
aws

Salim Tütüncü

SENIOR SOLUTIONS ARCHITECT - ANALYTICS & AI/ML

AMAZON WEB SERVICES

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)

AI

MACHINE LEARNING

DEEP LEARNING

FOUNDATION MODELS

SIMPLE

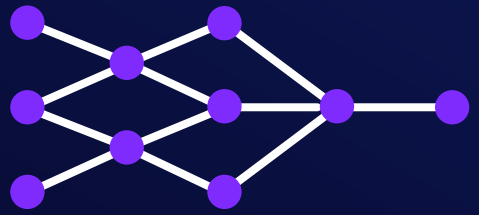
INPUTS

SIMPLE

OUTPUTS

COMPLEX

INPUTS

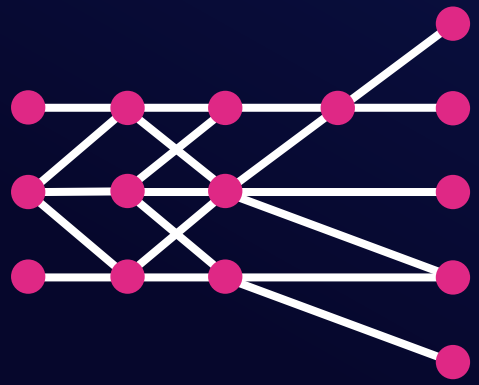


SIMPLE

OUTPUTS

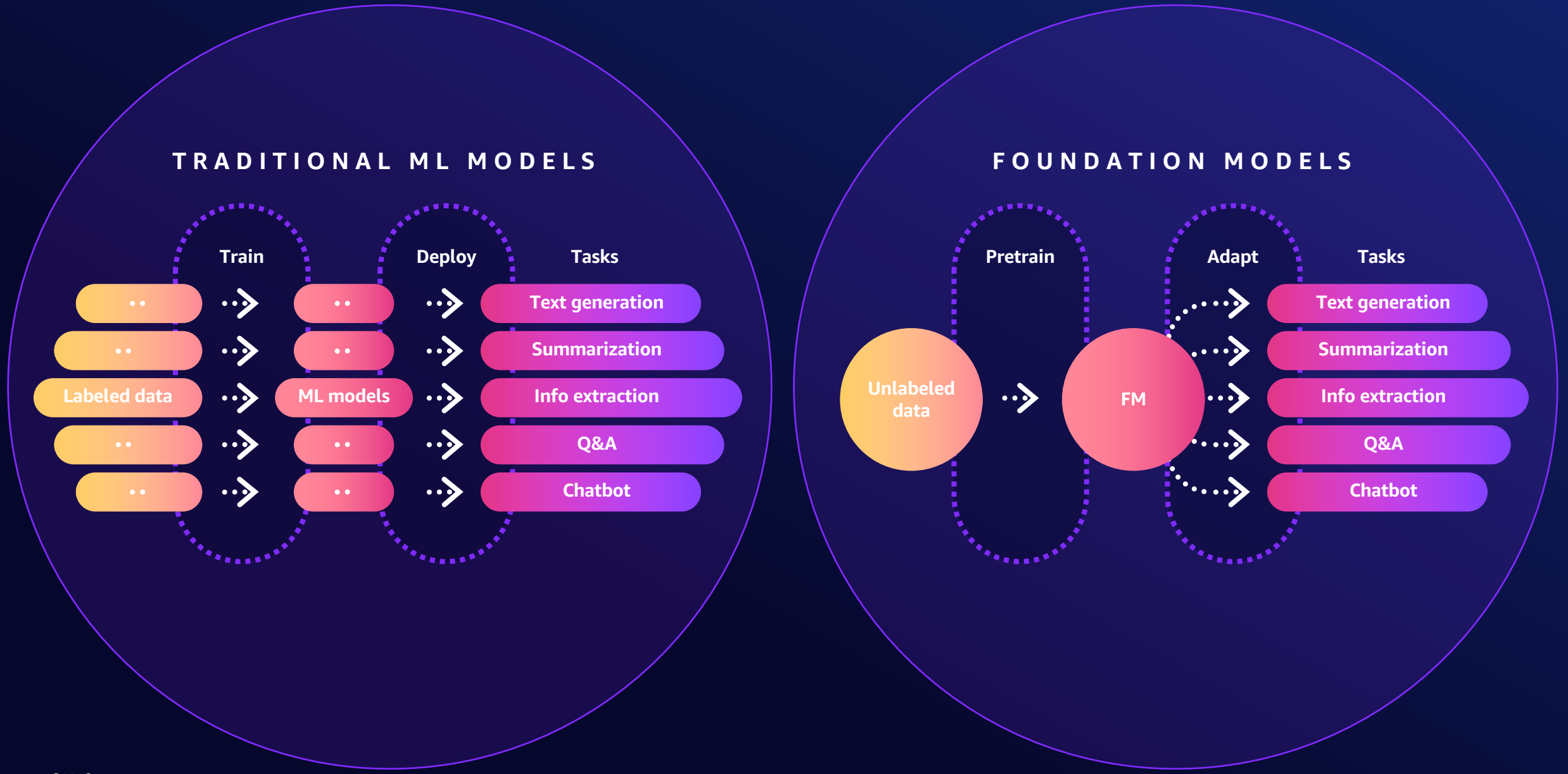
COMPLEX

INPUTS



COMPLEX

OUTPUTS



Innovation can **transform industries**

The tipping point for **Generative AI**

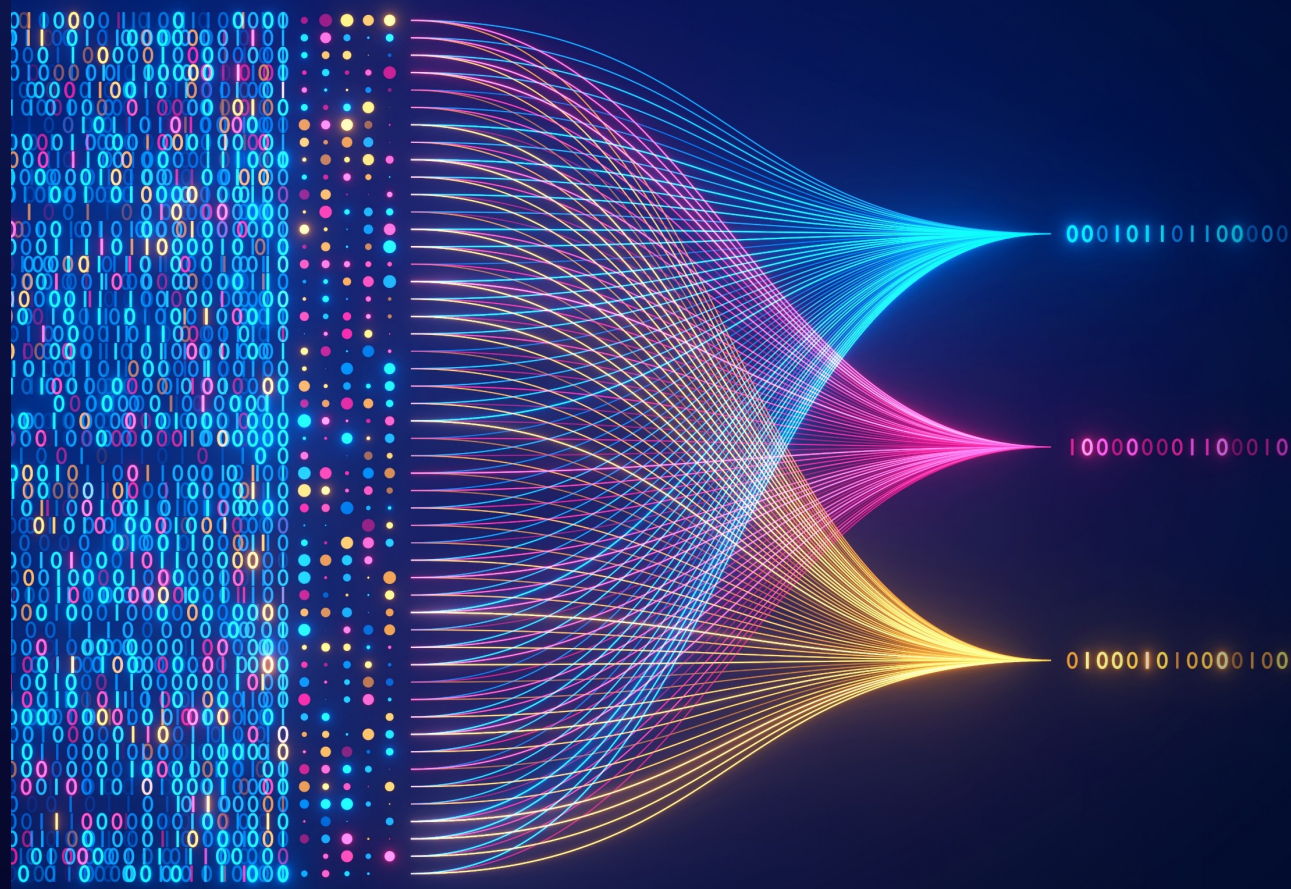
Generative AI is powered by foundation models

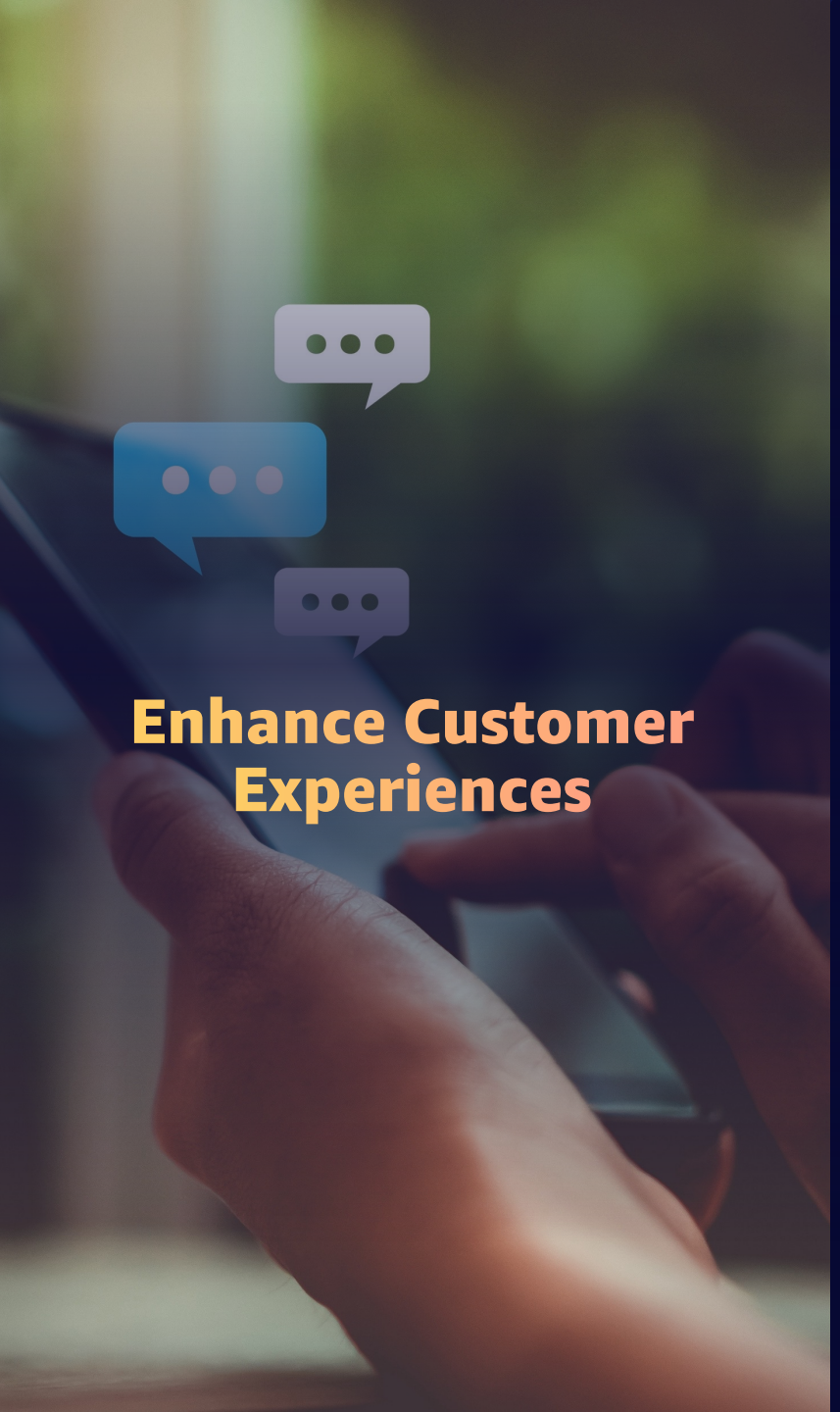
Pretrained on vast amounts of unstructured data

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

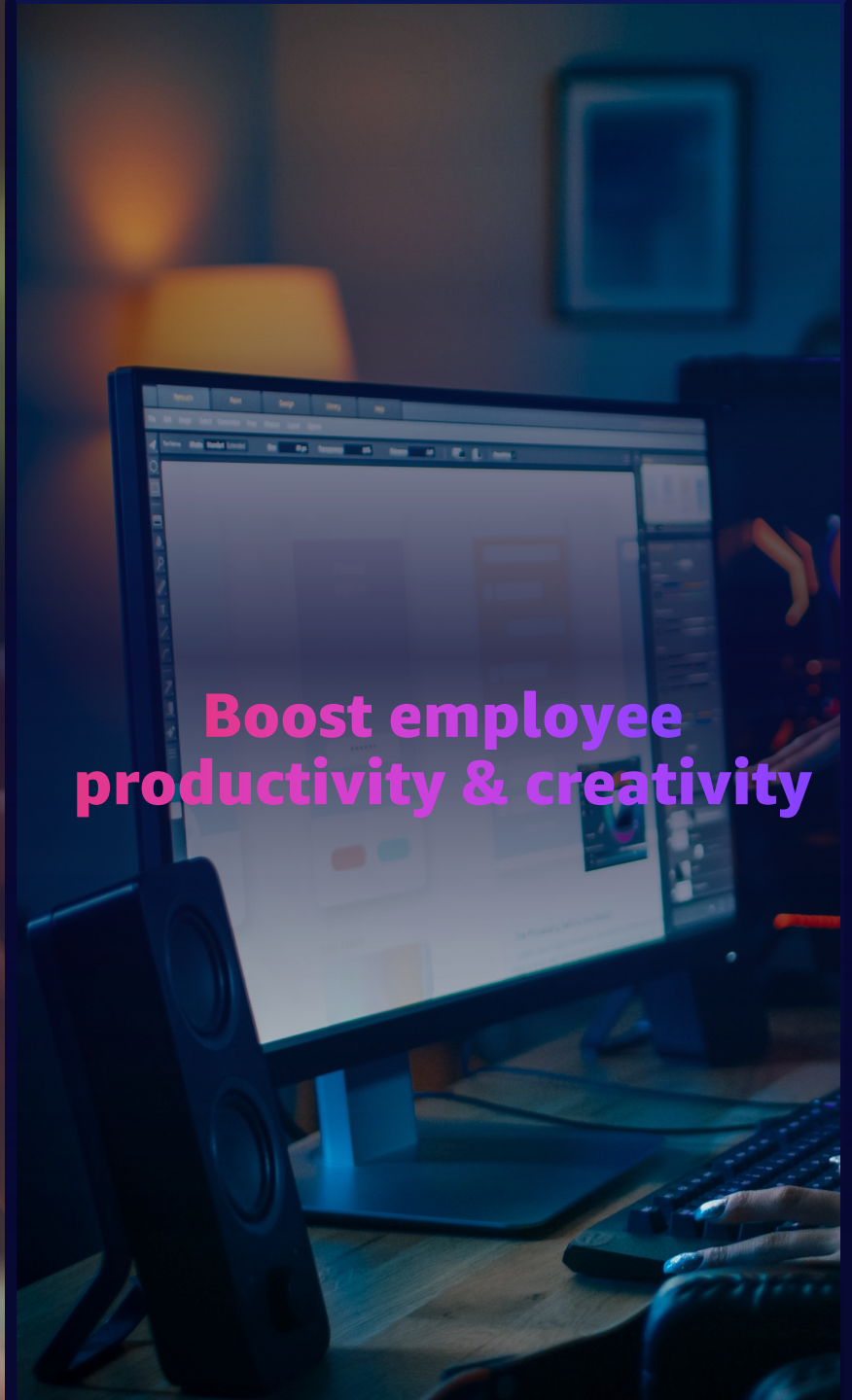
Can be applied in a wide range of contexts

Customize FMs using your data for domain specific tasks



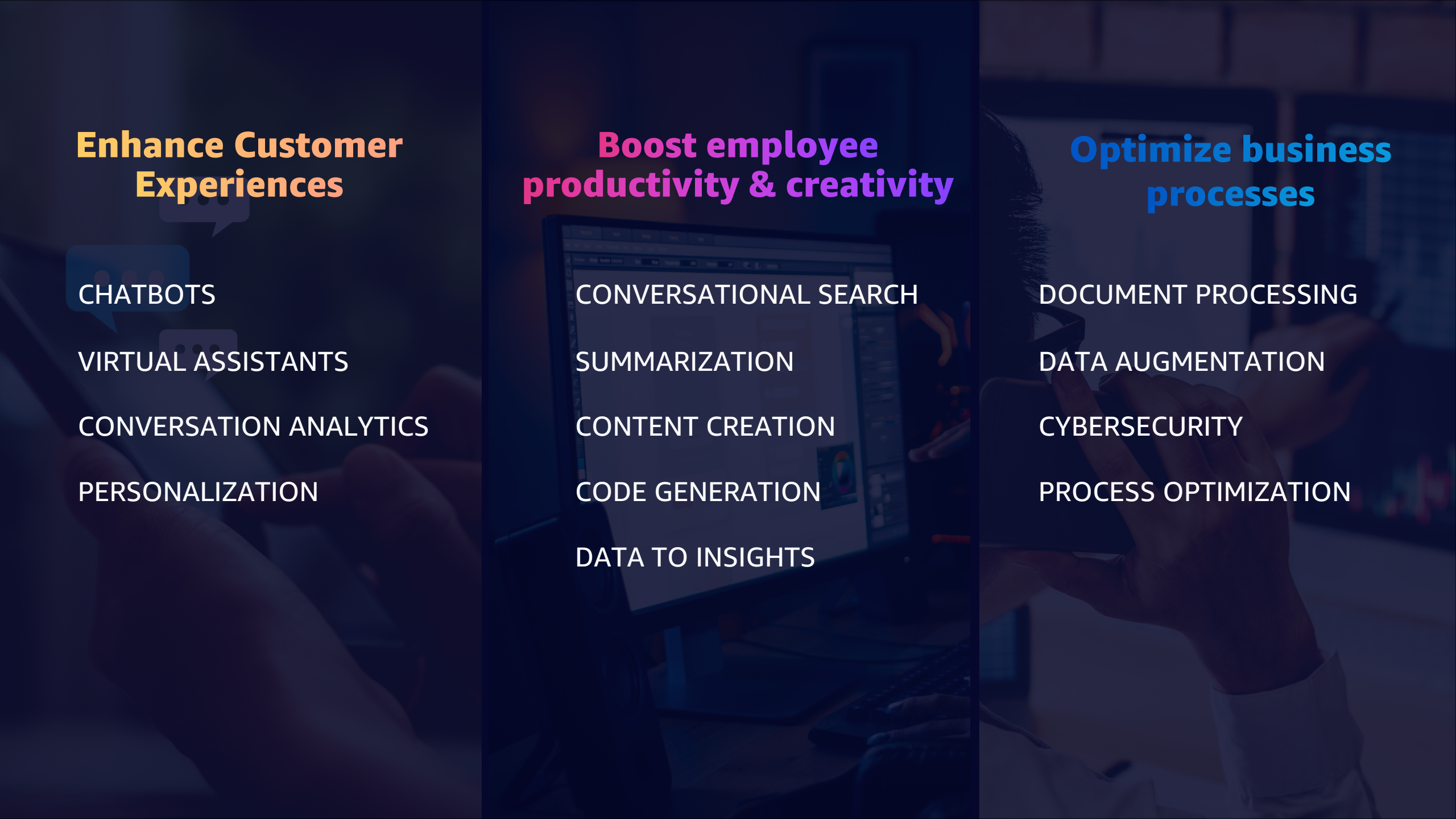


Enhance Customer Experiences



Boost employee productivity & creativity

Optimize business processes



Enhance Customer Experiences

CHATBOTS

VIRTUAL ASSISTANTS

CONVERSATION ANALYTICS

PERSONALIZATION

Boost employee productivity & creativity

CONVERSATIONAL SEARCH

SUMMARIZATION

CONTENT CREATION

CODE GENERATION

DATA TO INSIGHTS

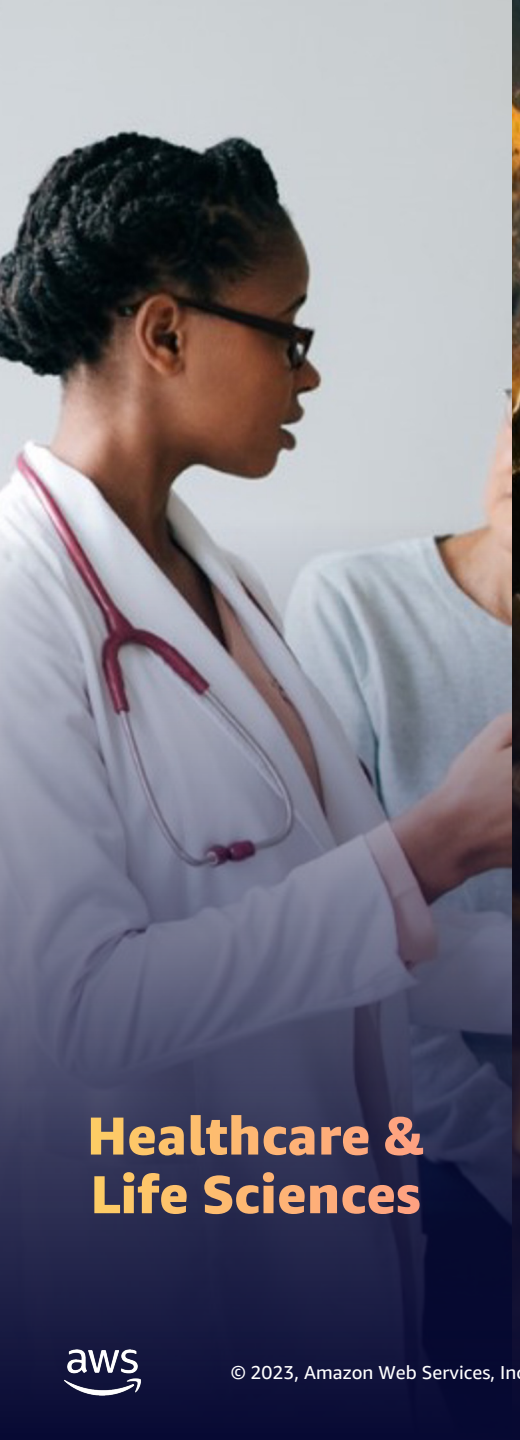
Optimize business processes

DOCUMENT PROCESSING

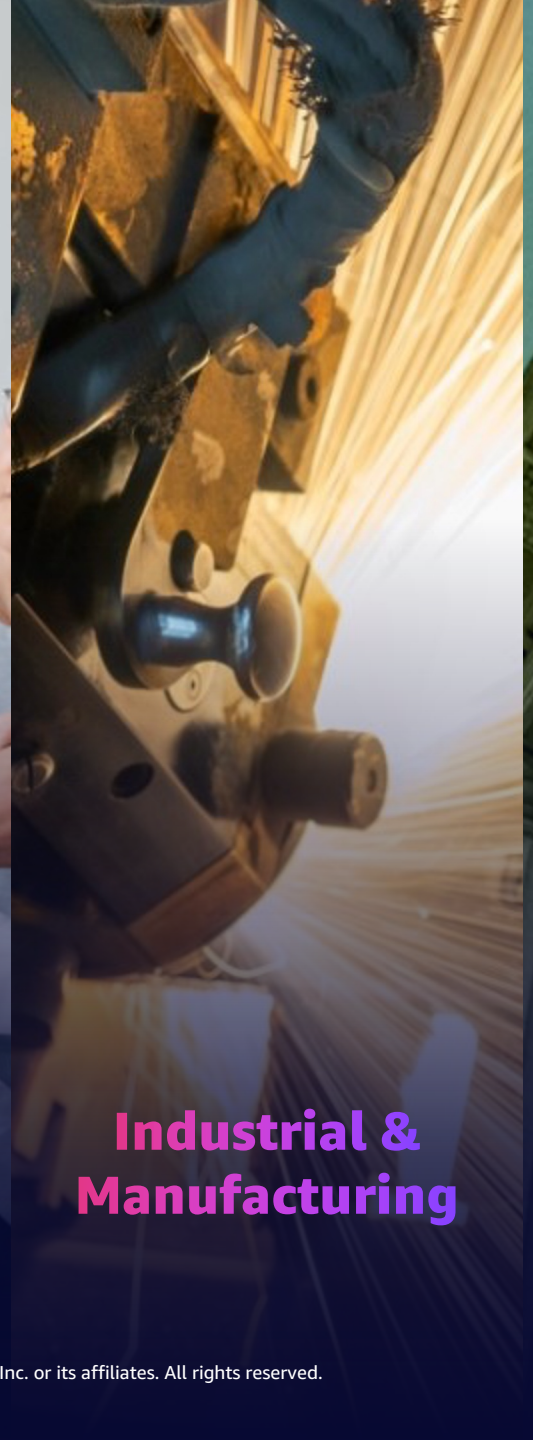
DATA AUGMENTATION

CYBERSECURITY

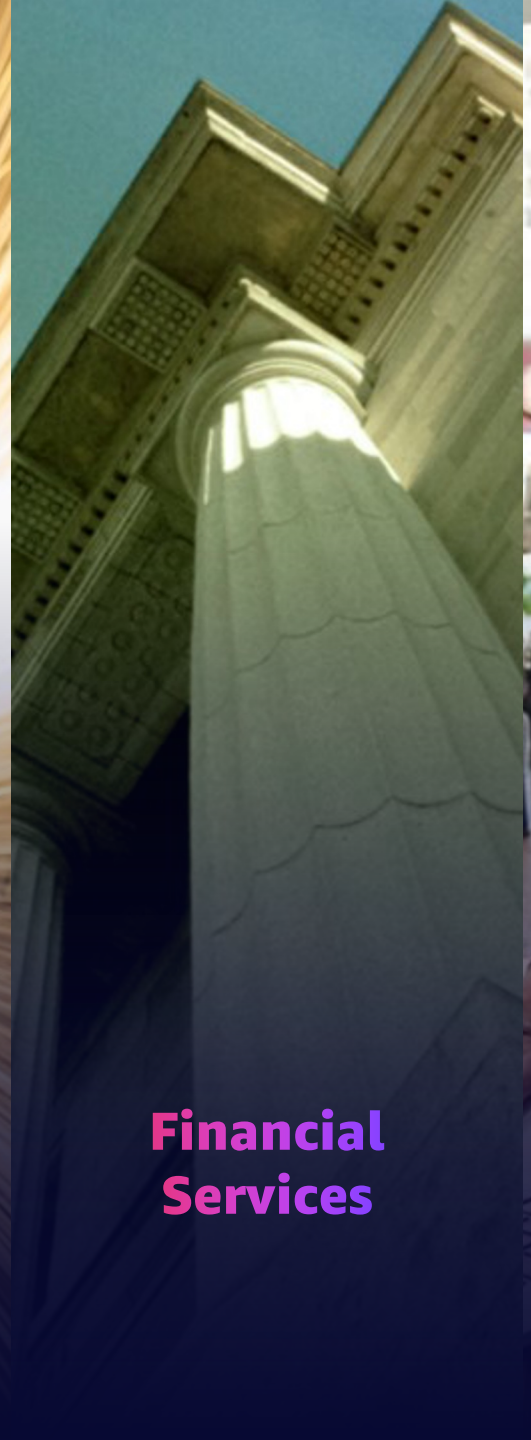
PROCESS OPTIMIZATION



Healthcare & Life Sciences

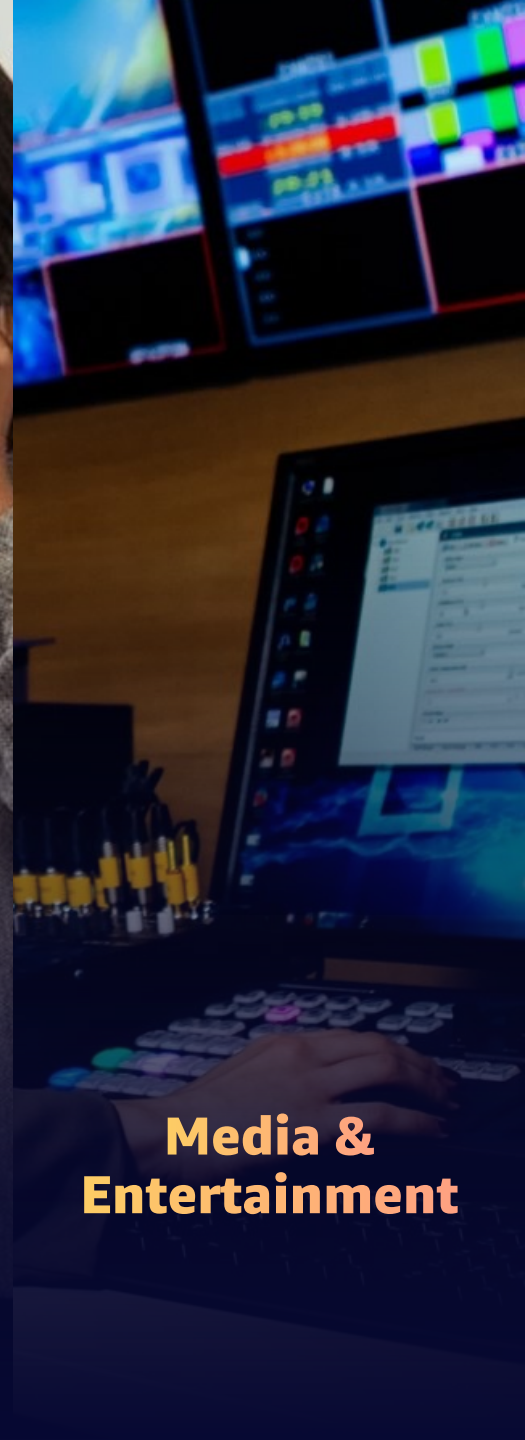


Industrial & Manufacturing



Financial Services

Retail



Media & Entertainment

Healthcare & Life Sciences

- Ambient digital scribe
- Medical imaging
- Drug discovery
- Enhance clinical trials
- Research reporting

Industrial & Manufacturing

- Product design
- Operational efficiency
- Maintenance Assistants
- Supply chain optimization
- Equipment diagnostics

Financial Services

- Portfolio management
- Financial documentation
- Intelligent advisory
- Fraud detection
- Compliance assistant

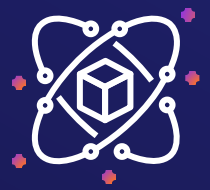
Retail

- Pricing optimization
- Virtual try-ons review
- Marketing Optimization
- Product descriptions
- Pers. Recommendations

Media & Entertainment

- HQ content at scale
- Enrich broadcast content
- Automated content tagging
- Optimize subscriber exper.
- Automated highlights gen.

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

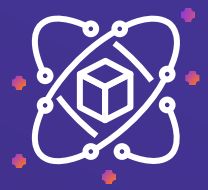
Increase productivity

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

JURASSIC-2

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

ANTHROPIC

CLAUDE 2

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

co:here

COMMAND

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

Meta AI

LLAMA 2

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

stability.ai

SDXL 1.0

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

amazon

AMAZON TITAN

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

aws

Foundation models alone

cannot execute tasks

Agents for **Amazon Bedrock**

Enable generative AI applications to complete tasks in just a few clicks

IN PREVIEW

Breaks down and orchestrates tasks

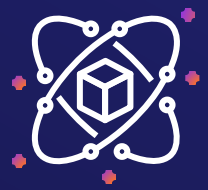
Securely accesses and retrieves company data

Takes action by executing API calls on your behalf

Provides fully managed infrastructure support

Driving innovation with Amazon Bedrock

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

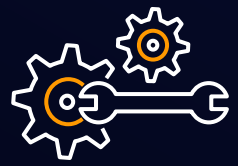
with generative AI applications and services

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

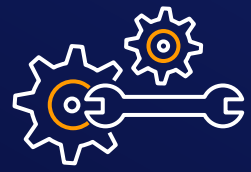
Build a data strategy to fuel your generative AI applications

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

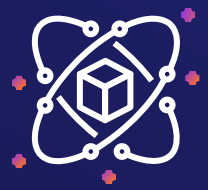
Keeping your data private and secure

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

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

AWS HealthScribe

A HIPAA-eligible automatic note generation service for clinical applications

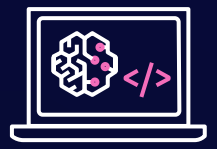
IN PREVIEW TODAY

- Enhances clinical productivity
- Enables AI to be used responsibly in clinical settings
- Includes built-in security, privacy, and compliance features

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
aws

COMING SOON

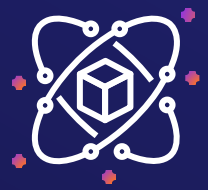
Amazon **CodeWhisperer customization capability**

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

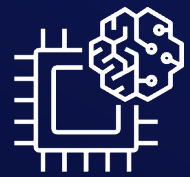
Most performant, low cost

infrastructure to scale generative AI

Deep investments
in global infrastructure

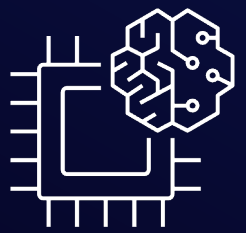
Broad choice of ML accelerators

High performance, low-cost ML infrastructure



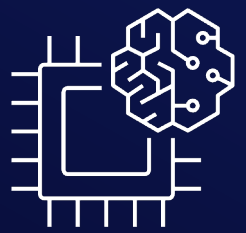
10+ years of silicon innovation

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

Demo