



# Innovating with Generative Artificial Intelligence (GenAI)

Saúl J. Torres Salazar

SOLUTIONS ARCHITECT | CARIBBEAN

AMAZON WEB SERVICES | PUBLIC SECTOR

Innovation can  
**transform industries**



GENERATIVE AI

# The tipping point for **Generative AI**



MASSIVE PROLIFERATION  
OF DATA

AVAILABILITY OF  
SCALABLE COMPUTE  
CAPACITY

MACHINE LEARNING  
INNOVATION

# AI

MACHINE  
LEARNING

**SIMPLE**

INPUTS



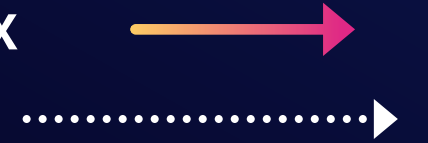
**SIMPLE**

OUTPUTS

DEEP  
LEARNING

**COMPLEX**

INPUTS



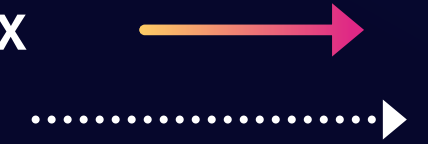
**SIMPLE**

OUTPUTS

FOUNDATION  
MODELS

**COMPLEX**

INPUTS



**COMPLEX**

OUTPUTS





# Generative AI is powered by foundation models

Pretrained on vast amounts of unstructured data

---

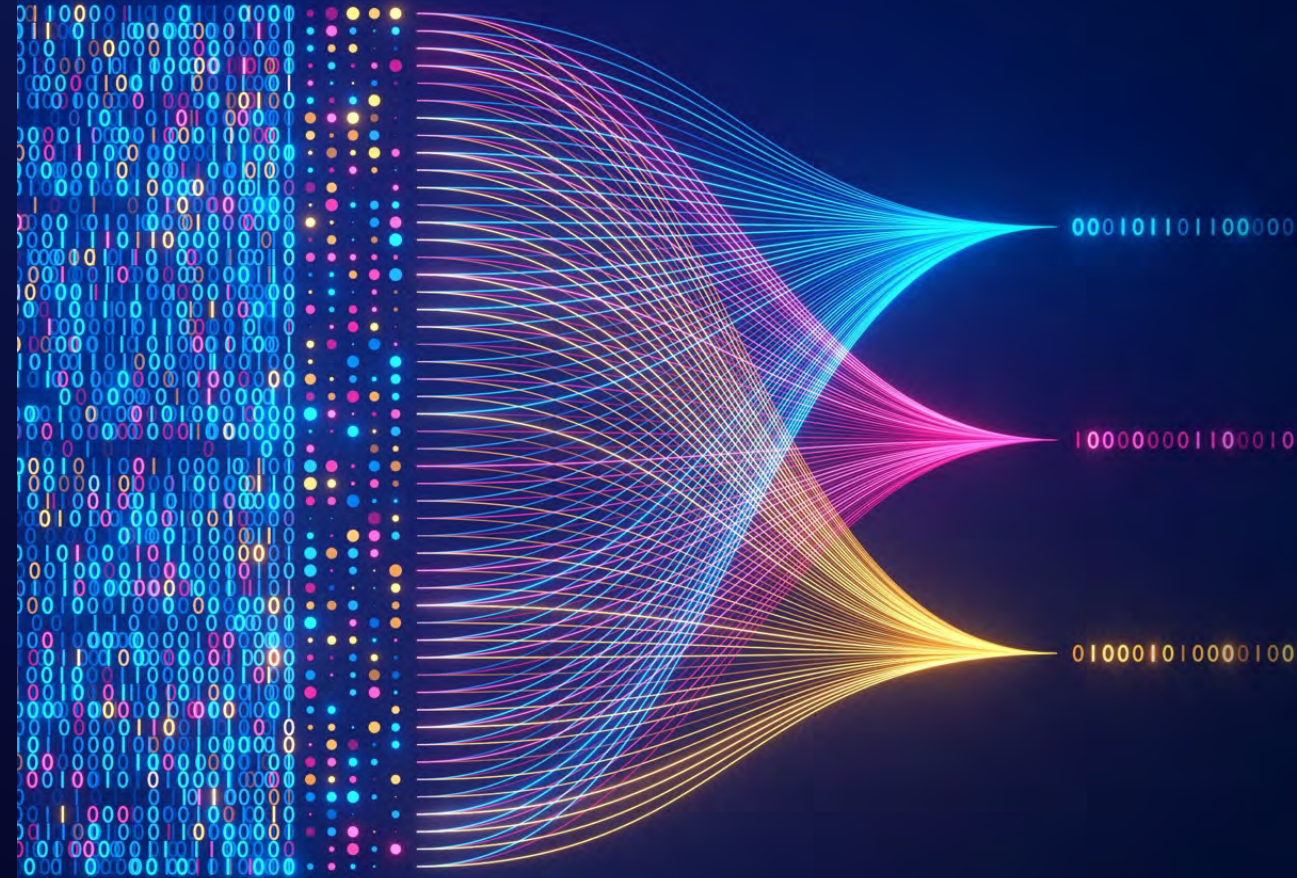
Contain large number of parameters that make them capable of learning complex concepts

---

Can be applied in a wide range of contexts

---

Customize FMs using your data for domain specific tasks





**Enhance Customer  
Experiences**

**Boost employee  
productivity & creativity**

**Optimize business  
processes**





CHATBOTS

VIRTUAL ASSISTANTS

**Enhance Customer  
Experiences**

CONVERSATION ANALYTICS  
PERSONALIZATION

CONVERSATIONAL SEARCH

SUMMARIZATION

**Boost employee  
productivity & creativity**

CONTENT CREATION  
CODE GENERATION

DATA TO INSIGHTS

DOCUMENT PROCESSING

DATA AUGMENTATION

**Optimize business  
processes**

CYBERSECURITY  
PROCESS OPTIMIZATION



**Healthcare &  
Life Sciences**



**Industrial &  
manufacturing**



**Financial  
services**



**Retail**



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.





AMBIENT DIGITAL SCRIBE

INTERPRET MEDICAL IMAGES

DRUG DISCOVERY

ENHANCE CLINICAL TRIALS

RESEARCH REPORTING

## Healthcare & Life Sciences



PRODUCT DESIGN

OPERATIONAL EFFICIENCY

MAINTENANCE ASSISTANTS

SUPPLY CHAIN OPTIMIZATION

EQUIPMENT DIAGNOSTICS

## Industrial & manufacturing



PORTFOLIO MANAGEMENT

FINANCIAL DOCUMENTATION

INTELLIGENT ADVISORY

PRODUCT INNOVATION

FINANCIAL DOCUMENTATION

## Financial services



PRICING OPTIMIZATION

VIRTUAL TRY-ONS

PRODUCT REVIEW SUMMARIES

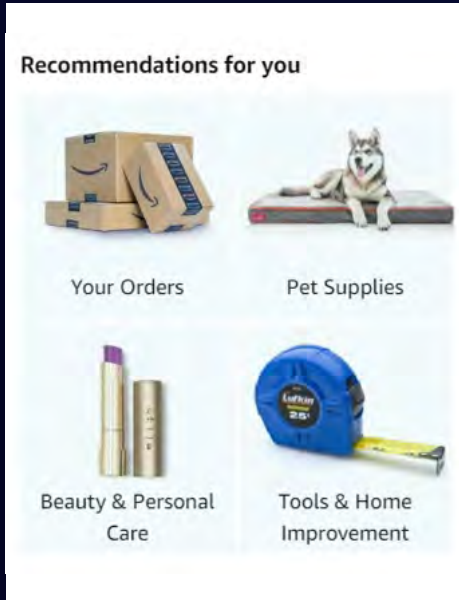
PRODUCT DESCRIPTIONS

PERSONALIZATION

## Retail



# ML innovation is in Amazon's DNA



**4,000 products  
per minute** sold  
on Amazon.com



**1.6M packages**  
every day



**Billions** of Alexa  
interactions each week



First Prime Air delivery  
on **December 7, 2016**

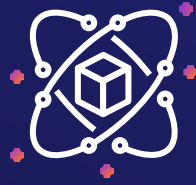




More than **100,000 customers** use AWS for ML



Everything you  
need to  
accelerate  
your generative  
AI journey



## **Easiest way to build**

with leading foundation models



## **Differentiate with your data**

in a secure and private environment



## **Increase productivity**

with generative AI applications and services

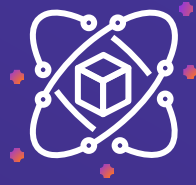


## **Most performant, low cost**

infrastructure to scale generative AI



Everything you  
need to  
accelerate  
your generative  
AI journey



## **Easiest way to build**

with leading foundation models



## **Differentiate with your data**

in a secure and private environment



## **Increase productivity**

with generative AI applications and services



## **Most performant, low cost**

infrastructure to scale generative AI

NOW GENERALLY AVAILABLE

# Amazon Bedrock

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



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



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



Privately customize FMs using your organization's data



# Amazon Bedrock

Choice of foundation models

**AI21labs**

**ANTHROPIC**

**co:here**

**∞ Meta AI**

**stability.ai**

**amazon**

## **JURASSIC-2**

Multilingual LLMs for text generation in Spanish, French, German, Portuguese, Italian, and Dutch

## **CLAUDE 2**

LLM for conversations, question answering, and workflow automation based on research into training honest and responsible AI systems

## **COMMAND**

Text generation model for business applications like summarization, copywriting, dialog, extraction, and question answering

## **LLAMA 2**

Pre-trained and fine-tuned LLMs for natural language tasks like question answering and reading comprehension

## **SDXL 1.0**

Generation of unique, realistic, high-quality images, art, logos, and designs

## **AMAZON TITAN**

Text summarization, generation, classification, open-ended Q&A, information extraction, embeddings and search



# Driving innovation with Amazon Bedrock

**Chegg**

lonely  planet

 **cimpres**

**PHILIPS**

**IBM** |  The Weather Company

**nexx:iot**

 **Sun Life**

*Neiman Marcus*

 **RYANAIR**

**hellmann**   
WORLDWIDE LOGISTICS

  
**WPS Office**  
Make It Simple

 **twilio**

**BRIDGEWATER**

 **Showpad**

  
**coda**

**Booking.com**



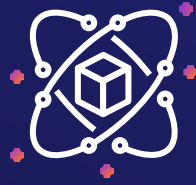
A photograph of an American football game in progress, showing players from both teams (one in orange and black, the other in white and black) lined up at the line of scrimmage. The image is dark and moody, with stadium lights visible in the background. The word "FOX" is overlaid in large, white, bold letters in the center of the image.

# FOX



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Everything you  
need to  
accelerate  
your generative  
AI journey



## **Easiest way to build**

with leading foundation models



## **Differentiate with your data**

in a secure and private environment



## **Increase productivity**

with generative AI applications and services



## **Most performant, low cost**

infrastructure to scale generative AI

Your data is  
**your differentiator**





# Privately customize foundation models using your organization's data



## Fine-tune

### PURPOSE

Maximizing accuracy for specific tasks

### DATA NEED

Small number of labeled examples

# Keeping your data private and secure



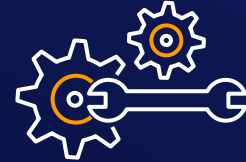
---

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



---

All data is encrypted at rest and PrivateLink support allows access to Bedrock APIs via customer's VPC endpoints



---

Customized foundation models and the customer-specific data that trains them remain private

# Build a data strategy to fuel your generative AI applications



## 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



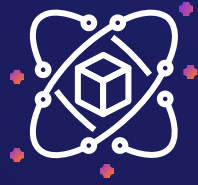
## Governed

---

End-to-end data governance capabilities



Everything you  
need to  
accelerate  
your generative  
AI journey



## **Easiest way to build**

with leading foundation models



## **Differentiate with your data**

in a secure and private environment



## **Increase productivity**

with generative AI applications and services



## **Most performant, low cost**

infrastructure to scale generative AI

# Generative BI capabilities in **Amazon QuickSight**

New FM-powered capabilities for business users  
to extract insights, collaborate and visualize data

PUBLIC PREVIEW



Easily author, fine-tune and add  
visuals to dashboards

COMING SOON



Automatically generate data  
stories with natural language



GENERALLY AVAILABLE

# Amazon CodeWhisperer

Build apps faster and more  
securely with an AI coding  
companion



Generate code suggestions  
in real-time



Scan code for hard-to-find  
vulnerabilities



Flag code that resembles open-  
source training data or filter by  
default

**FREE FOR INDIVIDUAL TIER**



CODEWHISPERER

Using  
CodeWhisperer  
to increase  
productivity

accenture

KOCH

Infosys

HCLTech

SmugMug

publicis  
sapient

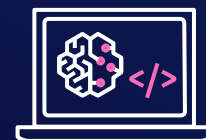
amazon ads



COMING SOON

# Amazon CodeWhisperer customization capability

Generate code recommendations based on  
your internal codebases



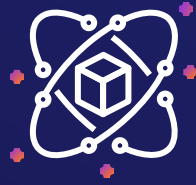
Generates organization-specific code  
recommendations based on their  
internal codebase



Will be available to customers as  
part of a new CodeWhisperer  
Enterprise Tier



Everything you  
need to  
accelerate  
your generative  
AI journey



## **Easiest way to build**

with leading foundation models



## **Differentiate with your data**

in a secure and private environment



## **Increase productivity**

with generative AI applications and services



## **Most performant, low cost**

infrastructure to scale generative AI



# Deep investments **in global infrastructure**



Broad choice of ML  
accelerators

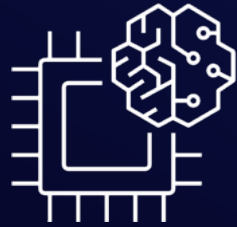


High performance,  
low-cost ML infrastructure



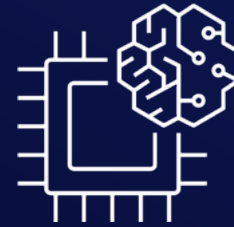
10+ years of silicon  
innovation

# Purpose-built accelerators for generative AI



## AWS Trainium

Up to 50% savings on training costs  
over comparable Amazon EC2 instances



## AWS Inferentia2

Up to 40% better price performance  
than comparable Amazon EC2 instances

# Start your generative AI journey today



Select the right use case

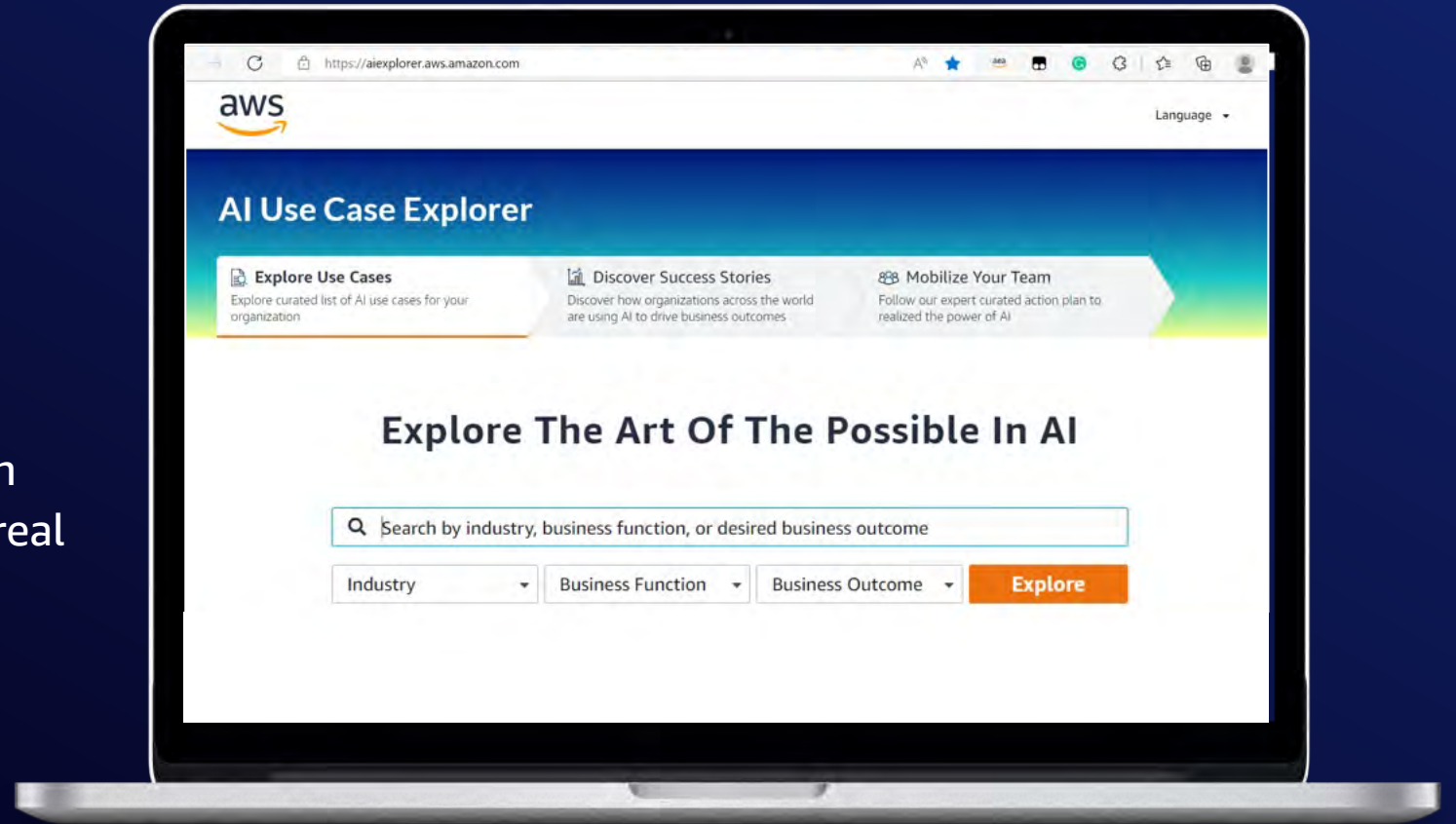






# AI Use Case Explorer

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



**aiexplorer.aws.amazon.com**



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Start your generative AI journey today



1

Select the right use case

2

Empower your teams  
through a variety of training  
opportunities

**SKILLBUILDER.AWS**

600+ digital  
courses

The screenshot displays the AWS Skill Builder interface. At the top, navigation tabs include 'COURSES AND LEARNING PLANS', 'SUBSCRIPTIONS', 'CLASSROOM TRAINING', 'AWS CERTIFICATION', and 'AWS PARTNER TRAINING'. The main header features the 'AWS Skill Builder' logo and the tagline 'Your learning center to build in-demand cloud skills'. Below this is a search bar with a 'FILTERS' button and a 'NEWEST TO OLDEST' sort option. The 'Featured content' section, titled 'Browse through the most popular courses offered in AWS Skill Builder', shows five course cards: 'Cloud for CEOs' (FREE, 10m 00s, 5.0 stars, Digital training), 'Build with Amazon MemoryDB for Redis' (FREE, 1h 00m, 5.0 stars, Digital training), 'AWS Best Practices for Hybrid Cloud Adoption' (FREE, 1h 00m, 5.0 stars, Digital training), 'AWS for Games Learning Plan: Cloud Game Development' (FREE, 12 courses, 11h 40m, Learning Plan), and 'AWS Certified Developer - Associate Official Practice Question Set (DVA-C02)' (FREE, 45m 00s, 5.0 stars, Digital training). The 'Featured subscription content' section, titled 'Get the most out of your AWS Skill Builder subscription with these training offerings available only with the subscription.', shows five more cards: 'Migrate a monolith web application to AWS using Application Migration...' (FREE, 1h 00m, 5.0 stars, Digital training), 'Challenge AWS Game Skills: Basic' (FREE, 1h 00m, 5.0 stars, Digital training), 'Exam Prep: AWS Certified Solutions Architect - Professional (SAP-C02) (with...)' (FREE, 1h 00m, 5.0 stars, Digital training), 'Exam Prep: AWS Certified Advanced Networking - Specialty (ANS-C01) (with...)' (FREE, 1h 00m, 5.0 stars, Digital training), and 'Industry Quest: Financial Services' (FREE, 1h 00m, 5.0 stars, Digital training).



# Hands-on generative AI training opportunities

Learn the fundamentals  
of generative AI for  
real-world applications

The screenshot shows the Coursera interface for the course "Generative AI with Large Language Models". The left sidebar contains navigation links: "Course Material" (with a sub-menu for Week 1, Week 2, and Week 3), "Grades", "Notes", "Messages" (with a blue notification bubble), and "Course Info". The main content area displays the course title and a search bar. Below this, the "Week 1" section is expanded, showing progress indicators for videos (1h 56m left), readings (45 min left), and graded assessments (2 left). The "Introduction to LLMs and the generative AI project lifecycle" is highlighted as the current topic, with 1 graded assessment left. A list of course items follows: "Course Introduction" (Video, 6 min), "Contributor Acknowledgments" (Reading, 10 min), "Introduction - Week 1" (Video, 5 min), and "Generative AI & LLMs" (Video, 4 min). A "Get started" button is visible next to the first video. At the bottom, a row of five course cards is partially visible, including "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...", and "Industry Quest: Financial Services".

**coursera** Search in course Search

**Generative AI with Large Language Models**

▼ **Course Material**

- Week 1
- Week 2
- Week 3

**Grades**

**Notes**

**Messages** 2

**Course Info**

▼ **Week 1**

1h 56m of videos left 45 min of readings left 2 graded assessments left

Generative AI use cases, project lifecycle, and model pre-training

▼ **Show Learning Objectives**

▼ **Introduction to LLMs and the generative AI project lifecycle** 1 graded assessment left

- Course Introduction Video • 6 min **Get started**
- Contributor Acknowledgments Reading • 10 min
- Introduction - Week 1 Video • 5 min
- Generative AI & LLMs Video • 4 min

[IMPORTANT] Have questions, issues or ideas? Join our Community!

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



# Start your generative AI journey today

1

Select the right use case

2

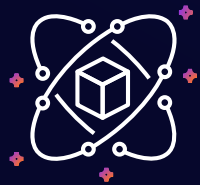
Empower your teams  
through a variety of training  
opportunities

3

Get started on a PoC for  
your top use cases



# Everything you need to accelerate **your generative AI journey**



Easiest and most  
secure way to build  
generative AI  
applications



Data as your  
differentiator and  
strategic asset for  
generative AI



Most performant, low  
cost infrastructure for  
generative AI



Generative AI  
applications to  
enhance  
productivity



# Thank you!

Saúl J. Torres Salazar

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

787-692-4668

Linkedin

