AWS Decision Guide

Choosing an AWS machine learning service

Choosing an AWS machine learning service: AWS Decision Guide

Table of Contents

Choosing an AWS machine learning service

Pick the right ML services and frameworks to support your work

Purpose	Help determine which AWS ML services are the best fit for your needs.
Last updated	May 3, 2024
Covered services	• Amazon Augmented AI • Amazon CodeGuru • Amazon Comprehend • Amazon DevOps Guru • Amazon Forecast • Amazon Kendra • Amazon Lex • Amazon Personalize • Amazon Polly • Amazon Rekognition • Amazon SageMaker AI • Amazon Textract • Amazon Transcribe • Amazon Translate

Introduction

At its most basic, machine learning (ML) is designed to provide digital tools and services to learn from data, identify patterns, make predictions, and then act on those predictions. Almost all artificial intelligence (AI) systems today are created using ML. ML uses large amounts of data to create and validate decision logic. This decision logic forms the basis of the AI *model*.

Scenarios where AWS machine learning services may be applied include:

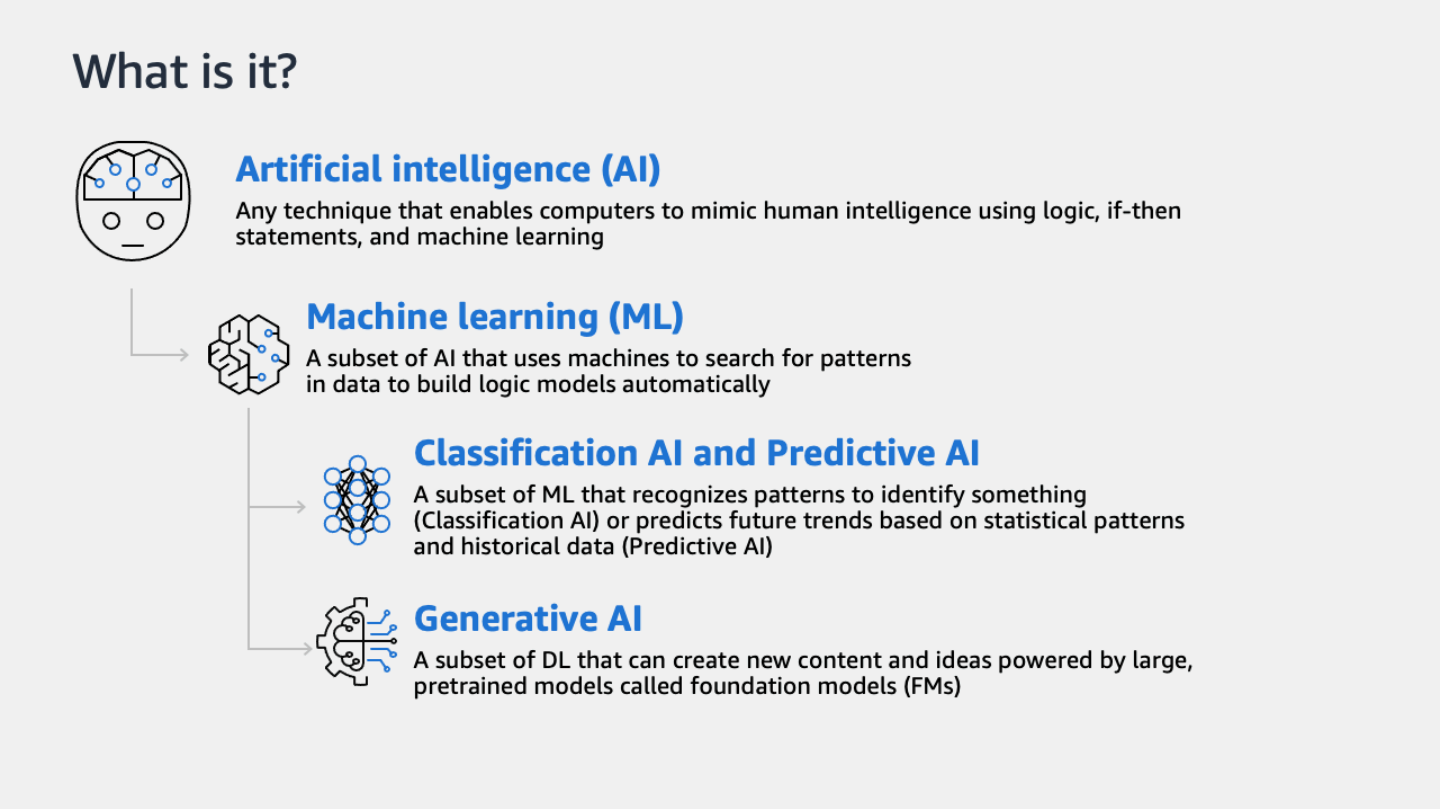
- **Specific use cases** — AWS machine learning services can support your AI powered use cases with a broad range of pre-built algorithms, models, and solutions for common use cases and industries. You have a choice of 23 pre-trained services, including Amazon Personalize, Amazon Kendra, and Amazon Monitron.
- **Customizing and scaling machine learning** — Amazon SageMaker AI is designed to help you build, train, and deploy ML models for any use case. You can build your own or access open source foundational models on AWS through Amazon SageMaker AI and Amazon Bedrock.
- **Accessing specialized infrastructure** — Use the ML frameworks and infrastructure provided by AWS when you require even greater flexibility and control over your machine learning workflows, and are willing to manage the underlying infrastructure and resources yourself.

This decision guide will help you ask the right questions, evaluate your criteria and business problem, and determine which services are the best fit for your needs.

Understand

As organizations continue to adopt AI and ML technologies, the importance of understanding and choosing among AWS ML services is an on-going challenge.

AWS provides a range of ML services designed to help organizations to build, train, and deploy ML models more quickly and easily. These services can be used to solve a wide range of business problems such as customer churn prediction, fraud detection, and image and speech recognition.



Before diving deeper into AWS ML services, let's look at the relationship between AI and ML.

- At a high level, *artificial intelligence* is a way to describe any system that can replicate tasks that previously required human intelligence. Most AI use cases are looking for a probabilistic outcome—making a prediction or decision with a high degree of certainty, similar to human judgement.
- Almost all AI systems today are created using *machine learning*. ML uses large amounts of data to create and validate decision logic, which is known as a model.
- Classification AI is a subset of ML that recognizes patterns to identify something. Predictive AI is a subset of ML that predicts future trends based on statistical patterns an historical data.
- Finally, generative AI is a subset of deep learning that can create new content and ideas, like conversations, stories, images, videos, and music. Generative AI is powered by very large models that are pretrained on vast corpora of data, called the Foundation Models or FMs. Amazon Bedrock is a fully managed service that offers a choice of high-performing FMs for building and scaling generative AI applications. Amazon Q Developer and Amazon Q Business are generative-AI powered assistants for specific use cases.

This guide is designed primarily to cover services in the *Classification AI* and *Predictive AI* machine learning categories.

In addition, AWS offers specialized, accelerated hardware for high performance ML training and inference.

- Amazon EC2 P5 instances are equipped with NVIDIA H100 Tensor Core GPUs, which are well-suited for both training and inference tasks in machine learning. Amazon EC2 G5 instances feature up to 8 NVIDIA A10G Tensor Core GPUs, and second generation AMD EPYC processors, for a wide range of graphics-intensive and machine learning use cases.
- AWS Trainium is the second-generation ML accelerator that AWS has purpose-built for deep learning (DL) training of 100B+ parameter models.
- AWS Inferentia2-based Amazon EC2 Inf2 instances are designed to deliver high performance at the lowest cost in Amazon EC2 for your DL and generative AI inference applications.

Consider

When solving a business problem with AWS ML services, consideration of several key criteria can help ensure success. The following section outlines some of the key criteria to consider when choosing a ML service.

Problem definition

Problem definition

The first step in the ML lifecycle is to frame the business problem. Understanding the problem you are trying to solve is essential for choosing the right AWS ML service, as different services are designed to address different problems. It is also important to determine whether ML is the best fit for your business problem.

Once you have determined that ML is the best fit, you can start by choosing from a range of purpose-built AWS AI services (in areas such as speech, vision and documents).

Amazon SageMaker AI provides fully managed infrastructure if you need to build and train your own models. AWS offers an array of advanced ML frameworks and infrastructure choices for the cases where you require highly customized and specialized ML models. AWS also offers a broad set of popular foundation models for building new applications with generative AI.

ML algorithm

ML algorithm

Choosing the ML algorithm for the business problem you are trying to solve depends on the type of data you are working with, as well as the desired outcomes. The following information outlines how each of the major AWS AI/ML service categories empowers you to work with its algorithms:

- Specialized AI services: These services offer a limited ability to customize the ML algorithm, as they are pre-trained models optimized for specific tasks. You can typically customize the input data and some parameters, but do not have access to the underlying ML models or the ability to build your own models.
- Amazon SageMaker AI: This service provides the most flexibility and control over the ML algorithm. You can use SageMaker AI to build custom models using your own algorithms and frameworks, or use pre-built models and algorithms provided by AWS. This allows for a high degree of customization and control over the ML process.
- Lower-level ML frameworks and infrastructure: These services offer the most flexibility and control over the ML algorithm. You can use these services to build highly customized ML models using their own algorithms and frameworks. However, using these services requires significant ML expertise and may not be feasible for all every use case.

Security

Security

If you need a private endpoint in your VPC, your options will vary based on the layer of AWS ML services you are using. These include:

- Specialized AI services: Most specialized AI services do not currently support private endpoints in VPCs. However, Amazon Rekognition Custom Labels and Amazon Comprehend Custom can be accessed using VPC endpoints.
- Core AI services: Amazon Translate, Amazon Transcribe, and Amazon Comprehend all support VPC endpoints.
- Amazon SageMaker AI: SageMaker AI provides built-in support for VPC endpoints, allowing you to deploy their trained models as an endpoint accessible only from within their VPC.
- Lower-level ML frameworks and infrastructure: You can deploy your models on Amazon EC2 instances or in containers within your VPC, providing complete control over the networking configuration.

Latency

Latency

Higher-level AI services, such as Amazon Rekognition and Amazon Transcribe, are designed to handle a wide variety of use cases and offer high performance in terms of speed. However, they might not meet certain latency requirements.

If you are using lower-level ML frameworks and infrastructure, we recommended leveraging Amazon SageMaker AI. This option is generally faster than building custom models due to its fully managed service and optimized deployment options. While a highly optimized custom model may outperform SageMaker AI, it will require significant expertise and resources to build.

Accuracy

Accuracy

The accuracy of AWS ML services varies based on the specific use case and level of customization required. Higher-level AI services, such as Amazon Rekognition, are built on pre-trained models that have been optimized for specific tasks and offer high accuracy in many use cases.

In some cases, you can choose to use Amazon SageMaker AI, which provides a more flexible and customizable platform for building and training custom ML models. By building your own models, you may be able to achieve even higher accuracy than what is possible with pre-trained models.

You can also choose to use ML frameworks and infrastructure, such as TensorFlow and Apache MXNet, to build highly customized models that offer the highest possible accuracy for your specific use case.

AWS and responsible AI

AWS and responsible AI

AWS builds foundation models (FMs) with responsible AI in mind at each stage of its development process. Throughout design, development, deployment, and operations we consider a range of factors including:

1. Accuracy (how closely a summary matches the underlying document; whether a biography is factually correct)
2. Fairness, (whether outputs treat demographic groups similarly)

3. Intellectual property and copyright considerations
4. Appropriate usage (filtering out user requests for legal advice, or medical diagnoses, or illegal activities)
5. Toxicity (hate speech, profanity, and insults)
6. Privacy (protecting personal information and customer prompts)

AWS builds solutions to address these issues into the processes used for acquiring training data, into the FMs themselves, and into the technology used to pre-process user prompts and post-process outputs.

Choose

Now that you know the criteria by which you will be evaluating your ML service options, you are ready to choose which AWS ML service is right for your organizational needs. The following table highlights which ML services are optimized for which circumstances. Use it to help determine the AWS ML service that is the best fit for your use case.

Categories	When would you use it?	What is it optimized for?	Related AI/ ML services or environments
Specific use cases These artificial intelligence services are intended to meet specific needs. They include personalization, forecasting, anomaly detection, speech transcription, and others. Since they are delivered as services, they can be embedded into applications without	Use the AI services provided by AWS when you require specific, pre-built functionalities to be integrated into your applications, without the need for extensive customizations or machine learning expertise. These services are designed to be easy to use and do not	These services are designed to be easy to use and do not require much coding, configuration, or ML expertise.	Amazon Augmented AI Amazon CodeGuru Amazon Comprehend Amazon Comprehend Medical Amazon DevOps Guru Amazon Forecast Amazon Kendra

Categories	When would you use it?	What is it optimized for?	Related AI/ ML services or environments
requiring any ML expertise.	require much coding or configuration.		Amazon Lex Amazon Personalize Amazon Polly Amazon Rekognition Amazon Textract Amazon Transcribe Amazon Translate
ML services These services can be used to develop customized machine learning models or workflows that go beyond the pre-built functionalities offered by the core AI services.	Use these services when when you need more customized machine learning models or workflows that go beyond the pre-built functiona lities offered by the core AI services.	These services are optimized for building and training custom machine learning models, large-scale training on multiple instances or GPU clusters, more control over machine learning model deployment, real-time inference, and for building end-to-end workflows.	Amazon SageMaker AI Amazon SageMaker AI JumpStart SageMaker AI Studio SageMaker AI Canvas SageMaker AI Studio Lab SageMaker AI Ground Truth PyTorch on AWS Apache MxNet Hugging Face TensorFlow on AWS

Categories	When would you use it?	What is it optimized for?	Related AI/ ML services or environments
Infrastructure To deploy machine learning in production, you need cost-effective infrastructure, which Amazon enables with AWS-built silicon.	Use when you want to achieve the lowest cost for training models and need to run inference in the cloud.	Optimized for supporting the cost-effective deployment of machine learning.	AWS Trainium AWS Inferentia and Inferentia2 Amazon SageMaker AI HyperPod
Tools and associated services These tools and associated services are designed to help you ease deployment of machine learning.	These services and tools are designed to help you accelerate deep learning in the cloud, providing Amazon machine images, docker images and entity resolution.	Optimized for helping you accelerate deep learning in the cloud.	AWS Deep Learning AMIss AWS Deep Learning Containers AWS Entity Resolution

Use

Now that you have a clear understanding of the criteria you need to apply in choosing an AWS ML service, you can select which AWS AI/ML service(s) are optimized for your business needs.

To explore how to use and learn more about the service(s) you have chosen, we have provided three sets of pathways to explore how each service works. The first set of pathways provides in-depth documentation, hands-on tutorials, and resources to get started with Amazon Comprehend, Amazon Textract, Amazon Translate, Amazon Lex, Amazon Polly, Amazon Rekognition, and Amazon Transcribe.

Amazon Comprehend

- **Get started with Amazon Comprehend**

  Use the Amazon Comprehend console to create and run an asynchronous entity detection job.

  Get started with the tutorial »

- **Analyze insights in text with Amazon Comprehend**

  Learn how to use Amazon Comprehend to analyze and derive insights from text.

  Get started with the tutorial »

- **Amazon Comprehend Pricing**

  Explore information on Amazon Comprehend pricing and examples.

  Explore the guide »

Amazon Textract

- **Getting Started with Amazon Textract**

  Learn how Amazon Textract can be used with formatted text to detect words and lines of words that are located close to each other, as well as analyze a document for items such as related text, tables, key-value pairs, and selection elements.

  Explore the guide »

- **Extract text and structured data with Amazon Textract**

  Learn how to use Amazon Textract to extract text and structured data from a document.

  Get started with the tutorial »

- **AWS Power Hour: Machine Learning**

Dive into Amazon Textract in this episode, spend time in the AWS Management Console, and review code samples that will help you understand how to make the most of service APIs.

Watch the video »

Amazon Translate

- **Getting started with Amazon Translate using the console**

  The easiest way to get started with Amazon Translate is to use the console to translate some text. Learn how to translate up to 10,000 characters using the console.

  Explore the guide »

- **Translate Text Between Languages in the Cloud**

  In this tutorial example, as part of an international luggage manufacturing firm, you need to understand what customers are saying about your product in reviews in the local market language - French.

  Get started with the tutorial »

- **Amazon Translate pricing**

  Explore Amazon Translate pricing, including Free Tier - which provides 2 million characters per month for 12 months.

  Explore the guide »

Amazon Lex

- **Amazon Lex V2 Developer Guide**

  Explore information about getting started, how it works, and pricing information for Amazon Lex V2.

  Explore the guide »

- **Introduction to Amazon Lex** We introduce you to the Amazon Lex conversational service, and walk you through examples that show you how to create a bot and deploy it to different chat services.

  Take the course » (sign-in required)

- **Exploring Generative AI in conversational experiences**

  Explore the use of generative AI in conversation experiences.

  Read the blog »

Amazon Polly

- **What is Amazon Polly?**

  Explore a complete overview of the cloud service that converts text into lifelike speech, and can be used to develop applications to increase your customer engagement and accessibility.

  Explore the guide »

- **Highlight text as it's being spoken using Amazon Polly**

  We introduce you to approaches for highlighting text as it's being spoken to add visual capabilities to audio in books, websites, blogs, and other digital experiences.

  Read the blog »

- **Create audio for content in multiple languages with the same TTS voice persona in Amazon Polly**

  We explain Neural Text-to-Speech (NTTS) and discuss how a broad portfolio of available voices, providing a range of distinct speakers in supported languages, can work for you.

  Read the blog »

Amazon Rekognition

- **What is Amazon Rekognition?**

Explore how you can use this service to add image and video analysis to your applications.

Explore the guide »

- **Hands-on Rekognition: Automated Image and Video Analysis**

  Learn how facial recognition works with streaming video, along with code examples and key points at a self-guided pace.

  Get started with the tutorial »

- **Amazon Rekognition FAQs**

  Learn the basics of Amazon Rekognition and how it can help you improve your deep learning and visually analyze your applications.

  Read the FAQs »

Amazon Transcribe

- **What is Amazon Transcribe?**

  Explore the AWS automatic speech recognition service using ML to convert audio to text. Learn how to use this service as a standalone transcription or add speech-to-text capability to any application.

  Explore the guide »

- **Amazon Transcribe Pricing**

  We introduce you to the AWS pay-as-you-go transcription, including custom language model options and the Amazon Transcribe Free Tier.

  Explore the guide »

- **Create an audio transcript with Amazon Transcribe**

  Learn how to use Amazon Transcribe to create a text transcript of recorded audio files using a real-world use case scenario for testing against your needs.

Get started with the tutorial »

- **Build an Amazon Transcribe streaming app**

  Learn how to build an app to record, transcribe, and translate live audio in real-time, with results emailed directly to you.

  Explore the guide »

The second set of AI/ML AWS service pathways provide in-depth documentation, hands-on tutorials, and resources to get started with the services in the Amazon SageMaker AI family.

SageMaker AI

- **How Amazon SageMaker AI works**

  Explore the overview of machine learning and how SageMaker AI works.

  Explore the guide »

- **Getting started with Amazon SageMaker AI**

  Learn how to join an Amazon SageMaker AI Domain, giving you access to Amazon SageMaker AI Studio and RStudio on SageMaker AI.

  Explore the guide »

- **Use Apache Spark with Amazon SageMaker AI**

  Learn how to use Apache Spark for preprocessing data and SageMaker AI for model training and hosting.

  Explore the guide »

- **Use Docker containers to build models**

  Explore how Amazon SageMaker AI makes extensive use of Docker containers for build and runtime tasks. Learn how to deploy the pre-built Docker images for its built-in algorithms and the supported deep learning frameworks used for training and inference.

Explore the guide »

- **Machine learning frameworks and languages**

  Learn how to get started with SageMaker AI using the Amazon SageMaker AI Python SDK.

  Explore the guide »

SageMaker AI Autopilot

- **Create an Amazon SageMaker AI Autopilot experiment for tabular data**

  Learn you how to create an Amazon SageMaker AI Autopilot experiment to explore, pre-process, and train various model candidates on a tabular dataset.

  Explore the guide »

- **Automatically create machine learning models**

  Learn how to use Amazon SageMaker AI Autopilot to automatically build, train, and tune a ML model, and deploy the model to make predictions.

  Get started with the tutorial »

- **Explore modeling with Amazon SageMaker AI Autopilot with these example notebooks**

  Explore example notebooks for direct marketing, customer churn prediction and how to bring your own data processing code to Amazon SageMaker AI Autopilot.

  Explore the guide »

SageMaker AI Canvas

- **Get started using Amazon SageMaker AI Canvas**

  Learn how to get started with using SageMaker AI Canvas.

  Explore the guide »

- **Generate machine learning predictions without writing code**

  This tutorial explains how to use Amazon SageMaker AI Canvas to build ML models and generate accurate predictions without writing a single line of code.

  Get started with the tutorial »

- **Dive deeper into SageMaker AI Canvas**

  Explore an in-depth look at SageMaker AI Canvas and its visual, no code ML capabilities.

  Read the blog »

- **Use Amazon SageMaker AI Canvas to make your first ML Model**

  Learn how to use Amazon SageMaker AI Canvas to create an ML model to assess customer retention, based on an email campaign for new products and services.

  Get started with the lab »

SageMaker AI Data Wrangler

- **Getting started with Amazon SageMaker AI Data Wrangler**

  Explore how set up SageMaker AI Data Wrangler and then provides a walkthrough using an existing example dataset.

  Explore the guide »

- **Prepare training data for machine learning with minimal code**

  Learn how to prepare data for ML using Amazon SageMaker AI Data Wrangler.

  Get started with the tutorial »

- **SageMaker AI Data Wrangler deep dive workshop**

  Learn how to apply appropriate analysis types on your dataset to detect anomalies and issues, use the derived results/insights to formulate remedial actions in the course of

transformations on your dataset, and test the right choice and sequence of transformations using quick modeling options provided by SageMaker AI Data Wrangler.

Get started with the workshop »

SageMaker AI Ground Truth

- **Getting Started with Amazon Ground Truth**

  Explore how to use the console to create a labeling job, assign a public or private workforce, and send the labeling job to your workforce. Learn how to monitor the progress of a labeling job.

  Explore the guide »

- **Label Training Data for Machine Learning**

  Learn how to set up a labeling job in Amazon SageMaker AI Ground Truth to annotate training data for your ML model.

  Get started with the tutorial »

- **Getting started with Amazon Ground Truth Plus** Explore how to complete the necessary steps to start an Amazon SageMaker AI Ground Truth Plus project, review labels, and satisfy SageMaker AI Ground Truth Plus prerequisites.

  Explore the guide »

- **Get started with Amazon Ground Truth** Watch how to get started with labeling your data in minutes through the SageMaker AI Ground Truth console.

  Watch the video »

- **Amazon SageMaker AI Ground Truth Plus – create training datasets without code or in-house resources**

  Learn about Ground Truth Plus, a turn-key service that uses an expert workforce to deliver high-quality training datasets fast, and reduces costs by up to 40 percent.

Read the blog »

SageMaker AI JumpStart

- **Get started with machine learning with SageMaker AI JumpStart**

  Explore SageMaker AI JumpStart solution templates that set up infrastructure for common use cases, and executable example notebooks for machine learning with SageMaker AI.

  Explore the guide »

- **Get Started with your machine learning project quickly using Amazon SageMaker AI JumpStart**

  Learn how to fast-track your ML project using pretrained models and prebuilt solutions offered by Amazon SageMaker AI JumpStart. You can then deploy the selected model through Amazon SageMaker AI Studio notebooks.

  Get started with the tutorial »

- **Get hands-on with Amazon SageMaker AI JumpStart with this Immersion Day workshop**

  Learn how the low-code ML capabilities found in Amazon SageMaker AI Data Wrangler, Autopilot and Jumpstart, make it easier to experiment faster and bring highly accurate models to production.

  Get started with the workshop »

SageMaker AI Pipelines

- **Getting Started with Amazon SageMaker AI Pipelines**

  Learn how to create end-to-end workflows that manage and deploy SageMaker AI jobs. SageMaker AI Pipelines comes with SageMaker AI Python SDK integration, so you can build each step of your pipeline using a Python-based interface.

  Explore the guide »

- **Automate machine learning workflows**

  Learn how to create and automate end-to-end machine learning (ML) workflows using Amazon SageMaker AI Pipelines, Amazon SageMaker AI Model Registry, and Amazon SageMaker AI Clarify.

  Get started with the tutorial »

- **How to create fully automated ML workflows with Amazon SageMaker AI Pipelines**

  Learn about Amazon SageMaker AI Pipelines, the world's first ML CI/CD service designed to be accessible for every developer and data scientist. SageMaker AI Pipelines brings CI/CD pipelines to ML, reducing the coding time required.

  Watch the video »

SageMaker AI Studio

- **Build and train a machine learning model locally**

  Learn how to build and train a ML model locally within your Amazon SageMaker AI Studio notebook.

  Get started with the tutorial »

- **SageMaker AI Studio integration with EMR workshop**

  Learn how to utilize distributed processing at scale to prepare data and subsequently train machine learning models.

  Get started with the workshop »

The third set of AI/ML AWS service pathways provide in-depth documentation, hands-on tutorials, and resources to get started with AWS Trainium, AWS Inferentia, and Amazon Titan.

AWS Trainium

- **Scaling distributed training with AWS Trainium and Amazon EKS**

Learn how you can benefit from the general availability of Amazon EC2 Trn1 instances powered by AWS Trainium—a purpose-built ML accelerator optimized to provide a high-performance, cost-effective, and massively scalable platform for training deep learning models in the cloud.

Read the blog »

- **Overview of AWS Trainium**

  Learn about AWS Trainium, the second-generation machine learning (ML) accelerator that AWS purpose built for deep learning training of 100B+ parameter models. Each Amazon Elastic Compute Cloud (EC2) Trn1 instance deploys up to 16 AWS Trainium accelerators to deliver a high-performance, low-cost solution for deep learning (DL) training in the cloud.

  Explore the guide »

- **Recommended Trainium Instances**

  Explore how AWS Trainium instances are designed to provide high performance and cost efficiency for deep learning model inference workloads.

  Explore the guide »

AWS Inferentia

- **Overview of AWS Inferentia**

  Understand how accelerators are designed by AWS to deliver high performance at the lowest cost for your deep learning (DL) inference applications.

  Explore the guide »

- **AWS Inferentia2 builds on AWS Inferentia1 by delivering 4x higher throughput and 10x lower latency**

  Understand what AWS Inferentia2 is optimized for - and explores how it was designed from the ground up to deliver higher performance while lowering the cost of LLMs and generative AI inference.

Read the blog »

- **Machine learning inference using AWS Inferentia**

  Learn how to create an Amazon EKS cluster with nodes running Amazon EC2 Inf1 instances and (optionally) deploy a sample application. Amazon EC2 Inf1 instances are powered by AWS Inferentia chips, which are custom built by AWS to provide high performance and lowest cost inference in the cloud.

  Explore the guide »

Amazon Titan

- **Overview of Amazon Titan**

  Explore how Amazon Titan FMs are pretrained on large datasets, making them powerful, general-purpose models. Learn how you can use them as is - or privately - to customize them with your own data for a particular task without annotating large volumes of data.

  Explore the guide »

Explore

- **Architecture diagrams**

  These reference architecture diagrams show examples of AWS AI and ML services in use.

  Explore architecture diagrams »
- **Whitepapers**

  Explore whitepapers to help you get started and learn best practices in choosing and using AI/ML services.

  Explore whitepapers »
- **AWS Solutions**

  Explore vetted solutions and architectural guidance for common use cases for AI and ML services.

Explore solutions »

Resources

Foundation models

Supported foundation models include:

- Anthropic Claude
- Cohere Command & Embed
- AI21 Labs Jurassic
- Meta Llama
- Mistral AI
- Stable Diffusion XL
- Amazon Titan

Using Amazon Bedrock, you can experiment with a variety of foundation models and privately customize them with your data.

Use case or industry-specific services

- Amazon Comprehend Medical
- Amazon Fraud Detector
- AWS HealthLake
- Amazon Lookout for Equipment
- Amazon Lookout for Metrics
- Amazon Lookout for Vision
- Amazon Monitron
- AWS HealthOmics
- AWS Panorama

Associated blog posts

- Significant new capabilities make it easier to use Amazon Bedrock to build and scale generative AI applications – and achieve impressive results
- AWS Inferentia and AWS Trainium deliver lowest cost to deploy Llama 3 models in Amazon SageMaker AI JumpStart
- Revolutionize Customer Satisfaction with tailored reward models for your business on Amazon SageMaker AI
- Amazon Personalize launches new recipes supporting larger item catalogs with lower latency

Document history

The following table describes the important changes to this decision guide. For notifications about updates to this guide, you can subscribe to an RSS feed.

Change	Description	Date
Minor update	Updated content for Amazon Q and Amazon's latest AI and ML stack.	May 3, 2024
Initial release	Initial release of decision guide.	July 24, 2023