



# Generative AI for Software Companies

How to start building and scaling generative AI on AWS



# Table of contents

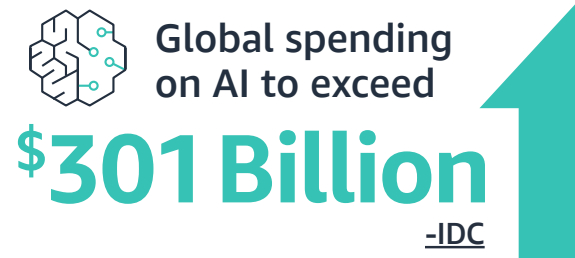
<b>Introduction: Helping Software Companies Harness Generative AI</b>	3
<b>Understanding Generative AI</b>	4
<b>The Benefits of Generative AI for Chief Product Officers</b>	5
<b>Business Application Use Cases</b>	6
<b>How AWS Helps Software Companies Easily Build and Scale Generative AI</b>	7
<b>Quickly Apply Generative AI to Create Practical Solutions</b>	8
<b>How to Get Started</b>	10

# Introduction: Helping Software Companies Harness Generative AI

A primary objective of software Chief Product Officers (CPOs) across the globe is to drive product innovation and velocity as part of efficiency and growth initiatives. Generative artificial intelligence (AI) can be a game-changer, accelerating innovation and value capture from AI and machine learning (ML) in support of your goals.

According to Accenture, 84% of business executives believe they need to use AI to achieve their growth objectives. Unfortunately, 76% acknowledge struggling with how to scale AI across their business. It can be difficult to determine how best to leverage generative AI for your organization. Which opportunities should you pursue? What types of resources do you need to apply to get the maximum benefits? How can you get started?

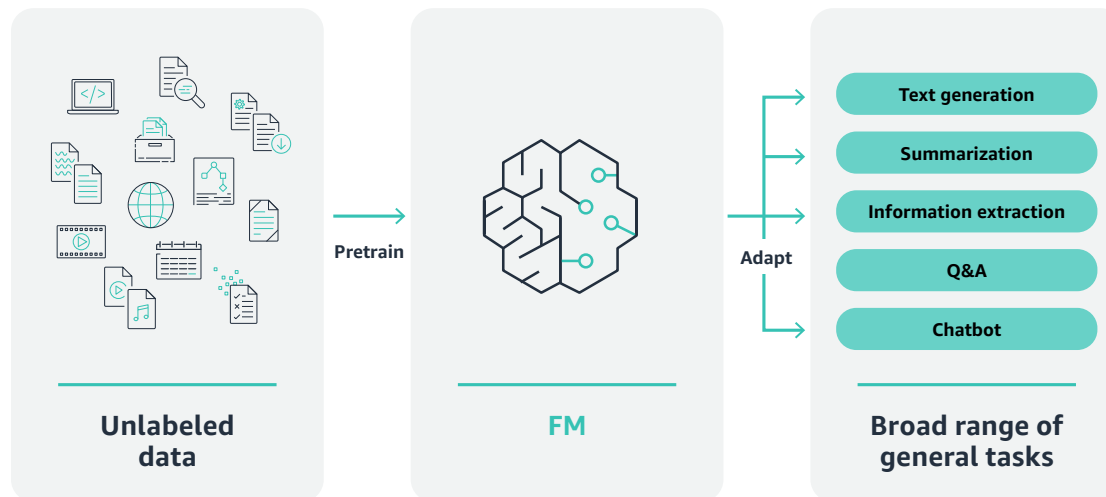
This ebook explores all these questions, providing basic information and guidelines that can help you identify how to start to use generative AI to improve your products, services, and general enterprise operations.



# Understanding Generative AI

Generative AI is the latest evolution in AI capable of creating new content and ideas, including conversations, stories, images, videos, and music. Like all AI, generative AI is powered by machine learning (ML) models. These ML models, commonly referred to as foundation models (FMs), are pre-trained on vast amounts of data containing a large number of parameters that allow them to learn complex concepts and apply their knowledge within a wide range of contexts.

## How foundation models work



Your software company can customize these models to perform domain-specific functions that will differentiate you in the market and add value to your business. This is an extremely efficient approach, taking only a small fraction of the data and compute that would be required to train a model from scratch. Because there are several different large language models available for implementing generative AI, each with its own unique strengths and characteristics, you need to make sure you are free to choose the model(s) that work best for your business.



Generative AI will change the nature of how we interact with all software, and given how many brands have significant software components in how they interact with customers, generative AI will drive and distinguish how more brands compete.”

“Generative AI Will Change Your Business. Here’s How to Adapt,” by David C. Edelman and Mark Abraham, Harvard Business Review, April 12, 2023.

# The Benefits of Generative AI for Chief Product Officers

Generative AI is expected to spark creativity, productivity, and economic growth. According to [Goldman Sachs](#), generative AI could drive a 7% (or almost \$7 trillion) increase in global GDP and lift productivity growth by 1.5 percentage points over a 10-year period. Generative AI can be used across all lines of business, including engineering, marketing, customer service, finance, and sales, to improve the efficiency and productivity of your internal teams and ultimately the experiences and outcomes of your end-customers. Many of your customers are starting to use generative AI for many different use cases – customers across industries, from healthcare, such as 3M and Philips, to financial services, such as BBVA, are already working on new applications. As a Chief Product Officer (CPO), you can use generative AI to enable your organizations to take advantage of ML quicker and apply it to a broad set of use cases.

## Engineering and Product (Developer) Use Cases

- Write code or provide code suggestions to increase productivity and develop applications faster and deliver services to market to drive revenue.
- Identify coding mistakes, automate refactoring, or explore alternative solutions to improve the consistency and quality of code.
- Build unit tests to help detect and resolve coding errors under a wide variety of operating parameters, including corner-cases that are typically difficult to discover.
- Explain poorly documented code or generate new documentation for code.

## Data Use Cases

- Augment data through synthetic data generation to enhance the quality of results.
- Improve the customer experience with chatbots and virtual assistants that can better anticipate and address customer needs.
- Drive personalization and content modernization to better respond to the needs of stakeholders.
- Boost employee productivity with conversational search, content creation, text summarization, and code creation, among others.
- Increase the efficiencies of business operations with intelligent document processing, predictive maintenance, quality control, and visual inspection.

# Business Application Use Cases

- Transform call centers using generative AI-powered chatbots/virtual assistants to answer customer questions with conversational interfaces and handle interactions in a fully-automated way.
- Provide live turn-by-turn sentiment analysis of calls/chats to enable effective escalations and interventions before the customer ends the interaction to prevent churn.
- Summarize calls and chat transcripts using large language models (LLM) to identify trends and opportunities to improve products and services.
- Detect real-time intent of customer enquiries to automatically generate next best recommended actions for agents.
- Automatically respond to customer enquiries, using a custom knowledge base, to improve agent productivity and accelerate issue resolution.
- Create ad copy and stunning visuals in minutes using natural language instructions.
- Produce financial reports to summarize financial results tailored to different business units.



For many developers, generative AI will become the most valuable coding partner they will ever know.”

**KPMG**



# How AWS Helps Software Companies Easily Build and Scale Generative AI

We've been working on AI and ML for over 20 years – it's part of our heritage, ethos, and future. We partner with our customers on their generative AI strategies to impart what we've learned from our own experiences – more than 100,000 customers use AWS for ML.

**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** 

We are democratizing ML and making it accessible to anyone who wants to use it as AWS services. At AWS, our commitment to develop AI and machine learning in a responsible way is integral to our approach to building and delivering these services. We will continue to iterate our understanding and implementation as AI evolves and matures, looking at bias and fairness, explainability, robustness, privacy and security, governance, and transparency to ensure we use it responsibly.

We offer the broadest and deepest portfolio of AI and ML services at all three layers of the stack. For starters, we are making AWS the easiest place for you to build and scale generative AI-based applications with foundation models (FMs). **Amazon Bedrock** provides model choice,

making FMs from Amazon and leading AI startups, including AI21 Labs, Anthropic, Stability AI, and Cohere accessible via an API. Capabilities and services, like **Amazon SageMaker** and Amazon S3, then make it easy for developers to use familiar controls to readily integrate and deploy FMs into applications and workloads on AWS.

Since the data you use for customization is incredibly valuable IP, you can configure your Amazon Virtual Private Cloud (Amazon VPC) settings to access Amazon Bedrock APIs and provide model fine-tuning data in a secure manner. All data is encrypted in transit (TLS1.2) and at rest through service managed keys to ensure it stays protected, secure, and private. AWS offers the most secure global cloud infrastructure on which to build, migrate, and manage applications and workloads. This is backed by a deep set of 300+ cloud security tools and the trust of millions of customers, including some of the most security-sensitive organizations, like government, healthcare, and financial services.

In addition, we have been investing in the most price-performant infrastructure for ML, investing in own silicon to push the envelope on performance and price for demanding workloads. AWS Trainium and Inferentia chips offer the lowest cost for training models and running inference in the cloud – Trn1 instances, powered by Trainium, can deliver up to 50% savings on training costs over any other EC2 instance.

When it comes to the actual coding, our AI coding companion, **HYPERLINK Amazon CodeWhisperer**, can do a lot of the heavy lifting and “write” a lot of the undifferentiated code. This allows developers to build faster, freeing them to focus on the more creative aspects of coding.



# Quickly Apply Generative AI to Create Practical Solutions

We can help you apply generative AI to your real-world applications at scale to create practical solutions to a variety of your business problems. We enable you to choose from cutting-edge foundation models, such as **Amazon Titan FMs**, as well as the models from leading AI startups, including AI21 Labs, Anthropic, Stability AI, and Cohere, to best meet the needs of your business. We also allow you to customize these models with your data to get much better performance than out of the box and build differentiated generative AI applications in an easy and secure way with the breadth and depth of AWS security controls and services.

**You can quickly integrate and deploy FMs into your applications and workloads running on AWS** using familiar controls and integrations with the depth and breadth of AWS capabilities and services like Amazon SageMaker and Amazon S3.

Of course, with AWS, you have a cost-effective infrastructure powered by AWS-built silicon chips tailored for generative AI training and inference to make it easy for you to run large models in production. We also provide Amazon **CodeWhisperer** to help you use generative AI to code and boost developer productivity.

## PRODUCTIVITY CHALLENGE

When Amazon ran a productivity challenge, participants who used Amazon CodeWhisperer were

**27%**  
more likely  
to complete tasks



AND  
**57%**  
faster







We are building our generative AI capabilities on AWS to benefit from scale, reliability, and rate of innovation. From innovation across data collection to AI model creation, AWS helped us pioneer cloud-powered cybersecurity, first disrupting endpoint security and now doing the same for cloud security. Today, with cybersecurity's richest data set, we're excited to further our AI collaboration with AWS."

**Raj Rajamani**  
Chief Product Officer DICE  
Data Identity Cloud Endpoint  
CrowdStrike



As a pioneer in Digital Engineering, we have always believed in staying ahead of the disruptive technology curve, such as Generative AI. Our proactive approach and deep collaboration with AWS have enabled us to provide game-changing engineering solutions that have accelerated time to value for our clients across industries.

With the adoption of CodeWhisperer by our engineering teams, we are witnessing valuable productivity gains across the software engineering lifecycle and expect these to improve significantly over time. With this alliance, we will work closely with AWS in onboarding our Enterprise and ISV clients to CodeWhisperer and provide them the necessary training and support to accelerate their digital transformation journeys."

**Sandeep Kalra**  
CEO and Executive Director  
Persistent Systems



# How to Get Started

To start taking advantage of the benefits of generative AI on AWS, you can:

- 1** Deploy Amazon CodeWhisperer to achieve immediate productivity gains.
- 2** Get started as soon as possible with FMs through the Amazon Bedrock preview.
- 3** Work towards a PoC with the AWS Generative AI Innovation Center.

We offer the following resources to learn more:

- Get [An introduction to generative AI](#): Read what Werner Vogels, Amazon CTO, has to say about Generative AI
- [Explore generative AI on AWS](#)
- [Demystifying generative AI](#): Hear what Werner Vogels sees as the impact of generative AI for businesses
- [Learn about Amazon Bedrock](#), the easiest way to build and scale generative AI applications with FMs
- [Learn about Amazon Titan](#), high-performing FMs from Amazon to help you innovate responsibly
- Discover the new [AWS Trainium based Trn1n instance](#) and the [AWS Inferentia based Inf2 instance](#)
- Learn how you can use [Amazon CodeWhisperer as your coding companion](#)
- Find out about [New Tools for Building with Generative AI on AWS](#)
- See how [C3 AI and AWS Are Delivering Enhanced Generative AI for the Enterprise](#)
- [7 generative AI innovations from AWS](#)

