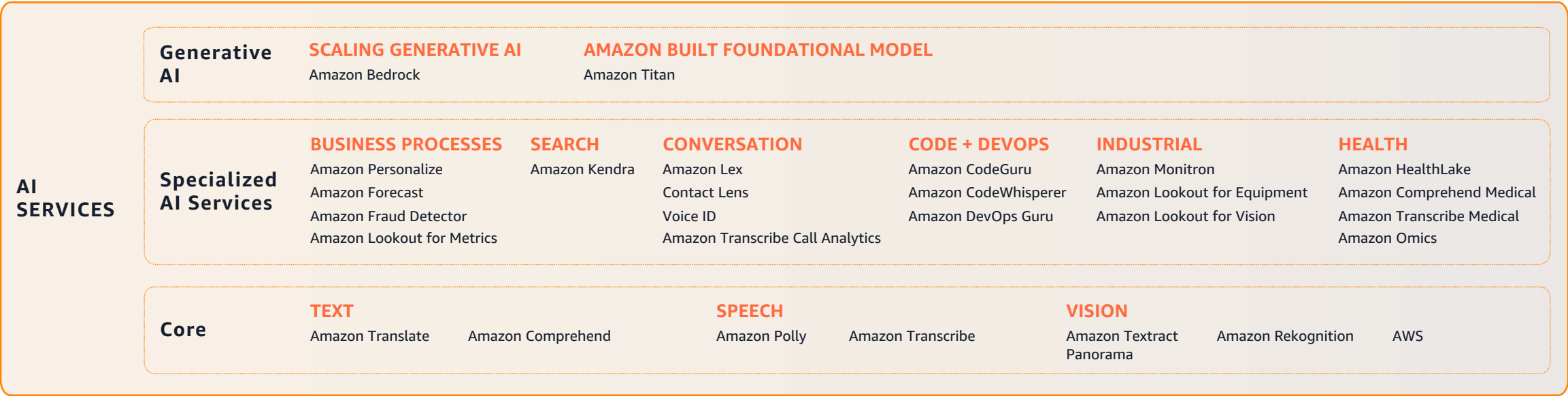
King County employees will focus on higher value, more satisfying work, and ultimately help the county realize its vision for connected communities, connected data, and connected government."



[This Photo](#) by Unknown Author is licensed under CC BY

# The AWS AI/ML Stack

Consumer



Tuner



Provider



# 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 use cases across industries

ENHANCE CUSTOMER EXPERIENCES	BOOST EMPLOYEE PRODUCTIVITY & CREATIVITY	OPTIMIZE BUSINESS PROCESSES
CHATBOTS	CONVERSATIONAL	DOCUMENT
VIRTUAL	SEARCH	PROCESSING
ASSISTANTS	SUMMARIZATION	DATA
CONVERSATION	CONTENT CREATION	AUGMENTATION
ANALYTICS	CODE GENERATION	FRAUD DETECTION
PERSONALIZATION	DATA TO INSIGHTS	PROCESS
		OPTIMIZATION

# However...

# Generative AI is not...

- The only way to leverage AI or ML
- Deterministic
- Perfect or error-free
- Sentient
- An independent decision maker



# 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

# 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

# Generative AI Stack

## APPLICATIONS THAT LEVERAGE LLMs AND OTHER FMs



Amazon Q  
Business



Amazon Q  
Developer



Amazon Q in  
QuickSight



Amazon Q in  
Connect

## TOOLS TO BUILD WITH LLMs AND OTHER FMs



**Amazon Bedrock**

Guardrails | Agents | Studio | Customization Capabilities | Custom Model Import

## INFRASTRUCTURE FOR FM TRAINING AND INFERENCE



GPUs



Trainium



Inferentia



SageMaker



UltraClusters

EFA



EC2 Capacity Blocks



Nitro



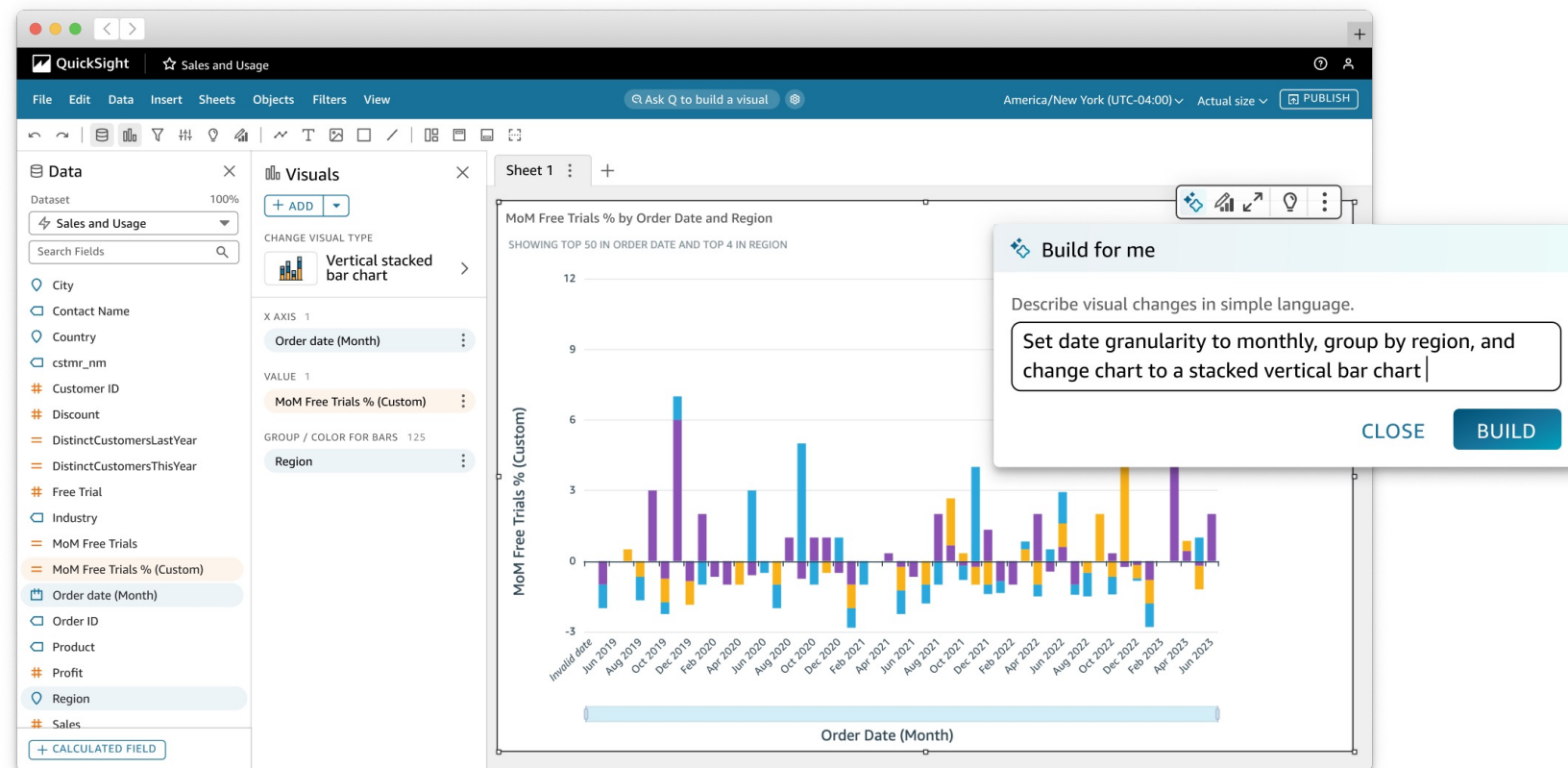
Neuron

# Demo



# Visual authoring in QuickSight

Use everyday language to generate and fine-tune visuals in seconds



Sheet 1 : +

YTD sales

\$189,948

Sales last quarter

\$81,908

Month-over-month sales

Sales for Jun 2023 increased by 97.41% (\$22,188.88) from \$22,779.48 to \$44,968.36.

Quarterly sales trend

Quarter	Sales (\$)
Q1 2019	75,000
Q2 2019	85,000
Q3 2019	140,000
Q4 2019	180,000
Q1 2020	70,000
Q2 2020	90,000
Q3 2020	125,000
Q4 2020	185,000
Q1 2021	95,000
Q2 2021	130,000
Q3 2021	140,000
Q4 2021	235,000
Q1 2022	115,000
Q2 2022	130,000
Q3 2022	200,000
Q4 2022	280,000
Q1 2023	80,000
Q2 2023	105,000

Product	Sales (Sum)
ContactMatcher	\$450.00K
FinanceHub	\$380.00K
Site Analytics	\$350.00K
Marketing Suite ...	\$240.00K
Big OI Database	\$200.00K
Data Smasher	\$180.00K
Alchemy	\$150.00K
Support	\$140.00K
Marketing Suite	\$120.00K
OneView	\$120.00K
SaaS Connector P...	\$100.00K
ChatBot Plugin	\$40.00K
SaaS Connector P...	\$30.00K
Storage	\$0.00K

SHOWING TOP 255 IN CITY AND TOP 2 IN SEGMENT

Segment

- Premium
- Standard

UNITED STATES

MEXICO

Mexico City

Vancouver

San Francisco

Los Angeles

Chicago

Toronto

New York

A Sankey diagram illustrating the distribution of 1000 employees across 10 business units and 3 regions. The business units are listed on the left, and the regions are listed on the right. The flows represent the number of employees in each category.

Business Unit	Region	Count
Finance	EMEA	100
Healthcare	EMEA	100
Manufacturing	EMEA	100
Communications	EMEA	100
Tech	EMEA	100
Energy	EMEA	100
Misc	EMEA	100
Retail	EMEA	100
Consumer Products	AMER	100
Transportation	AMER	100
Finance	APJ	100
Healthcare	APJ	100
Manufacturing	APJ	100
Communications	APJ	100
Tech	APJ	100
Energy	APJ	100
Misc	APJ	100
Retail	APJ	100
Consumer Products	APJ	100
Transportation	APJ	100

<b>Usage</b> MTD usage sessions	<b>Active customers</b> MTD	<b>Trial sign-ups</b> MTD	<b>Month-over-month usage</b> Total Usage Sessions for Jun 2023	<b>Product usage by industry</b>
------------------------------------	--------------------------------	------------------------------	--	----------------------------------



# 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

# How are customers leveraging AI/ML?

# Kern Community College District partners with AWS to build AI/ML-enabled Guided Pathways tool



## CHALLENGE

[Kern Community College District](#) (KCCD) district in California aimed to design and build a student-centered course planning tool to support the district's [Guided Pathways](#) model in order to optimize student course-taking patterns to decrease time to degree, improve enrollment forecasting, and focus limited staff resources where and when they are needed most

## SOLUTION

KCCD and AWS designed the Guided Pathways with AI/ML technologies to optimize advising and counseling resources, helping students with their needs while also maximizing budget and impact. The solution brought together data from siloed systems, and provided data analysis resources in a way that was timely and easy to access, understand, and make decisions based on student behaviors.

## OUTCOME

- ✓ KCCD leadership and practitioners gained dynamic insights into successful course-taking and student enrollment patterns used to provide focused interventions and understand course-offering needs and systemic blockers.
- ✓ Reduced manual burden on advising officers, allowing team members to devote time and resources to higher priority student services

# Student Advising Assistant



## CHALLENGE

The Faculty of Science Advising office at UBC aimed to improve the advising quality and student experience. They identified that answers to some student inquiries could be found in the Academic Calendar, but students found this information difficult to find or interpret. Existing advising system was impersonal and time-consuming, leaving many student questions unanswered.

## SOLUTION

Built with the UBC-CIC, ASK Cali chatbot was developed to provide 24/7 academic advising support to students. It uses natural language processing (Vicuna 7b LLM) on AWS EC2, to extract information from the Academic Calendar and other reliable UBC sources. AWS services: AWS Lambda, Amazon ECS, Amazon RDS, AWS ElasticBeanstalk

## OUTCOME

- ✓ ASK Cali is now advising questions from student, reducing advising staff workload and increasing user satisfaction by allowing quick, round-the-clock access to information on degree requirements

# Responsible AI



# 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

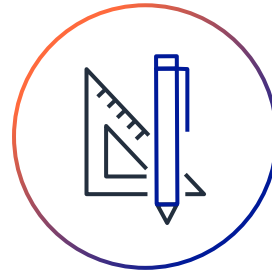
# Educating people on responsible AI is paramount



**AWS Machine Learning University (MLU)—free to all**



**Responsible use of machine learning guide**



**AWS AI and ML Scholarship program**



**Training and certification**

# New AWS Machine Learning University (MLU) Course on Fairness and Bias Mitigation

NEW

Free, public, hands-on training on fairness and bias

Taught by the same Amazon scientists that train our own AWS employees on machine learning

Over 9 hours of content available to everyone

>> **Get started today**



# What can you do now?

## Continue to explore responsible AI

Check out learning resources from AWS, including [training](#), [guidance](#), and [research](#).

## Educate your organization

Pass along what you learn to members of your team.

## Identify needs & consider risks

Think carefully about your organization's needs and where AI fits. Engage with [ML experts](#) at AWS to get started.

## Choose diverse talent

Strengthen your AI team by reflecting [diversity](#) within it.



# The Generative AI Innovation Center can help



## Art of the possible

Explore and experiment safely, reliably, and cost-effectively

## Responsible use

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



# Thank you!

## **Anthony Harvey**

Sr. Security Specialist Solutions Architect  
Amazon Web Services  
aharveyr@amazon.com  
Linkedin: anthonyrh

**Please complete the survey  
for this session**



**Artificial Intelligence**  
Generative AI for Public Sector