



AWS Guia de decisão

Escolhendo um serviço AWS de aprendizado de máquina



Escolhendo um serviço AWS de aprendizado de máquina: AWS Guia de decisão

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

As marcas comerciais e imagens de marcas da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestige a Amazon. Todas as outras marcas comerciais que não pertencem à Amazon pertencem a seus respectivos proprietários, que podem ou não ser afiliados, conectados ou patrocinados pela Amazon.

Table of Contents

Guia de decisão 1

Introdução 1

Compreendo 2

Considere 4

Escolher 7

Use 10

Explore 22

Recursos 23

Histórico do documentos 25

..... xxvi

Escolhendo um serviço AWS de aprendizado de máquina

Escolha os serviços e estruturas de ML certos para apoiar seu trabalho

Finalidade	Ajude a determinar quais serviços de AWS ML são mais adequados às suas necessidades.
Última atualização	3 de maio de 2024
Serviços cobertos	• Amazon Augmented AI • Amazon CodeGuru • Amazon Comprehend • DevOpsGuru da Amazon • Amazon Forecast • Amazon Kendra • Amazon Lex • Amazon Personalize • Amazon Polly • Amazon Rekognition • SageMaker IA da Amazon • Amazon Textract • Amazon Transcribe • Amazon Translate

Introdução

Basicamente, o aprendizado de máquina (ML) foi projetado para fornecer ferramentas e serviços digitais para aprender com dados, identificar padrões, fazer previsões e, em seguida, agir de acordo com essas previsões. Quase todos os sistemas de inteligência artificial (IA) atuais são criados usando ML. O ML usa grandes quantidades de dados para criar e validar a lógica de decisão. Essa lógica de decisão forma a base do modelo de IA.

Os cenários em que os serviços de aprendizado de AWS máquina podem ser aplicados incluem:

- Casos de uso específicos — os serviços de aprendizado de AWS máquina podem apoiar seus casos de uso baseados em IA com uma ampla variedade de algoritmos, modelos e soluções pré-criados para casos de uso e setores comuns. Você pode escolher entre 23 serviços pré-treinados, incluindo Amazon Personalize, Amazon Kendra e Amazon Monitron.
- Personalização e escalabilidade do aprendizado de máquina — A Amazon SageMaker AI foi projetada para ajudar você a criar, treinar e implantar modelos de ML para qualquer caso de uso. Você pode criar seus próprios modelos básicos ou acessar modelos básicos de código aberto AWS por meio do Amazon SageMaker AI e do Amazon Bedrock.
- Acesso à infraestrutura especializada — Use as estruturas e a infraestrutura de ML fornecidas por AWS quando você precisar de ainda mais flexibilidade e controle sobre seus fluxos de trabalho de aprendizado de máquina e estiver disposto a gerenciar a infraestrutura e os recursos subjacentes por conta própria.

Este guia de decisão ajudará você a fazer as perguntas certas, avaliar seus critérios e problemas comerciais e determinar quais serviços são mais adequados às suas necessidades.

Compreendo

À medida que as organizações continuam adotando tecnologias de IA e ML, a importância de entender e escolher entre os serviços de AWS ML é um desafio contínuo.

AWS fornece uma variedade de serviços de ML projetados para ajudar as organizações a criar, treinar e implantar modelos de ML com mais rapidez e facilidade. Esses serviços podem ser usados para resolver uma ampla variedade de problemas comerciais, como previsão de rotatividade de clientes, detecção de fraudes e reconhecimento de imagem e fala.

What is it?



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



Classification AI and Predictive AI

A subset of ML that recognizes patterns to identify something (Classification AI) or predicts future trends based on statistical patterns and historical data (Predictive AI)



Generative AI

A subset of DL that can create new content and ideas powered by large, pretrained models called foundation models (FMs)

Antes de nos aprofundarmos nos serviços de AWS ML, vamos analisar a relação entre IA e ML.

- Em um alto nível, a inteligência artificial é uma forma de descrever qualquer sistema que possa replicar tarefas que antes exigiam inteligência humana. A maioria dos casos de uso de IA busca um resultado probabilístico: fazer uma previsão ou decisão com um alto grau de certeza, semelhante ao julgamento humano.
- Quase todos os sistemas de IA atuais são criados usando aprendizado de máquina. O ML usa grandes quantidades de dados para criar e validar a lógica de decisão, conhecida como modelo.
- A IA de classificação é um subconjunto do ML que reconhece padrões para identificar algo. A IA preditiva é um subconjunto do ML que prevê tendências futuras com base em padrões estatísticos e dados históricos.
- Por fim, a IA generativa é um subconjunto do aprendizado profundo que pode criar novos conteúdos e ideias, como conversas, histórias, imagens, vídeos e músicas. A IA generativa é alimentada por modelos muito grandes que são pré-treinados em vastos conjuntos de dados, chamados de Modelos Fundamentais ou FMs. [O Amazon Bedrock](#) é um serviço totalmente gerenciado que oferece uma opção de alto desempenho FMs para criar e escalar aplicativos generativos de IA. [O Amazon Q Developer](#) e o [Amazon Q Business](#) são assistentes com inteligência artificial generativa para casos de uso específicos.

Este guia foi desenvolvido principalmente para cobrir serviços nas categorias de aprendizado de máquina de IA de classificação e IA preditiva.

Além disso, AWS oferece hardware especializado e acelerado para treinamento e inferência de ML de alto desempenho.

- As instâncias [Amazon EC2 P5](#) são equipadas com o NVIDIA H100 Tensor Core GPUs, que é adequado para tarefas de treinamento e inferência em aprendizado de máquina. As instâncias [Amazon EC2 G5](#) apresentam até 8 processadores NVIDIA A10G Tensor Core e AMD EPYC de segunda geração GPUs, para uma ampla variedade de casos de uso de aprendizado de máquina e com uso intensivo de gráficos.
- [AWS Trainium](#) é o acelerador de ML de segunda geração desenvolvido especificamente para o AWS treinamento de aprendizado profundo (DL) de mais de 100 bilhões de modelos de parâmetros.
- AWS Inferentia As [instâncias Amazon EC2 Inf2 baseadas em 2](#) foram projetadas para oferecer alto desempenho ao menor custo na Amazon EC2 para seus aplicativos de inferência de DL e IA generativa.

Considere

Ao resolver um problema comercial com serviços de AWS ML, a consideração de vários critérios-chave pode ajudar a garantir o sucesso. A seção a seguir descreve alguns dos principais critérios a serem considerados ao escolher um serviço de ML.

Problem definition

Definição do problema

A primeira etapa do ciclo de vida do ML é enquadrar o problema comercial. Entender o problema que você está tentando resolver é essencial para escolher o serviço de AWS ML certo, pois serviços diferentes são projetados para resolver problemas diferentes. Também é importante determinar se o ML é a melhor opção para seu problema comercial.

Depois de determinar que o ML é a melhor opção, você pode começar escolhendo entre uma variedade de serviços de AWS IA específicos (em áreas como fala, visão e documentos).

A Amazon SageMaker AI fornece uma infraestrutura totalmente gerenciada se você precisar criar e treinar seus próprios modelos. AWS oferece uma variedade de estruturas de ML avançadas e opções de infraestrutura para os casos em que você precisa de modelos de ML altamente

personalizados e especializados. AWS também oferece um amplo conjunto de modelos básicos populares para criar novos aplicativos com IA generativa.

ML algorithm

algoritmo de ML

A escolha do algoritmo de ML para o problema comercial que você está tentando resolver depende do tipo de dados com os quais você está trabalhando, bem como dos resultados desejados. As informações a seguir descrevem como cada uma das principais categorias de serviços de AWS IA/ML permite que você trabalhe com seus algoritmos:

- Serviços especializados de IA: esses serviços oferecem uma capacidade limitada de personalizar o algoritmo de ML, pois são modelos pré-treinados otimizados para tarefas específicas. Normalmente, você pode personalizar os dados de entrada e alguns parâmetros, mas não tem acesso aos modelos de ML subjacentes nem à capacidade de criar seus próprios modelos.
- Amazon SageMaker AI: esse serviço oferece a maior flexibilidade e controle sobre o algoritmo de ML. Você pode usar a SageMaker IA para criar modelos personalizados usando seus próprios algoritmos e estruturas ou usar modelos e algoritmos pré-criados fornecidos pela AWS. Isso permite um alto grau de personalização e controle sobre o processo de ML.
- Estruturas e infraestrutura de ML de nível inferior: esses serviços oferecem a maior flexibilidade e controle sobre o algoritmo de ML. Você pode usar esses serviços para criar modelos de ML altamente personalizados usando seus próprios algoritmos e estruturas. No entanto, o uso desses serviços requer uma experiência significativa em ML e pode não ser viável para todos os casos de uso.

Security

Segurança

Se você precisar de um endpoint privado em sua VPC, suas opções variarão de acordo com a camada AWS de serviços de ML que você está usando. Isso inclui:

- Serviços especializados de IA: a maioria dos serviços especializados de IA atualmente não oferece suporte a endpoints privados em VPCs. No entanto, o Amazon Rekognition Custom Labels e o Amazon Comprehend Custom podem ser acessados usando VPC endpoints.
- Os principais serviços de IA: Amazon Translate, Amazon Transcribe e Amazon Comprehend oferecem suporte a endpoints de VPC.

- Amazon SageMaker SageMaker AI: A IA fornece suporte integrado para VPC endpoints, permitindo que você implante seus modelos treinados como um endpoint acessível somente de dentro da VPC.
- Estruturas e infraestrutura de ML de nível inferior: você pode implantar seus modelos em EC2 instâncias da Amazon ou em contêineres dentro de sua VPC, fornecendo controle total sobre a configuração da rede.

Latency

Latência

Serviços de IA de alto nível, como o Amazon Rekognition e o Amazon Transcribe, foram projetados para lidar com uma ampla variedade de casos de uso e oferecer alto desempenho em termos de velocidade. No entanto, eles podem não atender a determinados requisitos de latência.

Se você estiver usando estruturas e infraestrutura de ML de nível inferior, recomendamos usar a Amazon AI. SageMaker Essa opção geralmente é mais rápida do que criar modelos personalizados devido ao serviço totalmente gerenciado e às opções de implantação otimizadas. Embora um modelo personalizado altamente otimizado possa superar a SageMaker IA, ele exigirá conhecimentos e recursos significativos para ser construído.

Accuracy

Precisão

A precisão dos serviços de AWS ML varia com base no caso de uso específico e no nível de personalização necessário. Serviços de IA de alto nível, como o Amazon Rekognition, são baseados em modelos pré-treinados que foram otimizados para tarefas específicas e oferecem alta precisão em muitos casos de uso.

Em alguns casos, você pode optar por usar o Amazon SageMaker AI, que fornece uma plataforma mais flexível e personalizável para criar e treinar modelos personalizados de ML. Ao criar seus próprios modelos, você poderá obter uma precisão ainda maior do que a possível com modelos pré-treinados.

Você também pode optar por usar estruturas e infraestrutura de ML, como TensorFlow o Apache MXNet, para criar modelos altamente personalizados que ofereçam a maior precisão possível para seu caso de uso específico.

AWS and responsible AI

AWS e IA responsável

AWS constrói modelos básicos (FMs) com a IA responsável em mente em cada estágio de seu processo de desenvolvimento. Ao longo do projeto, desenvolvimento, implantação e operações, consideramos uma série de fatores, incluindo:

1. Precisão (até que ponto um resumo corresponde ao documento subjacente; se a biografia está factualmente correta)
2. Justiça (se os resultados tratam grupos demográficos de forma semelhante)
3. Considerações sobre propriedade intelectual e direitos autorais
4. Uso adequado (filtrando as solicitações do usuário por aconselhamento jurídico, diagnósticos médicos ou atividades ilegais)
5. Toxicidade (discurso de ódio, palavrões e insultos)
6. Privacidade (protegendo informações pessoais e solicitações do cliente)

AWS cria soluções para resolver esses problemas nos processos usados para adquirir dados de treinamento, FMs neles mesmos e na tecnologia usada para pré-processar as solicitações do usuário e as saídas pós-processamento.

Escolher

Agora que você conhece os critérios pelos quais avaliará suas opções de serviço de ML, você está pronto para escolher qual serviço de AWS ML é adequado às suas necessidades organizacionais. A tabela a seguir destaca quais serviços de ML são otimizados para quais circunstâncias. Use-o para ajudar a determinar o serviço de AWS ML mais adequado ao seu caso de uso.

Categorias	Quando você o usaria?	Para que é otimizado ?	AI/ML Serviços ou ambientes relacionados
Casos de uso específicos	Use os serviços de IA fornecidos pela AWS quando	Esses serviços foram projetados para serem fáceis de	Amazon Augmented AI Amazon CodeGuru

Categorias	Quando você o usaria?	Para que é otimizado ?	AI/ML Serviços ou ambientes relacionados
Esses serviços de inteligência artificial têm como objetivo atender a necessidades específicas. Eles incluem personalização, previsão, detecção de anomalias, transcrição de fala e outros. Como são fornecidos como serviços, eles podem ser incorporados aos aplicativos sem exigir nenhum conhecimento em ML.	precisar que funcionalidades específicas e pré-criadas sejam integradas aos seus aplicativos, sem a necessidade de personalizações extensas ou experiência em aprendizado de máquina. Esses serviços foram projetados para serem fáceis de usar e não exigem muita codificação ou configuração.	usar e não exigem muita experiência em codificação, configuração ou ML.	Amazon Comprehend Amazon Comprehend Medical DevOpsGuru da Amazon Amazon Forecast Amazon Kendra Amazon Lex Amazon Personalize Amazon Polly Amazon Rekognition Amazon Textract Amazon Transcribe Amazon Translate

Categorias	Quando você o usaria?	Para que é otimizado ?	AI/ML Serviços ou ambientes relacionados
Serviços de ML Esses serviços podem ser usados para desenvolver modelos ou fluxos de trabalho personalizados de aprendizado de máquina que vão além das funcionalidades pré-criadas oferecidas pelos principais serviços de IA.	Use esses serviços quando precisar de modelos ou fluxos de trabalho de aprendizado de máquina mais personalizados que vão além das funcionalidades pré-criadas oferecidas pelos principais serviços de IA.	Esses serviços são otimizados para criar e treinar modelos personalizados de aprendizado de máquina, treinamento em grande escala em várias instâncias ou clusters de GPU, maior controle sobre a implantação do modelo de aprendizado de máquina, inferência em tempo real e criação de fluxos de trabalho end-to-end	SageMaker IA da Amazon SageMaker IA da Amazon JumpStart SageMaker Estúdio AI SageMaker Tela de IA SageMaker Laboratório AI Studio SageMaker AI Ground Truth PyTorch em AWS Apache MxNet Hugging Face TensorFlow em AWS
Infraestrutura Para implantar o aprendizado de máquina na produção, você precisa de uma infraestrutura econômica, que a Amazon possibilita com o silício AWS incorporado.	Use quando quiser obter o menor custo para modelos de treinamento e precisar executar inferências na nuvem.	Otimizado para oferecer suporte à implantação econômica do aprendizado de máquina.	AWS Trainium AWS Inferentia e Inferentia2 SageMaker IA da Amazon HyperPod

Categorias	Quando você o usaria?	Para que é otimizado ?	AI/ML Serviços ou ambientes relacionados
Ferramentas e serviços associados Essas ferramentas e os serviços associados foram projetados para ajudar você a facilitar a implantação do aprendizado de máquina.	Esses serviços e ferramentas foram projetados para ajudar você a acelerar o aprendizado profundo na nuvem, fornecendo imagens de máquinas, imagens docker e resolução de entidades da Amazon.	Otimizado para ajudar você a acelerar o aprendizado profundo na nuvem.	AMIs de deep learning da AWS AWS Contêineres de Deep Learning AWS Resolução de entidades

Use

Agora que você tem uma compreensão clara dos critérios que precisa aplicar na escolha de um serviço de AWS ML, você pode selecionar quais serviços de AWS AI/ML são otimizados para suas necessidades comerciais.

Para explorar como usar e saber mais sobre os serviços que você escolheu, fornecemos três conjuntos de caminhos para explorar como cada serviço funciona. O primeiro conjunto de caminhos fornece documentação detalhada, tutoriais práticos e recursos para começar a usar o Amazon Comprehend, o Amazon Textract, o Amazon Translate, o Amazon Lex, o Amazon Polly, o Amazon Rekognition e o Amazon Transcribe.

Amazon Comprehend

- Comece a usar o Amazon Comprehend

Use o console do Amazon Comprehend para criar e executar um trabalho assíncrono de detecção de entidades.

[Comece com o tutorial](#)

- Analise insights em texto com o Amazon Comprehend

Saiba como usar o Amazon Comprehend para analisar e extrair insights do texto.

[Comece com o tutorial](#)

- Preços do Amazon Comprehend

Explore informações sobre preços e exemplos do Amazon Comprehend.

[Explore o guia](#)

Amazon Textract

- Comece a usar o Amazon Textract

Saiba como o Amazon Textract pode ser usado com texto formatado para detectar palavras e linhas de palavras localizadas próximas umas das outras, bem como analisar um documento em busca de itens como texto relacionado, tabelas, pares de valores-chave e elementos de seleção.

[Explore o guia](#)

- Extraia texto e dados estruturados com o Amazon Textract

Saiba como usar o Amazon Textract para extrair texto e dados estruturados de um documento.

[Comece com o tutorial](#)

- AWS Power Hour: Machine Learning

Mergulhe no Amazon Textract neste episódio, passe algum tempo nele AWS Management Console e analise exemplos de código que ajudarão você a entender como aproveitar ao máximo o serviço. APIs

[Assista ao vídeo](#)

Amazon Translate

- Comece a usar o Amazon Translate usando o console

A maneira mais fácil de começar com o Amazon Translate é usar o console para traduzir um texto. Saiba como traduzir até 10.000 caracteres usando o console.

[Explore o guia”](#)

- Traduza texto entre idiomas na nuvem

Neste exemplo de tutorial, como parte de uma empresa internacional de fabricação de malas, você precisa entender o que os clientes estão dizendo sobre seu produto em avaliações no idioma do mercado local - francês.

[Comece com o tutorial”](#)

- Preços do Amazon Translate

Explore os preços do Amazon Translate, incluindo o nível gratuito, que fornece 2 milhões de caracteres por mês durante 12 meses.

[Explore o guia”](#)

Amazon Lex

- Guia do desenvolvedor do Amazon Lex V2

Explore informações sobre como começar, como funciona e informações sobre preços do Amazon Lex V2.

[Explore o guia”](#)

- Introdução ao Amazon Lex Apresentamos o serviço conversacional Amazon Lex e mostramos exemplos que mostram como criar um bot e implantá-lo em diferentes serviços de bate-papo.

[Faça o curso”](#) (é necessário fazer login)

- Explorando a IA generativa em experiências de conversação

Explore o uso da IA generativa em experiências de conversação.

[Leia o blog”](#)

Amazon Polly

- O que é Amazon Polly?

Explore uma visão geral completa do serviço de nuvem que converte texto em fala realista e pode ser usado para desenvolver aplicativos para aumentar o engajamento e a acessibilidade do cliente.

[Explore o guia”](#)

- Destaque o texto enquanto ele está sendo falado usando o Amazon Polly

Apresentamos abordagens para destacar o texto enquanto ele é falado para adicionar recursos visuais ao áudio em livros, sites, blogs e outras experiências digitais.

[Leia o blog”](#)

- Crie áudio para conteúdo em vários idiomas com a mesma persona de voz TTS no Amazon Polly

Explicamos o Neural Text-to-Speech (NTTS) e discutimos como um amplo portfólio de vozes disponíveis, fornecendo uma variedade de falantes distintos em idiomas compatíveis, pode funcionar para você.

[Leia o blog”](#)

Amazon Rekognition

- O que é o Amazon Rekognition?

Explore como você pode usar esse serviço para adicionar análise de imagem e vídeo aos seus aplicativos.

[Explore o guia”](#)

- Rekognition prático: análise automatizada de imagens e vídeos

Saiba como o reconhecimento facial funciona com streaming de vídeo, junto com exemplos de código e pontos-chave em um ritmo autoguiado.

[Comece com o tutorial”](#)

- Amazon Rekognition FAQs

Aprenda os conceitos básicos do Amazon Rekognition e como ele pode ajudá-lo a melhorar seu aprendizado profundo e analisar visualmente seus aplicativos.

[Leia o FAQs”](#)

Amazon Transcribe

- O que é o Amazon Transcribe?

Explore o serviço de reconhecimento AWS automático de fala usando ML para converter áudio em texto. Saiba como usar esse serviço como uma transcrição independente ou adicionar speech-to-text recursos a qualquer aplicativo.

[Explore o guia”](#)

- Preços do Amazon Transcribe

Apresentamos a AWS pay-as-you-go transcrição, incluindo opções personalizadas de modelo de linguagem e o nível gratuito do Amazon Transcribe.

[Explore o guia”](#)

- Crie uma transcrição de áudio com o Amazon Transcribe

Aprenda a usar o Amazon Transcribe para criar uma transcrição de texto de arquivos de áudio gravados usando um cenário de caso de uso real para testar de acordo com suas necessidades.

[Comece com o tutorial”](#)

- Crie um aplicativo de streaming Amazon Transcribe

Aprenda a criar um aplicativo para gravar, transcrever e traduzir áudio ao vivo em tempo real, com os resultados enviados diretamente para você.

[Explore o guia”](#)

O segundo conjunto de AWS service (Serviço da AWS) caminhos de IA/ML fornece documentação detalhada, tutoriais práticos e recursos para começar a usar os serviços da família Amazon AI. SageMaker

SageMaker AI

- Como funciona a Amazon SageMaker AI

Explore a visão geral do aprendizado de máquina e como a SageMaker IA funciona.

[Explore o guia”](#)

- Começando com a Amazon SageMaker AI

Saiba como ingressar em um domínio Amazon SageMaker AI, dando acesso ao Amazon SageMaker AI Studio e RStudio à SageMaker IA.

[Explore o guia”](#)

- Use o Apache Spark com a Amazon AI SageMaker

Saiba como usar o Apache Spark para pré-processar dados e a SageMaker IA para treinamento e hospedagem de modelos.

[Explore o guia”](#)

- Use contêineres Docker para criar modelos

Explore como a Amazon SageMaker AI faz uso extensivo de contêineres Docker para tarefas de criação e execução. Saiba como implantar as imagens pré-criadas do Docker para seus algoritmos integrados e as estruturas de aprendizado profundo suportadas usadas para treinamento e inferência.

[Explore o guia”](#)

- Linguagens e estruturas de aprendizado de máquina

Saiba como começar a usar a SageMaker IA usando o Amazon SageMaker AI Python SDK.

[Explore o guia”](#)

SageMaker AI Autopilot

- Crie um experimento Amazon SageMaker AI Autopilot para dados tabulares

Saiba como criar um experimento Amazon SageMaker AI Autopilot para explorar, pré-processar e treinar vários candidatos a modelo em um conjunto de dados tabular.

[Explore o guia”](#)

- Crie modelos de aprendizado de máquina automaticamente

Aprenda a usar o Amazon SageMaker AI Autopilot para criar, treinar e ajustar automaticamente um modelo de ML e implantar o modelo para fazer previsões.

[Comece com o tutorial”](#)

- Explore a modelagem com o Amazon SageMaker AI Autopilot com esses exemplos de notebooks

Explore exemplos de cadernos para marketing direto, previsão de rotatividade de clientes e como trazer seu próprio código de processamento de dados para o Amazon SageMaker AI Autopilot.

[Explore o guia”](#)

SageMaker AI Canvas

- Comece a usar o Amazon SageMaker AI Canvas

Saiba como começar a usar o SageMaker AI Canvas.

[Explore o guia”](#)

- Gere previsões de aprendizado de máquina sem escrever código

Este tutorial explica como usar o Amazon SageMaker AI Canvas para criar modelos de ML e gerar previsões precisas sem escrever uma única linha de código.

[Comece com o tutorial”](#)

- Mergulhe mais fundo no SageMaker AI Canvas

Explore uma visão aprofundada do SageMaker AI Canvas e seus recursos visuais de ML sem código.

[Leia o blog”](#)

- Use o Amazon SageMaker AI Canvas para criar seu primeiro modelo de ML

Saiba como usar o Amazon SageMaker AI Canvas para criar um modelo de ML para avaliar a retenção de clientes, com base em uma campanha de e-mail para novos produtos e serviços.

[Comece com o laboratório”](#)

SageMaker AI Data Wrangler

- Introdução ao Amazon SageMaker AI Data Wrangler

Explore como configurar o SageMaker AI Data Wrangler e, em seguida, forneça um passo a passo usando um exemplo de conjunto de dados existente.

[Explore o guia”](#)

- Prepare dados de treinamento para aprendizado de máquina com o mínimo de código

Saiba como preparar dados para ML usando o Amazon SageMaker AI Data Wrangler.

[Comece com o tutorial”](#)

- SageMaker Workshop de aprofundamento do AI Data Wrangler

Aprenda a aplicar os tipos de análise apropriados em seu conjunto de dados para detectar anomalias e problemas, usar o derivado results/insights para formular ações corretivas durante as transformações em seu conjunto de dados e testar a escolha e a sequência corretas de

transformações usando as opções de modelagem rápida fornecidas pelo AI Data Wrangler.
SageMaker

[Comece com o workshop”](#)

SageMaker AI Ground Truth

- Introdução ao Amazon Ground Truth

Explore como usar o console para criar um trabalho de etiquetagem, atribuir uma força de trabalho pública ou privada e enviar o trabalho de etiquetagem para sua força de trabalho. Saiba como monitorar o progresso de um trabalho de etiquetagem.

[Explore o guia”](#)

- Dados de treinamento de etiquetas para Machine Learning

Saiba como configurar um trabalho de rotulagem no Amazon SageMaker AI Ground Truth para anotar dados de treinamento para seu modelo de ML.

[Comece com o tutorial”](#)

- Introdução ao Amazon Ground Truth Plus Explore como concluir as etapas necessárias para iniciar um projeto Amazon SageMaker AI Ground Truth Plus, revisar rótulos e satisfazer os pré-requisitos do SageMaker AI Ground Truth Plus.

[Explore o guia”](#)

- Comece a usar o Amazon Ground Truth Veja como começar a rotular seus dados em minutos por meio do console SageMaker AI Ground Truth.

[Assista ao vídeo”](#)

- Amazon SageMaker AI Ground Truth Plus — crie conjuntos de dados de treinamento sem código ou recursos internos

Conheça o Ground Truth Plus, um serviço pronto para uso que usa uma força de trabalho especializada para fornecer conjuntos de dados de treinamento de alta qualidade com rapidez e reduzir custos em até 40%.

[Leia o blog”](#)

SageMaker AI JumpStart

- Comece a usar o aprendizado de máquina com SageMaker IA JumpStart

Explore modelos de JumpStart soluções de SageMaker IA que configuram a infraestrutura para casos de uso comuns e exemplos de notebooks executáveis para aprendizado de máquina com SageMaker IA.

[Explore o guia”](#)

- Comece a usar seu projeto de aprendizado de máquina rapidamente usando o Amazon SageMaker AI JumpStart

Saiba como acelerar seu projeto de ML usando modelos pré-treinados e soluções pré-criadas oferecidas pela Amazon AI. SageMaker JumpStart Em seguida, você pode implantar o modelo selecionado por meio dos notebooks Amazon SageMaker AI Studio.

[Comece com o tutorial”](#)

- Experimente a inteligência SageMaker artificial da Amazon JumpStart com este workshop do Immersion Day

Saiba como os recursos de ML de baixo código encontrados no Amazon SageMaker AI Data Wrangler, no Autopilot e no Jumpstart facilitam a experimentação com mais rapidez e trazem modelos altamente precisos para a produção.

[Comece com o workshop”](#)

SageMaker AI Pipelines

- Introdução ao Amazon SageMaker AI Pipelines

Saiba como criar end-to-end fluxos de trabalho que gerenciam e implantam trabalhos de SageMaker IA. SageMaker O AI Pipelines vem com a integração do SageMaker AI Python

SDK, para que você possa criar cada etapa do seu pipeline usando uma interface baseada em Python.

[Explore o guia”](#)

- Automatize os fluxos de trabalho de aprendizado de máquina

Saiba como criar e automatizar fluxos de trabalho end-to-end de aprendizado de máquina (ML) usando o Amazon SageMaker AI Pipelines, o Amazon AI Model Registry e o Amazon SageMaker SageMaker AI Clarify.

[Comece com o tutorial”](#)

- Como criar fluxos de trabalho de ML totalmente automatizados com o Amazon SageMaker AI Pipelines

Saiba mais sobre o Amazon SageMaker AI Pipelines, o primeiro CI/CD serviço de ML do mundo projetado para ser acessível a todos os desenvolvedores e cientistas de dados. SageMaker O AI Pipelines traz CI/CD pipelines para o ML, reduzindo o tempo de codificação necessário.

[Assista ao vídeo”](#)

SageMaker AI Studio

- Crie e treine um modelo de aprendizado de máquina localmente

Aprenda a criar e treinar um modelo de ML localmente em seu notebook Amazon SageMaker AI Studio.

[Comece com o tutorial”](#)

- SageMaker Workshop de integração do AI Studio com EMR

Aprenda a utilizar o processamento distribuído em grande escala para preparar dados e, posteriormente, treinar modelos de aprendizado de máquina.

[Comece com o workshop”](#)

O terceiro conjunto de AWS service (Serviço da AWS) caminhos de IA/ML fornece documentação detalhada, tutoriais práticos e recursos para começar a usar o Amazon Titan e o Amazon Titan. AWS Trainium AWS Inferentia

AWS Trainium

- Escalando o treinamento distribuído com o AWS Trainium Amazon EKS

Saiba como você pode se beneficiar da disponibilidade geral das instâncias Amazon EC2 Trn1 baseadas em AWS Trainium— um acelerador de ML desenvolvido especificamente, otimizado para fornecer uma plataforma de alto desempenho, econômica e altamente escalável para treinar modelos de aprendizado profundo na nuvem.

[Leia o blog”](#)

- Visão geral do AWS Trainium

Conheça AWS Trainium o acelerador de aprendizado de máquina (ML) de segunda geração criado AWS especificamente para o treinamento de aprendizado profundo de mais de 100 bilhões de modelos de parâmetros. Cada instância Trn1 do Amazon Elastic Compute Cloud (EC2) implanta até 16 AWS Trainium aceleradores para oferecer uma solução de alto desempenho e baixo custo para treinamento de aprendizado profundo (DL) na nuvem.

[Explore o guia”](#)

- Recommended Trainium Instances

Explore como as AWS Trainium instâncias são projetadas para fornecer alto desempenho e economia para cargas de trabalho de inferência de modelos de aprendizado profundo.

[Explore o guia”](#)

AWS Inferentia

- Visão geral do AWS Inferentia

Entenda como os aceleradores são projetados AWS para oferecer alto desempenho ao menor custo para seus aplicativos de inferência de aprendizado profundo (DL).

[Explore o guia”](#)

- AWS Inferentia 2 se baseia em AWS Inferentia 1, oferecendo taxa de transferência 4x maior e latência 10x menor

Entenda para que o AWS Inferentia 2 é otimizado e explore como ele foi projetado desde o início para oferecer maior desempenho e, ao mesmo tempo, reduzir o custo LLMs e a inferência generativa de IA.

[Leia o blog”](#)

- Inferência de aprendizado de máquina usando AWS Inferentia

Saiba como criar um cluster Amazon EKS com nós executando instâncias Amazon EC2 Inf1 e (opcionalmente) implantar um aplicativo de amostra. As instâncias Amazon EC2 Inf1 são alimentadas por AWS Inferentia chips, que são personalizados AWS para fornecer inferência de alto desempenho e menor custo na nuvem.

[Explore o guia”](#)

Amazon Titan

- Visão geral do Amazon Titan

Explore como os Amazon Titan FMs são pré-treinados em grandes conjuntos de dados, tornando-os modelos poderosos e de uso geral. Saiba como você pode usá-los como estão — ou de forma privada — para personalizá-los com seus próprios dados para uma tarefa específica sem anotar grandes volumes de dados.

[Explore o guia”](#)

Explore

- Diagramas de arquitetura

Esses diagramas de arquitetura de referência mostram exemplos de serviços de AWS IA e ML em uso.

[Explore os diagramas de arquitetura”](#)

- Documentos técnicos

Explore os whitepapers para ajudar você a começar e aprender as melhores práticas na escolha e uso AI/ML de serviços.

[Explore os whitepapers”](#)

- AWS Soluções

Explore soluções aprovadas e orientações arquitetônicas para casos de uso comuns de serviços de IA e ML.

[Explore as soluções”](#)

Recursos

Modelos de fundação

Os modelos de fundação compatíveis incluem:

- [Claude antrópico](#)
- [Comando e incorporando o Cohere](#)
- [AI21 Laboratórios Jurassic](#)
- [Llama de metal](#)
- [IA Mistral](#)
- [Difusão estável XL](#)
- [Titã da Amazon](#)

Usando o Amazon Bedrock, você pode experimentar uma variedade de modelos básicos e personalizá-los de forma privada com seus dados.

Caso de uso ou serviços específicos do setor

- [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](#)

Postagens de blog associadas

- [Novos recursos significativos facilitam o uso do Amazon Bedrock para criar e escalar aplicativos generativos de IA — e alcançar resultados impressionantes](#)
- [AWS Inferência e AWS Trainium entrega o menor custo para implantar modelos Llama 3 na Amazon AI SageMaker JumpStart](#)
- [Revolucione a satisfação do cliente com modelos de recompensa personalizados para sua empresa na Amazon AI SageMaker](#)
- [Amazon Personalize lança novas receitas que oferecem suporte a catálogos de itens maiores com menor latência](#)

Histórico do documento

A tabela a seguir descreve as mudanças importantes nesse guia de decisão. Para receber notificações sobre atualizações deste guia, você pode assinar um feed RSS.

Alteração	Descrição	Data
Atualização secundária	Conteúdo atualizado para o Amazon Q e a mais recente pilha de IA e ML da Amazon.	3 de maio de 2024
Lançamento inicial	Lançamento inicial do guia de decisão.	24 de julho de 2023

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.