

13/49

Finalizar teste

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

Beta

[Compartilhar feedback](#)

R=AJUSTAR O PROMPT

Modo de prática



20/49

Finalizar teste

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 processar respostas detalhadas

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

Beta

Compartilhar feedback

**R=Experimentar e refinar o prompt até que o FM produza as respostas desejadas**

Modo de prática



2/49

Finalizar teste

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 multilíngue

☐ Construir um sistema automático de reconhecimento de entidades nomeadas:

**R=Desenvolver um chatbot de sumarização**

Modo de prática

12/49

Finalizar teste

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

Beta

Compartilhar feedback

R=AWS CloudTrail

Modo de prática

48/49

Finalizar teste

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.

Obs: Escreva exatamente conforme a imagem, respeitando Maiúsculas e Minúsculas.

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...

Select...

Deploy model

Develop model

Monitor model

Define business goal and frame ML problem


Beta

Compartilhar feedback

R:Define business goal and frame ML problem,

**Develop model,  
Deploy model,  
Monitor model**

Modo de prática ⓘ

47/49  Finalizar teste

INTELLIGENCE

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
- ☒ Extrair informações das gravações de chamadas usando o Amazon SageMaker Model Monitor
- ☐ Criar rótulos de classificação usando o Amazon Comprehend

Beta [Compartilhar feedback](#)

**R=Transcrever gravações de chamadas usando o Amazon Transcribe**

46/49

Finalizar teste

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 Beta [de terceiros](#) Complete feedback

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?**

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

Modo de prática



31/49

Finalizar teste

🔖 Pergunta 31:

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

## R=AWS Artifact

38/49

Finalizar teste

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 recursos e encerrar servidores

Beta

[Compartilhar feedback](#)

R=Criar trechos de software, rastreamento de referências e rastreamento de licenças de código aberto

39/49

Finalizar teste

Pergunta 39:

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 reforçoBeta [Compartilhar feedback](#)

**R=Aprendizado não supervisionado (Unsupervised learning)**

Modo de prática ⓘ

35/49

Finalizar teste

Pergunta 35:

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 CloudFrontBeta [Compartilhar feedback](#)

**R=AWS PrivateLink**



Modo de prática ⓘ

27/49

Finalizar teste

🔖 Pergunta 27:

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)

**R=Tempo médio de resposta (Average response time)**

36/49

Finalizar teste

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.

Beta [Compartilhar feedback](#)

**R=Adicionar mensagens ao prompt do modelo.**



Pergunta 26:

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 relacionadas.

**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)

**R=Usar aprendizado por transferência (Transfer learning)**

Pergunta 33:

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

**R=Modelo de incorporação multimodal (Multi-modal embedding model)**

34/49

Finalizar teste

Pergunta 34:

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

**R=Amazon SageMaker Feature Store**

25/49

Finalizar teste

Pergunta 25:

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

**R=Gerenciamento escalável de índices e capacidade de busca por vizinhos mais próximos**

18/49

Finalizar teste

Pergunta 18:

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

R=Desenvolver um sistema de detecção de anomalias

13/30

Finalizar teste

Pergunta 13:

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.

R=Conjuntos de dados de benchmark. (referência)

7/30

Finalizar teste

🔖 Pergunta 7:

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)

**R=Inferência (Inference)**

17/30

Finalizar teste

🔖 Pergunta 17:

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)

**R=Sob demanda (On-Demand)**

10/30

Finalizar teste

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 conformidade usando o Amazon Inspector
- ☐ Criptografar e proteger os dados de treinamento usando o Amazon Macie

**R=Gerar métricas simples, relatórios e exemplos usando o Amazon SageMaker Clarify**

1/30

Finalizar teste

interações passadas e recursos online.

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

**Alternativas traduzidas:**

- ☐ Aprendizado por reforço com recompensas para 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:

**R=Aprendizado por reforço com recompensas para feedback positivo dos clientes**

11/30

Finalizar teste

Pergunta 11:

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

R= Amazon Comprehend  
Amazon Bedrock

20/30

Finalizar teste

Pergunta 20:

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

R=Rede adversária generativa (GAN).



🔖 Pergunta 15:

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:

**R=Implementar APIs de moderação (Implement moderation APIs).**

2/30

Finalizar teste

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 do 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

Beta

[Compartilhar feedback](#)

**R=Importe os dados para o Amazon SageMaker Canvas. Construa modelos de ML e previsões de demanda selecionando os valores nos dados do SageMaker Canvas.**



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

Beta [Compartilhar feedback](#)

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.)**

R=Avaliar o comportamento do modelo para que a empresa possa fornecer transparência aos stakeholders.

Detectar desequilíbrios ou disparidades nos dados

6/30

Finalizar teste

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.

Beta [Compartilhar feedback](#)

☐ Selecionar um conjunto de dados

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

21/30

Finalizar teste

🔖 Pergunta 21:

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):

**R=Aumentar o número de épocas (epochs)**

22/30

Finalizar teste

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 estão 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)

Beta

[Compartilhar feedback](#)

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 estão enviesados e que atributos específicos afetam a geração de imagens, criando vieses no modelo.

Qual técnica resolverá o problema?

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

19/30

Finalizar teste

🔖 Pergunta 19:

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.

Beta

[Compartilhar feedback](#)

## R=Guardrails for Amazon Bedrock

23/30

Finalizar teste

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):**

Beta

[Compartilhar feedback](#)

Uma empresa deseja implantar um chatbot conversacional para responder às 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.)**

R=Proteção de dados (Data protection):

Detecção de ameaças (Threat detection):

28/30

Finalizar teste

Pergunta 28:

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 Beta **AI Model Manager** para trabalhos automáticos de avaliação

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

27/30

Finalizar teste

🔖 Pergunta 27:

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 Trn series

☐ Amazon EC2 G series

☐ Amazon EC2 C series

R= Amazon EC2 Trn series

26/30

Finalizar teste

🔖 Pergunta 26:

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).

R=Plagiarism (Plágio)





## Simulado 3

Modo de prática



2/30

Finalizar teste

Pergunta 2:

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:

R=Embeddings

4/30

Finalizar teste

Pergunta 4:

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

R= Amazon Q no Amazon QuickSight

🔖 Pergunta 6:

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:

**R=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:**

7/30

[Finalizar teste](#)

🔖 Pergunta 7:

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.

**R=Alucinação**

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.

Beta [Compartilhar feedback](#)

**R=Automação de tarefas repetitivas e orquestração de fluxos de trabalho complexos.**

🔖 Pergunta 10:

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.

Beta [Compartilhar feedback](#)

**R=Reduzir o número de tokens no prompt.**

🔖 Pergunta 11:

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

**R=Número de tokens consumidos**

12/30

Finalizar teste

🔖 Pergunta 12:

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:

**R=Amazon Textract**

15/30

Finalizar teste

Uma empresa está desenvolvendo um modelo de 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):

**R=Transformação em lote (Batch transform)**

18/30

Finalizar teste

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):

**R=Amazon S3**

Uma empresa está desenvolvendo um modelo de 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?**

Modo de prática ⓘ

19/30

Finalizar teste

🔖 Pergunta 19:

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

**R=Melhora o desempenho do modelo ao longo do tempo:**