



Generative AI with AWS

“Why you need to get started now”

Snehanshu Bhaisare (Sneh)

Partner Solutions Architect
Australia and New Zealand

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



What we will discuss

- Innovation opportunities
- Data is your differentiator
- Generative AI with AWS
- What's next?



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



90% of Kiwi Companies will be AI-enabled by 2028

79% of workers want to develop
AI skills to advance their career

AI skills could boost pay between 30% - 41%

49% productivity payoff
with an AI-skilled workforce



2024 Report by Access Partnership and AWS
1,671 employees and 513 employers across New Zealand



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

2023

The Year of POCs



What is generative AI?

Is this secure?

Do I need to become a prompt engineer?

How do I choose a model?

Where do I get started?



What does this mean for my business?

What is a Foundation Model?



Which models should we try out?

What is FM?

What is a Large Language Model?



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

2024

The Year of Production

(FOR SOME)



How do I prioritize my projects?

How can I lower my costs?

How do I make this real?

What customization method should I use?



How I can I scale this?

Which models should I use?

Should I train my own model?

How do I manage risks?



How can we move faster?



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

Generative AI has potential to create significant business value



NEW EXPERIENCES

Create new innovative and engaging ways of interacting with your customers and employees



PRODUCTIVITY

Radically improve productivity across all lines of business



INSIGHTS

Extract insights and clear answers from all your corporate information, enabling faster and better decisions



CREATIVITY

Create new content and ideas, including conversations, stories, images, videos, and music



Generative AI use cases across industries

ENHANCE CUSTOMER EXPERIENCES

CHATBOTS

VIRTUAL ASSISTANTS

CONVERSATION

ANALYTICS

PERSONALIZATION

BOOST EMPLOYEE PRODUCTIVITY & CREATIVITY

CONVERSATIONAL

SEARCH

SUMMARIZATION

CONTENT CREATION

CODE GENERATION

DATA TO INSIGHTS

OPTIMIZE BUSINESS PROCESSES

DOCUMENT PROCESSING

DATA AUGMENTATION

FRAUD DETECTION

PROCESS OPTIMIZATION



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

How can you get ahead of this?



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



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

RAG provides relevant context

No context

Question:

"What metrics does AnyCompany track to measure project success?"

Answer:

" AnyCompany tracks project success by monitoring timelines, budget adherence, customer satisfaction, team performance, and deliverable quality."

Answer is based only on the model's **pretrained knowledge**

Context:

".... AnyCompany measures project success using three key metrics: on-time delivery (92%), budget adherence (5% variance), and client satisfaction (4.7/5) from post-delivery surveys. These metrics are tracked across major projects and reviewed quarterly....."

Question:

"What metrics does AnyCompany track to measure project success?"

Answer:

" AnyCompany tracks on-time delivery (92%), budget variance (5%), and client satisfaction (4.7/5), with quarterly reviews."

Answer is based on the **context**



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

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 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

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



What does **the future**
hold for generative AI?

Agents

Multi-modal

Multiple models

AI Policies & Standards

Comprehensive, customized training for every stage of your cloud journey



Skills assessment and personalized training plans

Our **Learning Needs Analysis** helps you understand where your gaps are and design a plan to target the AI and ML skills that are most important to your business goals.



Blended training designed by AWS experts

Digital courses, instructor-led classes, and game-based learning meet teams' diverse training needs.

AWS Skill Builder: 100+ learning resources on AI and ML, including generative AI. Interactive challenges in a secure sandbox environment.



Industry-recognized credentials

Validate expertise with certifications like **AWS Certified Machine Learning – Specialty**. Retain top talent by funding exam fees.

Get moving with generative AI

01

Select the
right use case

02

Engage your technology
partner via Westcon-Comstor

03

Get started on
your top use cases

AI is not going to replace
humans, but humans with AI
are going to replace humans
without AI

THANK YOU

