

Datathon Virtual AWS workshop

Generating the Future:

Introduction to Generative AI



Mauricio Mendoza (Mao)

Senior Solutions Architect
maomend@amazon.com



**BANQUE
NATIONALE
DU CANADA**



POLYFINANCES



Agenda

- Who is Mao (tronzeus)
- What is AI?
- What is Generative IA?
- Generative AI at Amazon
- Generative AI Sample use cases
Finance
- AWS Generative AI Stack
 - Sagemaker
 - Amazon Q
 - CodeWhisperer
 - Bedrock
- Extra Bonus (CheatCode)
 - Tools look steep learning
PartyRock
 - Training
 - How access to workshop

Connaissances en IA générative

Quelle est votre familiarité avec l'IA générative, là?



Quelle est votre familiarité avec l'IA générative, là?

4.9

Avez-vous déjà utilisé des outils d'IA générative comme Claude, ChatGPT, Perplexity, Etc.

8.0

Comment évalueriez-vous votre compréhension des algorithmes

5.1

Avez-vous déjà implémenté un modèle d'apprentissage automatique?

4.8

Quel est votre niveau en programmation?

5.9

C'est quoi ça?

Chu un vrai pro!

Mauricio Mendoza (he/his)

Background

- Maîtrise Génie Informatique
- (M.Eng.), Computer Engineering,
GPA 3.9/4 Jun 2015
- Chargé de cours Polytechnique Montréal
(INF8402, INF8405, INF6404A)
- Spécialiste du réseau et sécurité, Hewlett-Packard
- Cyber Security Architect, CGI
- Expert Security Architect CN
- Senior Solutions Architect – Security AWS

Lets go to
play!



What is AI?



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



Deep learning (DL)

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



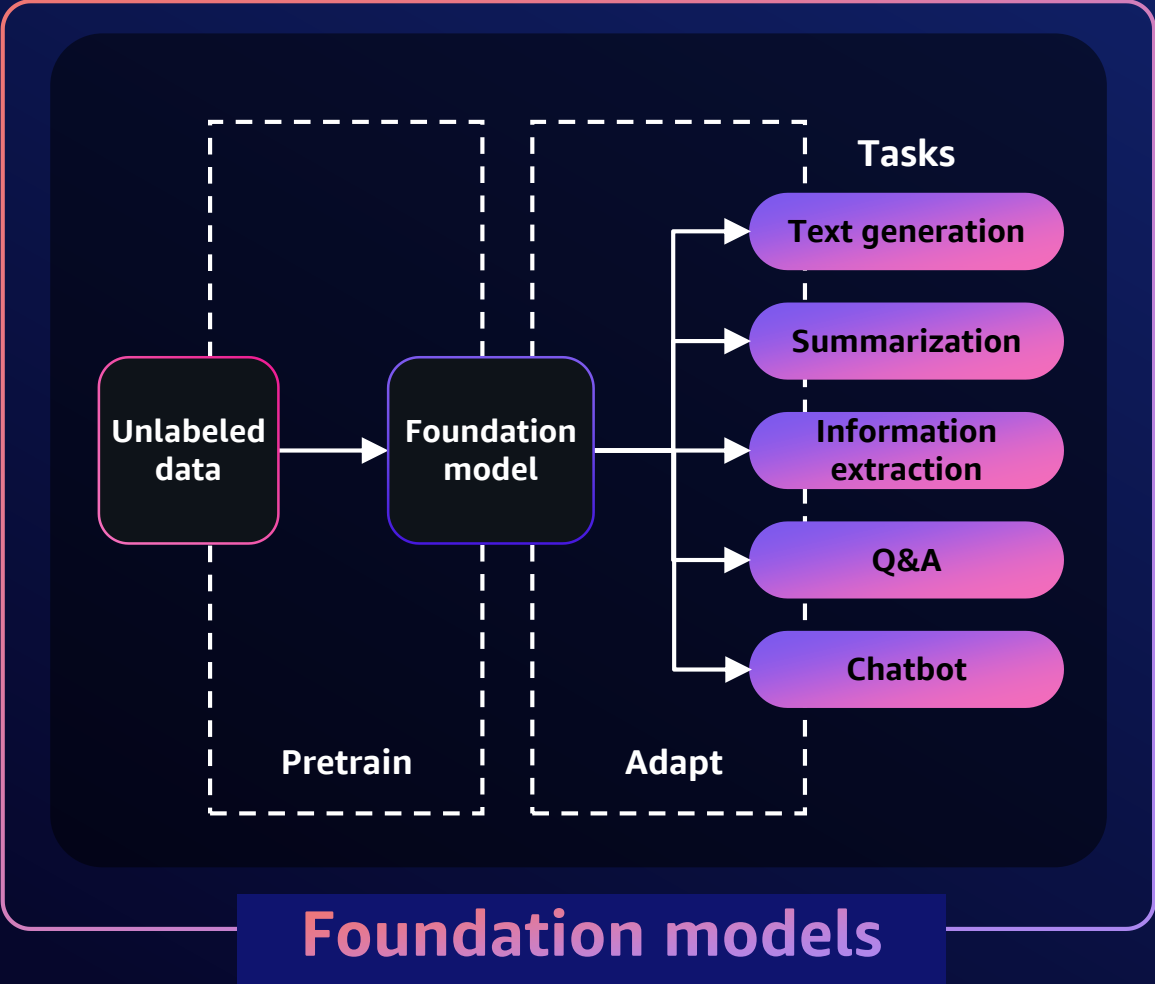
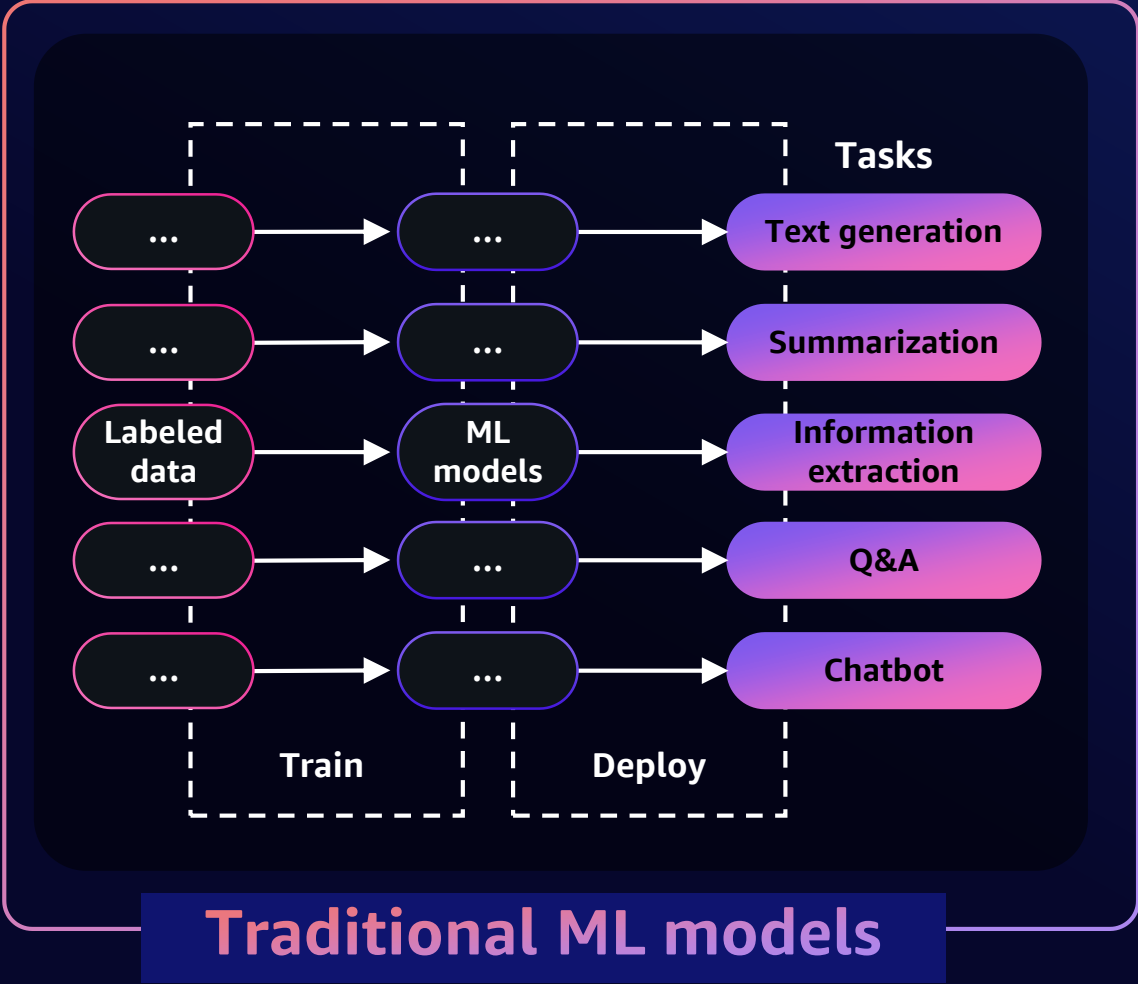
Generative AI

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

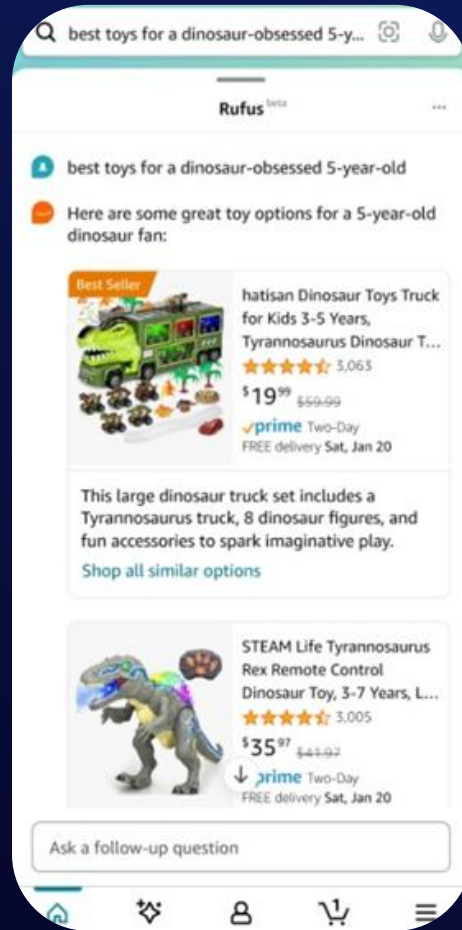
Generative AI

“Generative AI is a type of AI that can create new content and ideas, including conversations, stories, images, videos, and music. Like all AI, generative AI is powered by ML models—very large models that are pre-trained on vast amounts of data and commonly referred to as Foundation Models (FMs).”

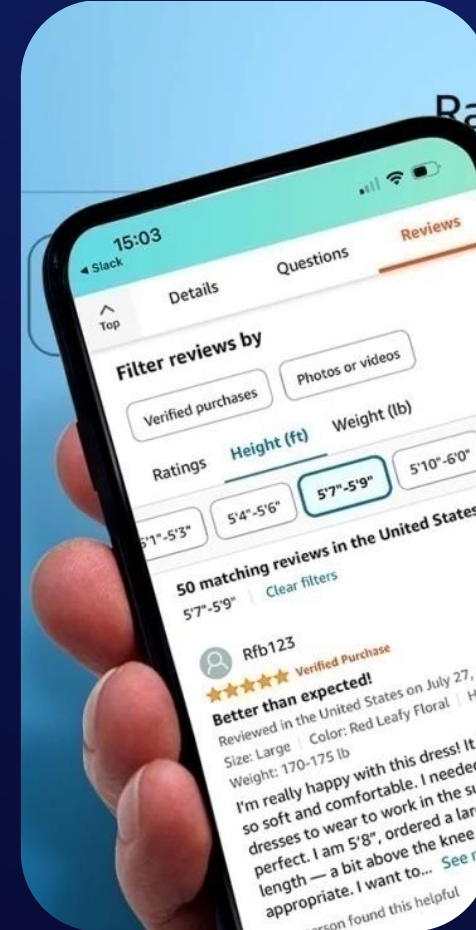
Traditional model vs foundation model



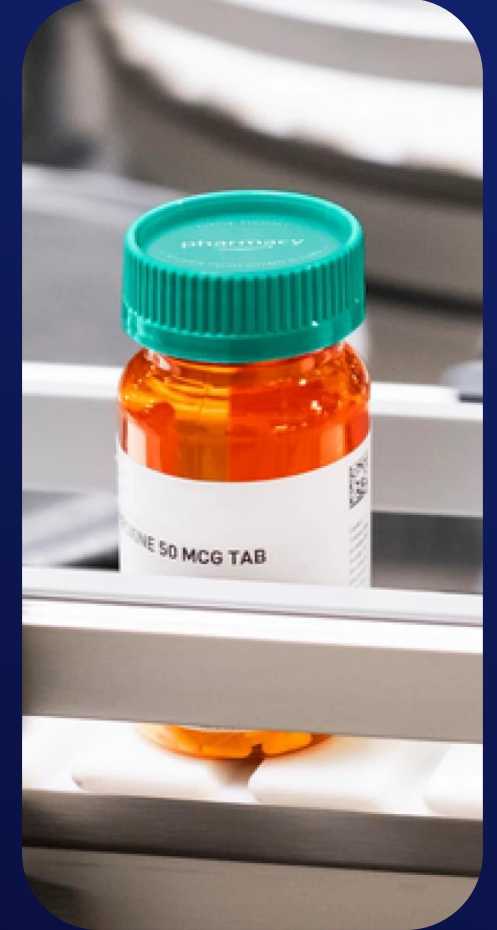
Generative AI Innovations at Amazon



Rufus
An expert
shopping assistant



**Customer
reviews**
On
Amazon.com



**Amazon
Pharmacy**
faster prescriptions and
more helpful support

GenAI can revolutionize the Finance organization to increase productivity & free up time to invest in strategic initiatives

From...



Work intense, **manual creation of financial reports...**



Time-consuming, **manual checks of accounting policies** with latest guidelines & standards ...



Hands-on search & extraction of relevant data from unstructured market data, competitor results, and macro-economic reports ...

To...



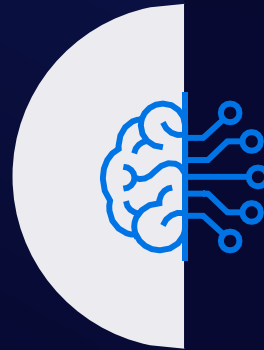
...**GenAI-created report drafts** (with humans in the loop as approvers) that enable teams to focus on value-adding activities



... **GenAI-enabled bots**, that utilize internal and external data, to answer policy-related questions in seconds



... **GenAI-enabled finance data platform**, that automatically ingests, processes and visualizes needed data to reduce time-to-insight



AWS can support implementation of several GenAI use cases within Finance function

AWS can support implementation of several GenAI use cases within Finance function

- Initial ideas for AWS supported GenAI use cases

Strategic leadership & support

Accounting & controlling

Stakeholder management & compliance

Working capital management

GenAI-enabled insight generation

Process, synthesize and extract analysis from large amounts of external & internal data (end-to-end solution)

Analytics co-pilot

Plug-in for employees analyzing vast amounts of raw data (e.g., from spreadsheets or data lakes)

Automated financial analysis & visualization

Automated financial analysis and visualization

Information summarizer

Synthesize summaries from text, slide decks, or video meetings

AI-based performance management

Answer performance-related questions, synthesize status and scenarios, understanding drivers and root cause of budget variances

Financial AI assistant

GenAI-powered finance bot to answer queries from executives and other decision-makers on public & private data

Automated expensing

Sorting and extracting expenses (incl. intelligent document recognition)

Advanced FinCrime detection

Identify potential fraud cases based on structured analysis of past data

Automated generation of investor reports

GenAI-enabled report and commentary creation for investors and broader stakeholders based on proprietary finance data

GenAI-supported drafting of tax documents

Drafts of tax-relevant reports (e.g., compensation agreements between parent and subsidiaries)

GenAI risk & compliance expert

Bot that gives easy access to latest public regulations across markets as well as internal process and compliance documents

AI assistant for accounts payable

GenAI bot to help facilitate procure to pay e.g., for employees and suppliers

Automated order-to-cash processes

AI-enablement to streamline order-to-cash processes (e.g., customer payment history risk assessment; pro-active customer credit limits based on market developments)



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



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



Build a data foundation to fuel your generative AI applications on AWS



Comprehensive

Comprehensive set of services for storing and querying structured, unstructured, and vector data



Integrated

Choices for integrating data including zero ETL so you can easily connect to all your data



Governed

End-to-end data governance capabilities, responsible AI, and regulating user interactions with LLMs

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

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

Generative AI Stack

APPLICATIONS THAT LEVERAGE LLMs AND 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



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



GPUs



Trainium



Inferentia



SageMaker



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

Guardrails for Amazon Bedrock

Implement safeguards customized to your application requirements and aligned to your responsible AI policies

Evaluate prompts and model responses for agents, knowledge bases, FMs in Amazon Bedrock, and custom or third-party FMs

Configure thresholds to filter harmful content, jailbreaks and prompt injection attacks

Define and disallow denied topics with short natural language descriptions

Remove personally identifiable information (PII) and sensitive information in gen AI apps

Filter hallucinations by detecting groundedness and relevance of model responses based on context

Demo “Bedrock Studio”

Lets create a Financial Virtual Chat assistant



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

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



AWS App Studio

The Fastest and Easiest Way
to Build Business Applications

Preview

Accelerate app development with a
generative AI-powered assistant

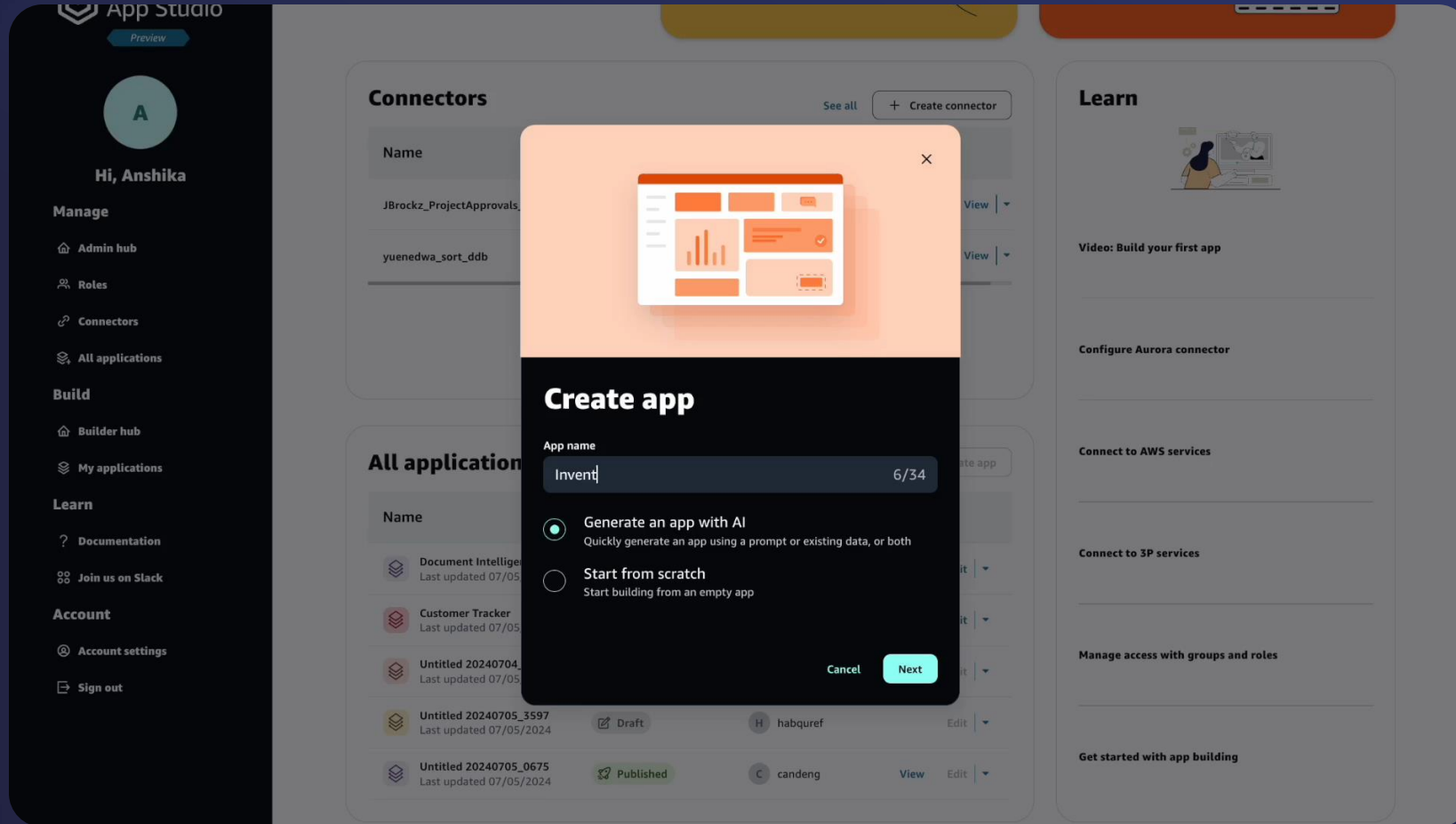
Deploy highly scalable applications
without operational overhead

Enhance security and governance
that delights IT

Save up to 80% vs.
comparable offerings



Accelerate application creation with a generative AI-powered assistant

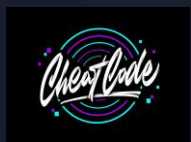


Describe the application you need, the generative AI-powered assistant will generate the full application, including UI, data models, and business logic.

Low-code environment with drag-and-drop canvas and AI helpers to customize UI, functionality, and roles, without professional software development skills.



Extra Bonus



My cheat sheet for the Generative AI hackathon is just a piece of paper that says 'Ask to AWS GenAI Tool'. Is that cheating or just being efficient?"

RETAIL
GEN_AI
HACKATH<>N

1- Use tools that do not require a steep learning curve.



What is PartyRock?

PartyRock, an Amazon Bedrock playground, is a shareable generative AI application-building playground

It's a hands-on, code-free application builder where you can build, share, and remix applications while playing with generative AI

Anyone can access PartyRock through its intuitive web-based UI



PartyRock

Everyone can build
AI apps.

Haiku Creator

Weird Tour Guide

Podcast Generator

Build your own app

App builder

Generate a 5-7-5 syllable haiku poem and a visual image based on user input.



User Input

winter

Generated Haiku LLM

Here is a 5-7-5 syllable haiku poem about winter:

Snowflakes falling down Blanketing
the ground in white Winter's cold
embrace

Visual LLM



Build your first application!
<https://partyrock.aws/>



2- Strategic recommendations



Strategic recommendations on generative AI

EYES WIDE OPEN

- Culture of continuous experimentation
- Prevent early dependencies

NO ONE SIZE FITS ALL

- Generative AI as an enhancement – not a replacement.
- Requirements in enterprises differ based on financial resources, security needs, governance and skills.
- Evaluate buy-vs-build

FLEXIBILITY IS KEY

- Innovation requires flexibility free from technical or contractual lock-ins.
- Infrastructure supporting 3rd-party Generative AI integration
- A breadth of services ensures long-term flexibility and business value

LISTEN AND ENABLE

- Work backward from your customers
- Listen to your domain experts
- Enable employees with the right set of tooling



3- Workshop Access



Getting started with the workshop

As a participant, you will have access to an AWS account shared with your Datathon team with any optional pre-provisioned infrastructure and IAM policies needed to complete this workshop.

The AWS account will only be available for the duration of this Datathon; you will lose access to the account thereafter

The optional pre-provisioned infrastructure will be deployed to a specific Region; check your workshop content to determine whether other Regions will be used

Be sure to review the terms and conditions of the event; do not upload any personal or confidential information in the account



Step 1: Sign-in on your laptop

The link to access will be bring on Datathon day



Sign in

Choose a preferred sign-in method

Email one-time password (OTP)

Enter your personal or corporate email to receive a one-time password

Login with Amazon

Login with your Amazon.com retail account

Amazon employee

Login with your Amazon Corporate account. Only for Amazon Employees.

Step 2: Review terms and join event

Terms and Conditions

Read and accept before joining the event

1. By using AWS Workshop Studio for the relevant event, you agree to the AWS Event Terms and Conditions and the AWS Acceptable Use Policy. You acknowledge and agree that are using an AWS-owned account that you can only access for the duration of the relevant event. If you find residual resources or materials in the AWS-owned account, you will make us aware and cease use of the account. AWS reserves the right to terminate the account and delete the contents at any time.
2. You will not: (a) process or run any operation on any data other than test data sets or lab-approved materials by AWS, and (b) copy, import, export or otherwise create derivate works of materials provided by AWS, including but not limited to, data sets.
3. AWS is under no obligation to enable the transmission of your materials through Event Engine and may, in its discretion, edit, block, refuse to post, or remove your materials at any time.
4. Your use of AWS Workshop Studio will comply with these terms and all applicable laws, and your access to AWS Workshop Studio will immediately and automatically terminate if you do not comply with any of these terms or conditions.

☒ I agree with the Terms and Conditions

Cancel

Previous

Join event



Step 3: Get started with the workshop

Workshop

Get started >

Title	Complexity level	AWS services	Topics
Data Mesh Immersion Day	300	AWS Glue, AWS Lake Formation, AWS Step Functions	Analytics

Description

Enable customers with hands-on experience to build a data mesh architecture on AWS step-by-step from scratch. **Instructor module: <https://gitlab.aws.dev/softicd/instructor-module-for-build-a-data-mesh-on-aws-workshop>**



4- Training



A man with dreadlocks, wearing a blue button-down shirt, is seated and gesturing with his hands while talking to a woman whose back is to the camera. They are in a modern office with large windows and other people working in the background. The text is overlaid on the left side of the image.

Technology enables
innovation, but
**people make it
happen.**

AWS Training and Certification
equips organizations with the
knowledge and skills to build
and innovate in the AWS Cloud
to **maximize business impact.**



Preselected free courses over AWS technologies 1/2

Generative AI Learning Plan Decision Makers (3 Hours 4 Courses)

- This learning plan is designed to introduce generative AI to business and technical decision makers.
- The digital training included will provide an overview of generative AI, as well as the approach to plan a generative AI project and build a generative AI-ready organization.



Machine Learning Plan (32 Hours 6 Courses)

- This Learning Plan is designed to help Data Scientists and Developers integrate machine learning (ML) and artificial intelligence (AI) into tools and applications.
- The training will expose you to the broadest and deepest set of machine learning services and supporting cloud infrastructure.
- This Learning Plan can also help prepare you for the AWS Certified Machine Learning - Specialty certification exam.



Preselected free courses over AWS technologies 2/2

(Included on the Learning Plans)
Introduction to Machine Learning: Art of the Possible
(30 Minutes)

- **Fundamental:** After taking this set of courses, you'll understand how artificial intelligence (AI) led to machine learning (ML), which then led to deep learning (DL).



Amazon Bedrock - Getting Started
(1 Hour)

- In this Getting Started course, you will learn about the benefits, features, typical use cases, technical concepts, and cost of Amazon Bedrock.
- You will also review an architecture that uses Amazon Bedrock, along with other Amazon Web Services (AWS) offerings, to build a chatbot solution.
- Through a guided tutorial consisting of a narrated video, step-by-step instructions, and transcript, you will try Amazon Bedrock in your AWS account.



Survey



https://amazonmr.au1.qualtrics.com/jfe/form/SV_2t6hoa9kUyA5ZzM



Thank you!

Mauricio M

maomend@amazon.com

