

Questões Assimilares a prova de IA - AWS

Questão 01

Uma empresa de dispositivos digitais deseja prever a demanda do cliente por hardware de memória. A empresa não possui experiência em programação ou conhecimento de algoritmos de aprendizado de máquina (ML) e precisa desenvolver um modelo preditivo orientado por dados. A empresa precisa realizar análises em dados internos e externos.

Qual solução atenderá a esses requisitos?

- Importe os dados para o Amazon SageMaker Canvas. Construa modelos de ML e previsões de demanda selecionando os valores nos dados SageMaker Canvas.
- Importe os dados para o Amazon SageMaker Data Wrangler. Construa modelos de ML e previsões de demanda usando uma receita Amazon Personalize Trending-Now.
- Importe os dados para o Amazon SageMaker Data Wrangler. Crie modelos de ML e previsões de demanda.
- Implemente modelos de ML e previsões de demanda usando o Amazon Forecast

Resposta correta: Importe os dados para o Amazon SageMaker Canvas. Construa modelos de ML e previsões de demanda selecionando os valores nos dados SageMaker Canvas.

- *Explicação: O Amazon SageMaker Canvas é uma ferramenta no-code da AWS que permite criar modelos de aprendizado de máquina sem necessidade de experiência em programação ou conhecimento profundo de ML. Ele facilita a importação de dados internos e externos, a construção de modelos preditivos e a geração de previsões com poucos cliques.*

Questão 02

Uma empresa está treinando um modelo fundamental (FM). A empresa deseja aumentar a precisão do modelo até um nível de aceitação específico.

Qual solução atenderá a esses requisitos?

- Reduzir o número de épocas (epochs):
- Reduzir o tamanho do lote (batch size)

- Aumentar o número de épocas (epochs)
- Aumentar o parâmetro de temperatura (temperature parameter)

Resposta correta: Aumentar o número de épocas (epochs)

- *Explicação: Treinar um modelo por mais épocas permite que ele aprenda melhor os padrões dos dados, o que pode aumentar a precisão. No entanto, isso deve ser feito com cautela para evitar overfitting.*

Questão 03

Uma empresa construiu um chatbot que pode responder a perguntas em linguagem natural com imagens. A empresa deseja garantir que o chatbot não retorne imagens inadequadas ou indesejadas.

- Implementar APIs de moderação (Implement moderation APIs)
- Automatizar a integração do feedback do usuário:
- Realizar validação do modelo
- Treinar novamente o modelo com um conjunto de dados público geral:

Resposta correta: Implementar APIs de moderação (Implement moderation APIs)

- *Explicação: APIs de moderação, como Amazon Rekognition e AWS Content Moderation, detectam conteúdo impróprio automaticamente, garantindo que o chatbot não exiba imagens inadequadas.*

Questão 04

Uma empresa de empréstimos está construindo uma solução baseada em IA generativa para oferecer descontos a novos solicitantes com base em critérios específicos de negócios. A empresa deseja construir e usar um modelo de IA de forma responsável para minimizar vieses que possam afetar negativamente alguns clientes. Quais ações a empresa deve tomar para atender a esses requisitos? (Escolha duas.)

- Garantir que o modelo seja executado frequentemente
- Usar a técnica Recall-Oriented Understudy for Gisting Evaluation (ROUGE) para garantir que o modelo seja 100% preciso.

- Avaliar o comportamento do modelo para que a empresa possa fornecer transparência aos stakeholders.
- Garantir que o tempo de inferência do modelo esteja dentro dos limites aceitáveis.
- Detectar desequilíbrios ou disparidades nos dados

Respostas corretas:

- **Avaliar o comportamento do modelo para que a empresa possa fornecer transparência aos stakeholders.**
 - *Permite entender como o modelo toma decisões e identificar possíveis vieses.*
- **Detectar desequilíbrios ou disparidades nos dados.**
 - *Garante que o modelo não esteja treinado com dados enviesados que possam prejudicar certos grupos.*

Questão 05

Uma empresa de e-commerce deseja criar uma solução para determinar os sentimentos dos clientes com base em análises escritas sobre produtos.

Quais serviços da AWS atendem a esses requisitos? (Escolha duas.)

- Amazon Comprehend
- Amazon Lex
- Amazon Rekognition
- Amazon Bedrock
- Amazon Polly

Respostas corretas:

- **Amazon Comprehend**
 - *Serviço da AWS para Processamento de Linguagem Natural (NLP), usado para análise de sentimentos.*
- **Amazon Bedrock**
 - *Permite usar modelos de IA generativa que podem ser ajustados para interpretar sentimentos em textos.*

Questão 06

Uma empresa médica está personalizando um modelo fundamental (FM) para fins de diagnóstico. A empresa precisa que o modelo seja transparente e explicável para atender aos requisitos regulatórios.

- Gerar métricas simples, relatórios e exemplos usando o Amazon SageMaker Clarify
- Coletar mais dados. Usar o Amazon Rekognition para adicionar rótulos personalizados aos dados
- Configurar a segurança e a conformidade usando o Amazon Inspector
- Criptografar e proteger os dados de treinamento usando o Amazon Macie

Resposta correta:

Gerar métricas simples, relatórios e exemplos usando o Amazon SageMaker Clarify.

- *O SageMaker Clarify ajuda a identificar e explicar previsões do modelo, garantindo conformidade regulatória e transparência.*

Questão 07

CLOUD KNITTERS COMMUNITY

Qual estratégia de aprendizado de IA fornece essa capacidade de autoaperfeiçoamento?

- Aprendizado por reforço com recompensas por feedback positivo dos clientes
- Aprendizado supervisionado com um conjunto de dados manualmente organizado de boas e más respostas
- Aprendizado supervisionado com um banco de dados de FAQ continuamente atualizado:
- Aprendizado não supervisionado para encontrar clusters de consultas similares dos clientes:

Resposta correta:

Aprendizado por reforço com recompensas por feedback positivo dos clientes.

- Permite que a IA melhore suas respostas com base em interações reais e recompensas, ajustando o comportamento ao longo do tempo.

Questão 08

Uma empresa está desenvolvendo uma aplicação que precisa gerar dados sintéticos com base em dados existentes.

Qual tipo de modelo a empresa pode usar para atender a esse requisito?

- Rede adversária generativa (GAN).
- XGBoost.
- WaveNet
- Rede neural residual



Resposta correta:

Rede adversária generativa (GAN).

- As GANs criam dados sintéticos realistas ao treinar dois modelos simultaneamente: um gerador e um discriminador.

Questão 09

Uma empresa deseja criar um aplicativo usando o Amazon Bedrock. A empresa possui um orçamento limitado e prefere flexibilidade sem compromisso de longo prazo.

Qual modelo de precificação do Amazon Bedrock atende a esses requisitos?

- Personalização de modelo (Model customization)
- Instância spot (Spot Instance)
- Taxa provisionada (Provisioned Throughput)
- Sob demanda (On-Demand)

Resposta correta:

Sob demanda (On-Demand).

- Permite pagar apenas pelo uso, sem necessidade de compromisso de longo prazo, ideal para empresas com orçamento limitado.

Questão 10

Uma empresa construiu um modelo de deep learning para detecção de objetos e implantou o modelo em produção.

Qual processo de IA ocorre quando o modelo analisa uma nova imagem para identificar objetos?

- Treinamento (Training)
- Implantação do modelo (Model deployment)
- Inferência (Inference)
- Correção de viés (Bias correction)

Resposta correta:

Inferência (Inference).

- Quando um modelo de IA já treinado analisa novos dados (como imagens) e gera previsões, esse processo é chamado de inferência.

Questão 11

Uma empresa de mídia social deseja usar um modelo de linguagem grande (LLM) para moderação de conteúdo. A empresa quer avaliar as saídas do LLM quanto a preconceitos e possível discriminação contra grupos ou indivíduos específicos.

Qual fonte de dados a empresa deve usar para avaliar as saídas do LLM com o MENOR esforço administrativo?

- Logs de moderação.
- Diretrizes de moderação de conteúdo.
- Conteúdo gerado por usuários
- Conjuntos de dados de benchmark.

Resposta correta:

Conjuntos de dados de benchmark.

- Os conjuntos de dados de benchmark contêm exemplos padronizados usados para medir viés e discriminação em modelos de IA, exigindo o menor esforço administrativo para avaliação.

Questão 12

Uma empresa deseja usar IA para proteger seu aplicativo contra ameaças. A solução de IA precisa verificar se um endereço IP é de uma fonte suspeita.

Qual solução atende a esses requisitos?

- Criar um sistema de previsão de fraude
- Criar um sistema de reconhecimento de entidades nomeadas (NLP)
- Desenvolver um sistema de detecção de anomalias
- Criar um sistema de reconhecimento de fala

Resposta correta:

Desenvolver um sistema de detecção de anomalias.

- A detecção de anomalias identifica padrões incomuns em dados, ajudando a reconhecer endereços IP suspeitos com base em comportamento atípico.

CLOUD KNITTERS
COMMUNITY

Questão 13

Uma empresa está utilizando modelos específicos de domínio. A empresa deseja evitar criar novos modelos do zero. Em vez disso, a empresa quer adaptar modelos pré-treinados para criar modelos para novas tarefas.

Qual estratégia de Machine Learning atende a esses requisitos?

- Aumentar o número de épocas
- Diminuir o número de épocas
- Usar aprendizado não supervisionado
- Usar aprendizado por transferência (Transfer learning)

Resposta correta:

Usar aprendizado por transferência (Transfer learning).

- *O aprendizado por transferência permite reaproveitar modelos já treinados, reduzindo o tempo e os recursos necessários para treinar um novo modelo do zero.*

Questão 14

Uma empresa deseja construir um modelo de Machine Learning (ML) usando o Amazon SageMaker. A empresa precisa compartilhar e gerenciar variáveis para o desenvolvimento de modelos entre várias equipes.

Qual funcionalidade do SageMaker atende a esses requisitos?

- Amazon SageMaker Feature Store
- Amazon SageMaker Model Cards
- Amazon SageMaker Clarify
- Amazon SageMaker Data Wrangler

Resposta correta:

Amazon SageMaker Feature Store.

- *O SageMaker Feature Store armazena e gerencia features de Machine Learning, permitindo que equipes compartilhem variáveis para o desenvolvimento de modelos.*

Questão 15

Qual recurso do Amazon OpenSearch Service oferece às empresas a capacidade de criar aplicativos de banco de dados vetorial?

- Suporte para indexação e consultas geoespaciais
- Gerenciamento escalável de índices e capacidade de busca por vizinhos mais próximos
- Capacidade de realizar análise em tempo real em dados de streaming
- Integração com o Amazon S3 para armazenamento de objetos

Resposta correta:

Gerenciamento escalável de índices e busca por vizinhos mais próximos.

- *Esse recurso permite armazenar e recuperar vetores de maneira eficiente, facilitando a busca por similaridade em grandes volumes de dados.*

Questão 16

Uma empresa está usando um modelo de linguagem grande (LLM) no Amazon Bedrock para construir um chatbot. O chatbot processa solicitações de suporte ao cliente. Para resolver uma solicitação, o cliente e o chatbot devem interagir algumas vezes.

Qual solução permite que o LLM use o conteúdo das mensagens anteriores do cliente?

- Ativar o registro de invocações do modelo para coletar mensagens.
- Usar Throughput Provisionado para o LLM.
- Usar o Amazon Personalize para salvar o histórico de conversas.
- Adicionar mensagens ao prompt do modelo.

Resposta correta:

Adicionar mensagens ao prompt do modelo.

- *Incluir mensagens anteriores no prompt do modelo garante que o chatbot mantenha o contexto da conversa e forneça respostas mais relevantes.*

Questão 17

Uma empresa deseja usar um modelo de IA generativa pré-treinada para gerar conteúdo para suas campanhas de marketing. A empresa precisa garantir que o conteúdo gerado esteja alinhado com a voz e os requisitos de mensagens da marca da empresa.

Qual solução atende a esses requisitos?

- Criar prompts eficazes que forneçam instruções claras e contexto para orientar a geração do modelo.

- Aumentar a complexidade do modelo adicionando mais camadas à arquitetura do modelo.
- Otimizar a arquitetura e os hiperparâmetros do modelo para melhorar o desempenho geral do modelo.
- Selecionar um conjunto de dados para aumentar a complexidade do modelo.

Resposta correta:

Criar prompts eficazes que fornecam instruções claras e contexto.

- *Prompts bem estruturados orientam a IA a gerar conteúdo alinhado à identidade da marca e aos requisitos de comunicação.*

Questão 18

Um praticante de IA deseja usar um modelo fundamental (FM) para projetar um aplicativo de busca. O aplicativo de busca deve lidar com consultas que contenham texto e imagens.

Qual tipo de FM o praticante de IA deve usar para alimentar o aplicativo de busca?

- Modelo de geração de imagens
- Modelo de incorporação de texto
- **Modelo de incorporação multimodal (Multi-modal embedding model)**
- Modelo de geração multimodal

Resposta correta:

Modelo de incorporação multimodal (Multi-modal embedding model).

- *Esse modelo processa múltiplos tipos de dados, permitindo buscas que combinem texto e imagens em um único sistema.*

Questão 19

Uma instituição financeira está usando o Amazon Bedrock para desenvolver um aplicativo de IA. O aplicativo está hospedado em uma VPC. Para atender aos

padrões de conformidade regulatória, a VPC não tem permissão para acessar nenhum tráfego da internet.

Qual serviço ou recurso da AWS atenderá a esses requisitos?

- Internet gateway
- Amazon Macie
- AWS PrivateLink
- Amazon CloudFront

Resposta correta:

AWS PrivateLink.

- O AWS PrivateLink permite a comunicação privada entre serviços dentro da VPC, garantindo conformidade sem a necessidade de acesso à internet.

Questão 20

Uma empresa de IA avalia periodicamente seus sistemas e processos com a ajuda de fornecedores independentes de software (ISVs). A empresa precisa receber notificações por e-mail quando os relatórios de conformidade de um ISV estiverem disponíveis.

Qual serviço da AWS a empresa pode usar para atender a esse requisito?

- AWS Audit Manager
- AWS Data Exchange
- AWS Trusted Advisor
- AWS Artifact

Resposta correta:

AWS Artifact.

- O AWS Artifact fornece relatórios de conformidade e permite configurar notificações quando novos relatórios ficam disponíveis.

Questão 21

Uma empresa possui petabytes de dados de clientes sem rótulos para usar em uma campanha publicitária. A empresa deseja classificar seus clientes em níveis (tiers) para anunciar e promover os produtos da empresa.

Qual metodologia a empresa deve usar para atender a esses requisitos?

- Aprendizado por reforço com feedback humano (RLHF)
- Aprendizado supervisionado
- Aprendizado não supervisionado (Unsupervised learning)
- Aprendizado por reforço

Resposta correta:

Aprendizado não supervisionado (Unsupervised learning).

- *O aprendizado não supervisionado identifica padrões e segmenta clientes automaticamente, sem necessidade de dados previamente rotulados.*

Questão 22

Uma empresa está usando a Matriz de Escopo de Segurança de IA Generativa para avaliar responsabilidades de segurança em suas soluções. A empresa identificou quatro diferentes escopos de solução com base na matriz. Qual escopo de solução dá à empresa a MAIOR responsabilidade sobre a segurança?

- Construir e treinar um modelo de IA generativa do zero usando dados específicos que o cliente possui
- Refinar um modelo fundamental (FM) de IA generativa existente de terceiros ajustando o modelo com dados específicos do negócio
- Usar um aplicativo corporativo de terceiros que tenha recursos de IA generativa embutidos
- Construir um aplicativo usando um modelo fundamental (FM) de IA generativa e usando tecnologias de terceiros

Resposta correta:

Construir e treinar um modelo de IA generativa do zero usando dados específicos que o cliente possui.

- *Esse escopo exige controle total sobre dados, infraestrutura e segurança, atribuindo à empresa a maior responsabilidade sobre a solução.*

Questão 23

Uma empresa deseja usar IA generativa para aumentar a produtividade dos desenvolvedores e o desenvolvimento de software. A empresa quer usar o Amazon Q Developer.

O que o Amazon Q Developer pode fazer para ajudar a empresa a atender a esses requisitos?

- Habilitar comandos de voz para codificação e oferecer pesquisa em linguagem natural
- Criar trechos de software, rastreamento de referências e rastreamento de licenças de código aberto
- Converter arquivos de áudio em documentos de texto usando modelos de Machine Learning
- Executar um aplicativo sem provisionar ou gerenciar servidores

Resposta correta:

Criar trechos de software, rastreamento de referências e rastreamento de licenças de código aberto.

- *O Amazon Q Developer auxilia desenvolvedores sugerindo trechos de código, rastreando referências e verificando conformidade com licenças de código aberto.*

Questão 24

Uma empresa está desenvolvendo um aplicativo de central de atendimento e deseja obter insights a partir de conversas com os clientes. A empresa quer analisar e extrair informações-chave do áudio das chamadas dos clientes.

Qual solução atende a esses requisitos?

- Construir um chatbot conversacional usando o Amazon Lex
- Transcrever gravações de chamadas usando o Amazon Transcribe
- Extraír informações das gravações de chamadas usando o Amazon SageMaker Model Monitor
- Criar rótulos de classificação usando o Amazon Comprehend

Resposta correta:

Transcrever gravações de chamadas usando o Amazon Transcribe.

- O Amazon Transcribe converte gravações de áudio em texto, permitindo a análise e extração de informações-chave das interações com clientes.

Questão 25

Uma empresa deseja construir uma aplicação de aprendizado de máquina (ML).

Selecione e ordene os passos corretos da lista a seguir para desenvolver uma carga de trabalho de ML bem arquitetada. Cada passo deve ser selecionado apenas uma vez.

Tradução:

Implantar o modelo (Deploy model)

Desenvolver o modelo (Develop model)

Monitorar o modelo (Monitor model)

Definir metas de negócios e formular o problema de ML (Define business goal and frame ML problem)

Step 1:

Select...

- Deploy model
- Develop model
- Monitor mode
- Define business goal and frame ML problem

Resposta correta:

- Step 1: Define business goal and frame ML problem
 - Step 2: Develop model
 - Step 3: Deploy model
 - Step 4: Monitor model
-
- *O ciclo de vida de um projeto de ML começa com a definição do problema, seguida pelo desenvolvimento e implantação do modelo, finalizando com o monitoramento contínuo.*

Questão 26

Uma empresa de segurança está usando o Amazon Bedrock para executar modelos fundamentais (FMs). A empresa quer garantir que apenas usuários autorizados

invoquem os modelos. A empresa precisa identificar qualquer tentativa de acesso não autorizado para definir políticas apropriadas de IAM (Identity and Access Management) e papéis para futuras iterações dos FMs.

Qual serviço da AWS a empresa deve usar para identificar usuários não autorizados tentando acessar o Amazon Bedrock?

- AWS Audit Manager
- AWS CloudTrail
- AWS Trusted Advisor
- Amazon Fraud Detector

Resposta correta:

AWS CloudTrail.

- *O AWS CloudTrail registra todas as ações realizadas em serviços da AWS, permitindo a identificação de acessos não autorizados ao Amazon Bedrock.*

Questão 27

Um escritório de advocacia deseja criar um aplicativo de IA usando grandes modelos de linguagem (LLMs). O aplicativo lerá documentos legais e extrairá pontos-chave desses documentos.

Qual solução atende a esses requisitos?

- Desenvolver um chatbot de sumarização
- Criar um mecanismo de recomendação
- Desenvolver um sistema de tradução multilingue
- Construir um sistema automático de reconhecimento de entidades nomeadas:

Resposta correta:

Desenvolver um chatbot de sumarização.

- *Um chatbot de sumarização pode ler documentos legais extensos e gerar resumos com os principais pontos, tornando a análise de informações mais eficiente.*

Questão 28

Uma empresa está usando um modelo de linguagem grande pré-treinado (LLM) para construir um chatbot para recomendações de produtos. A empresa precisa que as respostas do LLM sejam curtas e escritas em um idioma específico.

Qual solução alinhará a qualidade das respostas do LLM com as expectativas da empresa?

- Aumentar a temperatura
- Escolher um LLM de tamanho diferente
- Ajustar o prompt
- Aumentar o valor de Top K

Resposta correta:

Ajustar o prompt.

- *Um prompt bem estruturado pode direcionar o LLM a gerar respostas curtas e no idioma desejado, garantindo melhor alinhamento com as necessidades da empresa.*

Questão 29

Qual métrica mede a eficiência de tempo de execução na operação de modelos de IA?

- Tempo de treinamento para cada época
- Pontuação de satisfação do cliente (CSAT)
- Número de instâncias de treinamento
- Tempo médio de resposta (Average response time)

Resposta correta:

Tempo médio de resposta (Average response time).

- *O tempo médio de resposta mede a eficiência operacional do modelo ao processar solicitações e gerar respostas rapidamente.*

Questão 30

Uma empresa deseja criar um chatbot para ajudar os clientes. O chatbot ajudará a resolver problemas técnicos sem intervenção humana. A empresa escolheu um modelo fundamental (FM) para o chatbot. O chatbot precisa produzir respostas que estejam alinhadas ao tom da empresa.

Qual solução atende a esses requisitos?

- Definir um limite baixo para o número de tokens que o FM pode produzir
- Definir um número maior para o parâmetro de temperatura
- Usar inferência em lote (batch inferencing) para produzir respostas detalhadas
- Experimentar e refinar o prompt até que o FM produza as respostas desejadas

Resposta correta:

Experimentar e refinar o prompt até que o FM produza as respostas desejadas.

- *Ajustar o prompt permite controlar o estilo, tom e qualidade das respostas do chatbot, garantindo que estejam alinhadas à identidade da empresa.*

Questão 31

Uma empresa percebe que seu modelo de base (FM) gera imagens que não estão relacionadas aos prompts fornecidos. A empresa deseja modificar as técnicas de prompts para reduzir a geração de imagens não relacionadas.

Qual solução atende a esses requisitos?

- Usar prompts ambíguos:
- Usar prompts negativos
- Usar prompts de zero-shot:
- Usar prompts positivos:

Resposta correta:

Usar prompts negativos.

- *Prompts negativos ajudam a restringir a geração de imagens ao especificar o que deve ser evitado na saída, melhorando a relevância das imagens geradas*

Questão 32

Uma empresa de manufatura utiliza IA para inspecionar produtos e identificar quaisquer danos ou defeitos.

Que tipo de aplicação de IA a empresa está utilizando?

- Visão computacional
- Processamento de linguagem natural (NLP)
- Processamento de imagem
- Sistema de recomendação

Resposta correta:

Visão computacional.

- *A visão computacional permite que a IA analise imagens ou vídeos para identificar danos ou defeitos nos produtos.*



Questão 33

Uma empresa desenvolveu um modelo de sumarização de texto gerativo usando Amazon Bedrock. A empresa utilizará as capacidades de avaliação automática de modelos do Amazon Bedrock.

Qual métrica a empresa deve usar para avaliar a precisão do modelo?

- Pontuação F1:
- Área sob a curva ROC (AUC)
- Pontuação BERT (BERTScore)
- Pontuação de conhecimento do mundo real (RWK)

Resposta correta:

Pontuação BERT (BERTScore).

- O BERTScore mede a similaridade entre o resumo gerado e o esperado, avaliando a precisão da sumarização.

Questão 34

Uma loja de varejo deseja prever a demanda de um produto específico para as próximas semanas usando o algoritmo de previsão Amazon SageMaker DeepAR.

Qual tipo de dado atenderá a esse requisito?

- Dados binários:
- Dados textuais:
- Dados de imagem:
- Dados de séries temporais

Resposta correta:

Dados de séries temporais.

- O DeepAR é projetado para prever tendências futuras com base em padrões históricos de séries temporais.



Questão 35

um sistema de aprendizado de máquina (ML) para ajudar a equipe de gerenciamento de risco a decidir sobre alocações de empréstimos para diferentes grupos demográficos.

O que o banco deve fazer para desenvolver um modelo de ML imparcial?

- Criar um modelo de ML diferente para cada grupo demográfico
- Reduzir o tamanho do conjunto de dados de treinamento
- Medir o desequilíbrio de classes no conjunto de dados de treinamento. Adaptar o processo de treinamento de acordo.
- Garantir que as previsões do modelo de ML sejam consistentes com os resultados históricos.

Resposta correta:

Medir o desequilíbrio de classes no conjunto de dados de treinamento.

Adaptar o processo de treinamento de acordo.

- Isso ajuda a reduzir viés e garantir previsões justas para diferentes grupos demográficos.

Questão 36

Uma empresa deseja criar uma nova solução usando o AWS Glue. A empresa tem experiência mínima em programação com o AWS Glue.

Qual serviço da AWS pode ajudar a empresa a usar o AWS Glue?

- Amazon Q Developer
- Amazon Comprehend
- Amazon Personalize
- AWS Config

Resposta correta:

Amazon Q Developer.

- *O Amazon Q Developer pode fornecer suporte e sugestões para trabalhar com o AWS Glue, facilitando o uso do serviço.*

Questão 37

Uma empresa deseja usar um modelo de linguagem grande (LLM) para gerar descrições concisas e específicas de recursos para os produtos da empresa.

Qual técnica de engenharia de prompt atende a esses requisitos?

- Fornecer prompts detalhados e específicos para cada produto, garantindo descrições precisas e personalizadas.
- Incluir uma variedade de recursos de produto em cada prompt:
- Criar um único prompt que abranja todos os produtos:
- Criar prompts para cada categoria de produto:

Resposta correta:

Fornecer prompts detalhados e específicos para cada produto, garantindo descrições precisas e personalizadas.

- Prompts bem definidos aumentam a precisão das respostas do LLM, resultando em descrições mais relevantes.

Questão 38

Qual estratégia determinará se um modelo fundacional (FM) atende efetivamente aos objetivos de negócios?

- Analisar a arquitetura e os hiperparâmetros do modelo.
- Avaliar o alinhamento do modelo com casos de uso específicos
- Medir os recursos computacionais necessários para a implantação do modelo.
- Avaliar o desempenho do modelo em conjuntos de dados de benchmark.

Resposta correta:

Avaliar o alinhamento do modelo com casos de uso específicos.

- *O alinhamento com os casos de uso garante que o modelo seja eficaz para os objetivos da empresa.*



Questão 39

Uma empresa está lançando um aplicativo móvel que ajuda os usuários a aprender idiomas estrangeiros. O aplicativo torna o texto mais coerente chamando um modelo de linguagem de grande porte (LLM). A empresa coletou um conjunto diversificado de textos e suplementou o conjunto com exemplos de versões mais legíveis. A empresa quer que a saída do LLM se assemelhe aos exemplos fornecidos.

Qual métrica a empresa deve usar para avaliar se o LLM atende a esses requisitos?

- Robustez semântica:

- Métrica Recall-Oriented Understudy for Gisting Evaluation (ROUGE)
- Latência da geração de texto
- Valor da função de perda

Resposta correta:

Métrica Recall-Oriented Understudy for Gisting Evaluation (ROUGE).

- *O ROUGE mede a similaridade entre o texto gerado pelo LLM e os exemplos esperados, avaliando a qualidade da saída.*
-

Questão 40

Uma empresa deseja criar um aplicativo de priorização de leads para que seus funcionários entrem em contato com potenciais clientes. O aplicativo deve permitir que os funcionários visualizem e ajustem os pesos atribuídos a diferentes variáveis no modelo com base no conhecimento e na experiência do domínio.

Qual tipo de modelo de ML atende a esses requisitos?

- Modelo k-vizinhos mais próximos (k-NN)
- Modelo de deep learning baseado em componentes principais
- Modelo de regressão logística
- Rede neural

Resposta correta:

Modelo de regressão logística.

- *A regressão logística é adequada para problemas de classificação binária, onde as variáveis de entrada podem ser ajustadas com base no conhecimento do domínio para prever a prioridade dos leads.*

Questão 41

Uma empresa está criando um chatbot para melhorar a experiência do usuário. A empresa está usando um modelo de linguagem grande (LLM) da Amazon Bedrock para detecção de intenção. A empresa deseja usar aprendizado de few-shot para melhorar a precisão da detecção de intenção.

Quais dados adicionais a empresa precisa para atender a esses requisitos?

- Pares de respostas do chatbot e intenções corretas do usuário
- Pares de mensagens do usuário e respostas corretas do chatbot:
- Pares de intenções do usuário e respostas corretas do chatbot:
- Pares de mensagens do usuário e intenções corretas do usuário:

Resposta correta:

Pares de mensagens do usuário e intenções corretas do usuário.

- *Esses dados são essenciais para treinar o modelo com poucos exemplos, permitindo que ele identifique corretamente as intenções de novos usuários.*

Questão 42

Uma empresa está usando o Amazon SageMaker Studio notebooks para construir e treinar modelos de Machine Learning (ML). A empresa armazena os dados em um bucket Amazon S3. A empresa precisa gerenciar o fluxo de dados do Amazon S3 para os notebooks do SageMaker Studio.

Qual solução atenderá a esse requisito?

- Use o Amazon Inspector para monitorar o SageMaker Studio;
- Use o Amazon Macie para monitorar o SageMaker Studio;
- Configure o SageMaker para usar o S3 Glacier Deep Archive
- Configure o SageMaker para usar uma VPC com um endpoint do S3

Resposta correta:

Configure o SageMaker para usar uma VPC com um endpoint do S3.

- *Essa configuração permite a comunicação segura entre o SageMaker e o S3, sem sair da rede privada, garantindo maior desempenho e segurança.*

Questão 43

Um praticante de IA está usando um modelo base do Amazon Bedrock para resumir chats de sessão do departamento de atendimento ao cliente. O praticante deseja

armazenar logs de invocação para monitorar os dados de entrada e saída do modelo.

Qual estratégia o praticante de IA deve usar?

- Configurar o registro de logs de invocação no Amazon EventBridge;
- Habilitar o registro de logs de invocação no Amazon Bedrock (Enable invocation logging in Amazon Bedrock);
- Configurar o AWS Audit Manager como destino dos logs para o modelo;
- Configurar o AWS CloudTrail como destino dos logs para o modelo;

Resposta correta:

Habilitar o registro de logs de invocação no Amazon Bedrock (Enable invocation logging in Amazon Bedrock).

- *Com essa configuração, você pode armazenar os logs das interações do modelo, facilitando a análise e monitoramento do desempenho.*

Questão 44

Qual é um benefício do pré-treinamento contínuo ao ajustar um modelo de fundação (FM)?

- Ajuda a reduzir a complexidade do modelo
- Otimiza o tempo de inferência do modelo
- Melhora o desempenho do modelo ao longo do tempo;
- Diminui o tempo necessário para o treinamento

Resposta correta:

Melhora o desempenho do modelo ao longo do tempo.

- *O pré-treinamento contínuo permite que o modelo se atualize com novos dados, adaptando-se melhor ao longo do tempo e aprimorando suas respostas.*

Questão 45

Uma empresa quer usar um modelo de linguagem grande (LLM) no Amazon Bedrock para análise de sentimentos. A empresa precisa que o LLM produza respostas mais consistentes para o mesmo prompt de entrada.

Qual ajuste em um parâmetro de inferência a empresa deve fazer para atender a esses requisitos?

- Aumentar o valor do parâmetro de temperatura:
- Diminuir o valor do parâmetro de temperatura.
- Reduzir o comprimento dos tokens de saída:
- Aumentar o comprimento máximo da geração:

Resposta correta:

Diminuir o valor do parâmetro de temperatura.

- *Diminuir a temperatura torna as respostas mais determinísticas, reduzindo a aleatoriedade e garantindo mais consistência.*

Questão 46

Uma empresa tem um modelo fundacional (FM) que foi customizado usando o Amazon Bedrock para responder a consultas de clientes sobre produtos. A empresa deseja validar as respostas do modelo para novos tipos de consultas. A empresa precisa carregar um novo dataset que o Amazon Bedrock possa usar para validação.

Qual serviço AWS atende a esses requisitos?

- AWS Snowcone:
- Amazon Elastic Block Store (Amazon EBS)
- Amazon S3
- Amazon Elastic File System (Amazon EFS):

Resposta correta:

Amazon S3.

- *O S3 é o serviço recomendado para armazenar e gerenciar datasets, permitindo que o Amazon Bedrock acesse facilmente os dados para validação do modelo.*

Questão 47

aprendizado de máquina (ML) para analisar dados arquivados. A empresa deve realizar inferências em grandes conjuntos de dados que possuem múltiplos GB de tamanho. A empresa não precisa acessar imediatamente as previsões do modelo.

Qual opção de inferência do Amazon SageMaker atenderá a esses requisitos?

- Transformação em lote (Batch transform)
- Inferência assíncrona (Asynchronous inference):
- Inferência sem servidor (Serverless inference):
- Inferência em tempo real (Real-time inference):

Resposta correta:

Transformação em lote (Batch transform).

- *O Batch Transform permite realizar inferências em grandes volumes de dados de forma assíncrona, ideal para dados arquivados e sem necessidade de resposta imediata.*

Questão 48

Uma empresa revisa manualmente todos os currículos enviados em formato PDF. À medida que a empresa cresce, espera-se que o volume de currículos exceda a capacidade de revisão da empresa. A empresa precisa de um sistema automatizado para converter os currículos em PDF para o formato de texto simples para processamento adicional.

Qual serviço da AWS atende a esse requisito?

- Amazon Lex:
- Amazon Personalize:
- Amazon Textract
- Amazon Transcribe:

Resposta correta:

Amazon Textract.

- O Textract é especializado na extração de texto de documentos, como PDFs de currículos, e converte-os em um formato utilizável para processamento posterior.

Questão 49

Uma empresa quer avaliar os custos associados ao uso de um modelo de linguagem grande (LLM) para gerar inferências. A empresa deseja usar o Amazon Bedrock para criar aplicações de IA generativa.

Qual fator irá influenciar os custos de inferência?

- Tempo total de treinamento
- Valor da temperatura
- Quantidade de dados usados para treinar o LLM:
- Número de tokens consumidos

Resposta correta:

Número de tokens consumidos.

- *O custo de inferência é calculado com base na quantidade de tokens que o modelo consome ao gerar a resposta, sendo diretamente impactado pelo tamanho da entrada e da saída.*

Questão 50

Um grande varejista recebe milhares de consultas de suporte ao cliente sobre produtos todos os dias. As consultas precisam ser processadas e respondidas rapidamente. A empresa deseja implementar Agents for Amazon Bedrock.

- Geração de modelos fundamentais personalizados (FMs) para prever as necessidades dos clientes.
- Seleção do modelo fundamental (FM) com base em critérios e métricas predefinidos.
- Automação de tarefas repetitivas e orquestração de fluxos de trabalho complexos.
- Chamadas automáticas a múltiplos modelos fundamentais (FMs) e consolidação dos resultados.

Resposta correta:

Automação de tarefas repetitivas e orquestração de fluxos de trabalho complexos.

- Os Agents ajudam a automatizar e orquestrar processos que exigem a interação de múltiplos modelos ou tarefas repetitivas, aumentando a eficiência.

Questão 51

Uma empresa está utilizando few-shot prompting em um modelo base hospedado no Amazon Bedrock. O modelo atualmente usa 10 exemplos no prompt. O modelo é invocado uma vez ao dia e está apresentando bom desempenho. A empresa deseja reduzir o custo mensal.

- Aumentar o número de tokens no prompt.
- Personalizar o modelo usando fine-tuning:
- Usar Throughput Provisionado
- Reduzir o número de tokens no prompt.

Resposta correta:

Reducir o número de tokens no prompt.

- Reduzir a quantidade de tokens utilizados no prompt diminui o custo de inferência, mantendo a eficácia do modelo sem aumentar os gastos.

Questão 52

Um praticante de IA está utilizando um modelo de linguagem grande (LLM) para criar conteúdo para campanhas de marketing. O conteúdo gerado parece plausível e factual, mas está incorreto.

- Vazamento de dados:
- Underfitting:
- Overfitting:
- Alucinação.

Resposta correta:

Alucinação.

- Alucinação ocorre quando o modelo gera informações plausíveis, mas incorretas ou inexistentes, sem base em dados reais.

Questão 53

O que são tokens no contexto de modelos de IA generativa?

- Tokens são os prompts específicos ou instruções fornecidos a um modelo de IA generativa para gerar saídas;
- Tokens são as unidades básicas de entrada e saída com as quais um modelo de IA generativa opera, representando palavras, subpalavras ou outras unidades linguísticas;
- Tokens são os pesos pré-treinados de um modelo de IA generativa ajustados para tarefas específicas;
- Tokens são as representações matemáticas de palavras ou conceitos usados em modelos de IA generativa;

Resposta correta:

Tokens são as unidades básicas de entrada e saída com as quais um modelo de IA generativa opera, representando palavras, subpalavras ou outras unidades linguísticas.

- *Tokens são as partes do texto (palavras ou subpalavras) que o modelo utiliza para processar e gerar respostas.*

Questão 54

Uma empresa deseja exibir as vendas totais de seus produtos mais vendidos em várias localidades de varejo nos últimos 12 meses.

Qual solução da AWS a empresa deve usar para automatizar a geração de gráficos?

- Amazon Q Developer
- Amazon Q no AWS Chatbot
- Amazon Q no Amazon EC2

- Amazon Q no Amazon QuickSight

Resposta correta:

Amazon Q no Amazon QuickSight.

- *O Amazon QuickSight oferece uma plataforma de BI para gerar e visualizar gráficos e dashboards automaticamente, sendo a melhor opção para exibição de vendas e análise de dados.*

Questão 55

Qual termo descreve as representações numéricas de objetos e conceitos do mundo real que modelos de processamento de linguagem natural (NLP) usam para melhorar a compreensão de informações textuais?

- Modelos:
- Embeddings
- Binários:
- Tokens:

Resposta correta:

Embeddings.

- *Embeddings são representações vetoriais numéricas de palavras ou frases que os modelos usam para entender melhor as relações semânticas no texto.*

Questão 56

Um estudante em uma universidade está copiando conteúdo de uma IA generativa para escrever redações.

Qual desafio de uso responsável de IA generativa esse cenário representa?

- Plagiarism (Plágio)
- Privacy (Privacidade)
- Hallucinations (Alucinações).
- Toxicity (Toxicidade).

Resposta correta:
Plagiarism (Plágio).

- *O estudante está copiando conteúdo gerado pela IA, o que configura plágio, um problema ético e legal em ambientes acadêmicos.*

Questão 57

Uma empresa precisa construir seu próprio modelo de linguagem de grande porte (LLM) com base apenas nos dados privados da empresa. A empresa está preocupada com o impacto ambiental do processo de treinamento.

Qual tipo de instância Amazon EC2 tem o **MENOR** impacto ambiental ao treinar LLMs?

- Amazon EC2 P series
- Amazon EC2 Tm series
- Amazon EC2 G series
- Amazon EC2 C series

Resposta correta:
Amazon EC2 Tm series.

- *A série Tm é otimizada para treinamento de modelos de IA, com eficiência energética superior, resultando em menor impacto ambiental.*

Questão 58

Como as empresas podem usar modelos de linguagem grande (LLMs) com segurança no Amazon Bedrock?

- Projetar prompts claros e específicos. Configurar funções e políticas do AWS Identity and Access Management (IAM) usando o princípio de menor privilégio.
- Usar o Amazon CloudWatch Logs para tornar os modelos explicáveis e monitorar vieses

- Habilitar trabalhos automáticos de avaliação de modelos no Amazon Bedrock
- Habilitar o AI Audit Manager para track trabalhos automáticos de avaliação

Resposta correta:

Projetar prompts claros e específicos. Configurar funções e políticas do AWS Identity and Access Management (IAM) usando o princípio de menor privilégio.

- *Essas práticas garantem que o uso dos LLMs seja seguro, com controle de acesso adequado e prevenção de uso indevido.*

Questão 59

Uma empresa deseja implantar um chatbot conversacional para responder as perguntas dos clientes. O chatbot é baseado em um modelo ajustado do Amazon SageMaker JumpStart. A aplicação deve cumprir vários frameworks regulatórios.

Quais capacidades a empresa pode demonstrar conformidade? (Escolha duas.)

- Microsserviços fracamente acoplados (Loosely coupled microservices):
- Proteção de dados (Data protection):
- Otimização de custos (Cost optimization):
- Detecção de ameaças (Threat detection):
- Pontos de inferência com escalabilidade automática (Auto scaling inference endpoints):

Resposta correta:

Proteção de dados (Data protection).

Detecção de ameaças (Threat detection).

- *Essas capacidades são essenciais para garantir a conformidade regulatória e a segurança de dados no desenvolvimento de modelos e aplicações de IA.*

Questão 60

Uma empresa deseja criar uma aplicação interativa para crianças que gera novas histórias baseadas em histórias clássicas. A empresa quer usar o Amazon Bedrock e precisa garantir que os resultados e tópicos sejam apropriados para crianças.

Qual serviço ou recurso da AWS atenderá a esses requisitos?

- Amazon Bedrock playgrounds
- Guardrails for Amazon Bedrock
- Agents for Amazon Bedrock
- Amazon Rekognition.

Resposta correta:

Guardrails for Amazon Bedrock.

- *Os Guardrails para Amazon Bedrock permitem configurar regras e filtros que garantem a segurança e adequação dos resultados gerados para públicos específicos, como crianças.*

Questão 61

Um praticante de IA está desenvolvendo um modelo para gerar imagens de humanos em várias profissões. O praticante de IA descobriu que os dados de entrada são enviesados e que atributos específicos afetam a geração de imagens, criando vieses no modelo.

Qual técnica resolverá o problema?

- Aumento de dados para classes desbalanceadas (Data augmentation for imbalanced classes)
- Monitoramento do modelo para distribuição de classes (Model monitoring for class distribution)
- Geração aumentada por recuperação
- Detecção de marca d'água para imagens (Watermark detection for images)

Resposta correta:

Aumento de dados para classes desbalanceadas (Data augmentation for imbalanced classes).

- *O aumento de dados ajuda a equilibrar o modelo, criando mais exemplos para classes menos representadas, evitando viéses indesejados na geração de imagens.*

Questão 62

Um praticante de IA está desenvolvendo um modelo para gerar imagens de humanos em várias profissões. O praticante de IA descobriu que os dados de entrada são enviesados e que atributos específicos afetam a geração de imagens, criando vieses no modelo.

Qual técnica resolverá o problema?

- Aumento de dados para classes desbalanceadas (Data augmentation for imbalanced classes)
- Monitoramento do modelo para distribuição de classes (Model monitoring for class distribution)
- Geração aumentada por recuperação
- Detecção de marca d'água para imagens (Watermark detection for images)

Resposta correta:

Aumento de dados para classes desbalanceadas (Data augmentation for imbalanced classes).

- O aumento de dados ajuda a equilibrar o modelo, criando mais exemplos para classes menos representadas, evitando vieses indesejados na geração de imagens.

VOCABULÁRIO / CONCEITOS DE AI PRACTITIONER

Alucinação: Refere-se ao fenômeno em que o modelo gera informações que parecem plausíveis, mas são incorretas, irrelevantes ou totalmente inventadas.

Inferência: Quando um modelo gera uma saída a partir de uma entrada.
Ou seja, quando ele faz o que ele foi projetado para fazer.

Tokens: Representam palavras inteiras, partes de palavras ou até caracteres, dependendo do modelo e do método de tokenização utilizado. Por exemplo, a palavra "cachorro" pode ser um token único em uma abordagem, mas em outra pode ser dividida em "cacho" e "rro".

Janela de contexto: Limite de tokens que um modelo de IA consegue processar de uma vez (prompt + resposta).

Modelo multimodal: Um modelo de IA multimodal é capaz de processar e integrar informações provenientes de múltiplas modalidades, como texto e imagem.
Ele pode ser multimodal na:
- Incorporação (prompt)
- Geração (resposta do modelo)

Pré-treinamento contínuo: É o processo de treinar um modelo de forma incremental, adicionando novas informações ao longo do tempo, para que ele possa se adaptar a mudanças e aprender continuamente.

Épocas: Representa uma passagem completa por todo o conjunto de dados de treinamento pelo modelo. Durante cada época, o modelo processa todos os dados de entrada e ajusta seus parâmetros para minimizar o erro em suas previsões. Treinar por muitas épocas pode levar ao sobreajuste (quando o modelo aprende muito bem os detalhes dos dados de treinamento, mas não generaliza bem para novos dados).

Aprendizagem supervisionada: A máquina aprende com exemplos rotulados. Isso significa que o algoritmo recebe dados de entrada e as respostas corretas (saídas esperadas), e aprende a fazer previsões.

Aprendizagem não supervisionada: A máquina aprende sem rótulos. Ela tenta descobrir padrões ou estruturas nos dados por conta própria.

Aprendizagem por reforço: A máquina aprende por tentativa e erro, recebendo recompensas ou penalidades com base nas suas ações.

Acurácia: Métrica usada para avaliar o desempenho de modelos de classificação em aprendizado de máquina. Ela mede a proporção de previsões corretas do modelo em relação ao total de previsões.

Tempo médio de resposta: Mede o tempo que o modelo leva, em média, para fornecer uma resposta após receber uma solicitação. Em IA, o tempo médio de resposta é uma métrica de eficiência fundamental para avaliar o desempenho de um modelo em produção.

Conjuntos de dados de benchmark: Dados que servem como uma referência padronizada para medir a eficácia de um modelo. Se a resposta do modelo em teste for igual ou semelhante aos dados de benchmark, ele está funcionando bem.

Exemplo: Forneço a imagem de uma onça retirada de um conjunto de dados de benchmark e peço para um modelo de classificação de animais me dizer qual é o animal. Caso a resposta seja 'Onça', o modelo está desempenhando bem.

Base de conhecimento: Em vez de depender apenas do conhecimento geral do modelo, uma base de conhecimento fornece dados adicionais que podem ser usados para adaptar o modelo a uma tarefa específica, como responder perguntas sobre os produtos de uma empresa.

Guardrails: Aplica um filtro na geração dos modelos, evitando a criação de respostas inadequadas, ofensivas ou não seguras.

Registro de invocação: Garante o monitoramento das interações realizadas com o modelo, incluindo os dados de entrada (prompt) e saída (resposta) durante cada invocação.

Precificação de tokens: Quanto mais tokens são consumidos durante uma inferência (ou seja, no processamento de uma entrada e geração de uma resposta), maior será o custo associado.

Temperatura: Parâmetro ajustável que controla o nível de aleatoriedade nas respostas do modelo. Quando a temperatura é mais baixa, o modelo escolhe as respostas mais consistentes. Com uma temperatura mais alta, o modelo gera respostas mais criativas.

RAG (Retrieval-Augmented-Generation):

Técnica utilizada para enriquecer as respostas (generation) dos modelos do Amazon Bedrock com informações específicas (augmented) retiradas de fontes personalizadas (retrieval), como PDFs armazenados no S3.

Essa técnica faz sentido quando a base de conhecimento do modelo não é suficiente para responder uma questão específica, como dúvidas sobre a política organizacional de uma empresa.

Tipos de inferência do Amazon SageMaker:

Em Tempo Real: O modelo é implantado em um endpoint, que está sempre ativo e pronto para responder a solicitações em tempo real. Quando o sistema recebe uma nova solicitação de predição, o modelo responde quase instantaneamente.

Serverless: Permite que o SageMaker aloque e ajuste automaticamente os recursos, sem a necessidade de provisionar ou gerenciar servidores. Você paga apenas pelo tempo e recursos usados durante a execução da inferência.

Assíncrona: Permite enviar a solicitação e obter a resposta depois, ideal para casos onde não é necessária uma resposta instantânea. A solicitação é colocada em uma fila, e as respostas podem ser recuperadas posteriormente, facilitando o processamento em paralelo de várias predições.

Em Lote: Para processamento em larga escala de dados previamente armazenados, ideal para análises periódicas ou em grandes conjuntos de dados.

Detectção de Objetos

Finalidade: Identificar e localizar objetos específicos dentro de imagens ou vídeos.

Essa técnica é amplamente utilizada em visão computacional e combina classificação de imagem e localização de objetos.

Aplicações:

Segurança (identificação de pessoas ou veículos em câmeras de vigilância).

Veículos autônomos (reconhecimento de pedestres, sinais de trânsito, etc.).

Diagnóstico médico (detecção de tumores em imagens radiológicas).

Comércio (identificação de produtos em imagens).

Detectção de Anomalias

Finalidade: Identificar padrões incomuns ou desvios significativos em um conjunto de dados que podem indicar erros, fraudes ou eventos inesperados. Essa técnica analisa comportamentos fora do padrão definido.

Aplicações:

Detecção de fraudes financeiras.

Monitoramento de sistemas industriais (detectar falhas em máquinas).

Segurança cibernética (identificação de atividades maliciosas).

Saúde (detecção de condições médicas incomuns em exames).

Reconhecimento de Entidades Nomeadas

Finalidade: Identificar e classificar entidades específicas dentro de textos, como nomes de pessoas, organizações, datas ou localizações, utilizando técnicas de Processamento de Linguagem Natural (NLP).

Aplicações:

Análise de contratos ou documentos legais (identificar partes envolvidas).

Extração de dados estruturados de textos não estruturados (e.g., e-mails, artigos).

Sistemas de atendimento ao cliente (compreender perguntas e respostas).

Buscadores (melhorar a precisão das buscas).

Preenchimento de Imagem

Finalidade: Preencher automaticamente áreas ausentes ou danificadas de uma imagem, reconstruindo os pixels com base no contexto visual da área ao redor.

Aplicações:

Restauração de fotos antigas ou danificadas.

Remoção de objetos indesejados de imagens (e.g., remoção de marcas d'água).

Produção de conteúdo criativo (edição visual avançada).

Efeitos visuais em filmes e videogames.

Sistema de Reconhecimento de Fala

Finalidade: Converter fala em texto ou interpretar comandos de voz utilizando modelos de IA treinados para processar áudio. Esses sistemas dependem de técnicas como aprendizado profundo e redes neurais.

Aplicações:

- Assistentes virtuais (e.g., Alexa, Google Assistant, Siri).
- Transcrição de entrevistas ou reuniões.
- Controle por voz em dispositivos IoT.
- Acessibilidade (para pessoas com dificuldades motoras ou auditivas).

Amazon Q Developer – Voltado para desenvolvedores, facilita a criação e integração de IA generativa em aplicações personalizadas, com foco em códigos e APIs Amazon Q Business – Direcionado para gestores e equipes de negócios, otimiza processos e decisões empresariais usando IA generativa, automatizando fluxos de trabalho.

Amazon Q Connect – Focado em clientes e parceiros, conecta IA generativa a sistemas corporativos como CRMs e ERPs para melhorar o atendimento e engajamento.

Amazon QuickSight – Usado por analistas de dados e equipes de TI, é uma ferramenta de business intelligence (BI) para criar dashboards interativos e relatórios analíticos.

1. SageMaker Studio → IDE Completa

Ambiente visual para criar, treinar e implantar modelos de ML, tudo em um só lugar.

2. SageMaker Notebook → Notebooks Gerenciados

Notebooks Jupyter pré-configurados para explorar dados e rodar código sem precisar gerenciar servidores.

3. SageMaker Data Wrangler → Limpeza e Preparação de Dados

Facilita a coleta, transformação e análise de dados antes do treino do modelo.

4. SageMaker Feature Store → Armazena Features de ML

Banco de dados para armazenar, compartilhar e reutilizar features entre modelos.

5. SageMaker Autopilot → AutoML (Treinamento Automático)

Cria e ajusta modelos automaticamente, sem necessidade de experiência em ML.

6. SageMaker JumpStart → Modelos Prontos. Oferece modelos pré-treinados e tutoriais para acelerar o desenvolvimento.

7. SageMaker Pipelines → Automação do ML (MLOps)

Permite criar pipelines automatizados para treinar e implantar modelos.

8. SageMaker Training → Treinamento de Modelos

Usa infraestrutura otimizada para treinar modelos em grande escala.

9. SageMaker Inference → Deploy de Modelos

Serve previsões de modelos treinados com opções de tempo real, batch ou assíncrono.

10. SageMaker Clarify → Detecção de Viés e Explicabilidade Identifica viés nos dados e no modelo e explica como as previsões são feitas.

Fev/2025

Material desenvolvido por:



CLOUD KNITTERS
COMMUNITY



<https://chat.whatsapp.com/GfxMzmFxNSMCLT4AbtmM3R>

Colaboradores:

Emerson Mafra - <https://www.linkedin.com/in/emersonmafraam/>

Wagner Lopes - <https://www.linkedin.com/in/wagner-lopes-silva/>