

Generative AI Leader

Guia do exame de certificação

Um Generative AI Leader com Certificação em Google Cloud é um profissional visionário com amplo conhecimento sobre como a IA generativa pode transformar uma empresa e ser usada dentro dela. Ele tem conhecimento em nível empresarial de produtos e serviços de IA generativa do Google Cloud e entende como a abordagem do Google, centrada em IA, junto com as soluções e os produtos modernos, pode ajudar sua organização a adotar a IA de maneira inovadora e responsável. Ele pode participar de conversas significativas tanto com equipes técnicas quanto não técnicas para promover a colaboração e influenciar iniciativas voltadas à tecnologia de IA generativa. Também é capaz de identificar potenciais casos de uso para a IA generativa em diversas funções de negócios e setores, aplicando o próprio conhecimento das soluções empresariais do Google Cloud para impulsionar a inovação. Tem experiência com liderança estratégica e influência, mas não com implementação técnica, embora entenda de forma geral os conceitos e as tecnologias da IA generativa.

Seção 1: conceitos básicos da IA generativa (cerca de 30% do exame)

1.1 Descrever os principais conceitos e casos de uso da IA generativa. As considerações incluem:

- Definir os principais conceitos da IA generativa (por exemplo, inteligência artificial, Aprocessamento de linguagem natural, machine learning, IA generativa, modelos de fundação, modelos de fundação multimodais, modelos de difusão, ajuste de comandos, engenharia de comando, modelos de linguagem grande).
- Descrever as abordagens de machine learning (por exemplo, supervisionada, não supervisionada, por reforço).
- Identificar as etapas do ciclo de vida de machine learning (ingestão de dados, preparação de dados, treinamento de modelo, implantação de modelo e gerenciamento de modelos) e as ferramentas do Google Cloud para cada etapa.
- Identificar como escolher o modelo de fundação adequado para um caso de uso empresarial (por exemplo, modalidade, janela de contexto, segurança, disponibilidade e confiabilidade, custo, desempenho, ajuste fino e personalização).
- Identificar casos de uso empresariais em que a IA generativa pode criar, resumir, descobrir e automatizar (por exemplo, geração de texto, geração de imagens, geração de código, geração de vídeo, análise de dados e experiência do usuário personalizada).
- Descrever como vários tipos de dados são usados na IA generativa e as implicações deles para os negócios.

- Explicar as características e a importância da qualidade e da acessibilidade dos dados para a IA (por exemplo, integridade, consistência, relevância, disponibilidade, custo, formato).
- Identificar as diferenças entre dados estruturados e não estruturado, além de exemplos reais de cada tipo.
- Identificar as diferenças entre dados rotulados e não rotulados.

1.2 Descrever como os diversos tipos de dados são usados na IA generativa e as implicações deles para os negócios. As considerações incluem:

- Explicar as características e a importância da qualidade e da acessibilidade dos dados para a IA (por exemplo, integridade, consistência, relevância, disponibilidade, custo, formato).
- Identificar as diferenças entre dados estruturados e não estruturado, além de exemplos reais de cada tipo.
- Identificar as diferenças entre dados rotulados e não rotulados.

1.3 Identificar as camadas principais do cenário da IA generativa e as implicações delas para os negócios. As considerações incluem:

- Infraestrutura
- Modelos
- Plataformas
- Agentes
- Aplicativos

1.4 Identificar os casos de uso e os pontos fortes dos modelos de fundação do Google. As considerações incluem:

- Gemini
- Gemma
- Imagen
- Veo

Seção 2: soluções de IA generativa do Google Cloud (cerca de 35% do exame)

2.1 Descrever os pontos fortes do Google Cloud no campo da IA generativa. As considerações incluem:

- Descrever como a abordagem da Google, centrada em IA, e o compromisso dele com a inovação futura se traduzem em soluções modernas de IA generativa.
- Descrever como o Google Cloud tem uma plataforma de IA pronta para empresas (por exemplo, responsável, segura, privada, confiável e escalonável).

- Reconhecer as vantagens do ecossistema de IA abrangente do Google (por exemplo, a integração da IA generativa em todos os produtos e serviços do Google).
- Descrever os benefícios da abordagem aberta do Google Cloud.
- Identificar os componentes essenciais da infraestrutura otimizada para IA do Google Cloud e os benefícios dela (por exemplo, hipercomputadores, TPUs personalizadas do Google, GPUs, data centers, computação em nuvem).
- Explicar como a plataforma de IA do Google Cloud oferece aos usuários controle sobre os dados próprios (por exemplo, segurança, privacidade, governança, modelos abertos e modelos próprios líderes de mercado, soluções pré-configuradas e personalizáveis, agentes).
- Descrever como a plataforma de IA do Google Cloud democratiza o desenvolvimento da IA (por exemplo, ferramentas de pouco código e sem código, modelos pré-treinados, APIs).

2.2 Descrever como as soluções de IA generativa pré-construídas do Google Cloud possibilitam o trabalho com inteligência artificial. As considerações incluem:

- Reconhecer a funcionalidade, os casos de uso e o valor comercial do app do Gemini e do Gemini Advanced (por exemplo, Gems).
- Reconhecer a funcionalidade, os casos de uso e o valor comercial do Gemini Enterprise (por exemplo, API Gemini Notebook, pesquisa multimodal e habilidades agênticas personalizadas).
- Reconhecer a funcionalidade, os casos de uso e o valor comercial do Gemini para Google Workspace.

2.3 Descrever como as soluções de IA generativa do Google Cloud melhoram a experiência do cliente. As considerações incluem:

- Reconhecer a funcionalidade, os casos de uso e os benefícios comerciais das soluções de pesquisa externa do Google Cloud (por exemplo, Agent Platform para Pesquisa, Pesquisa Google).
- Reconhecer a funcionalidade, os casos de uso e o valor comercial do Pacote de engajamento do cliente do Google (por exemplo, Conversational Agents, Agent Assist, Conversational Insights, Central de atendimento como serviço do Google Cloud).

2.4 Descrever como o Google Cloud capacita os desenvolvedores a criar soluções com IA. As considerações incluem:

- Reconhecer a funcionalidade, os casos de uso e o valor comercial da Agent Platform (por exemplo, Model Garden, Agent Platform para Pesquisa, AutoML).
- Reconhecer a funcionalidade, os casos de uso e o valor comercial das soluções de RAG do Google Cloud (por exemplo, RAG pré-configurada com Agent Platform para Pesquisa, APIs RAG).

- Reconhecer a funcionalidade, os casos de uso e o valor comercial do uso do Agent Platform do Agent Builder para criar agentes personalizados.

2.5 Definir a finalidade e os tipos de ferramentas para agentes de IA generativa. As considerações incluem:

- Identificar como os agentes utilizam ferramentas para interagir com o ambiente externo e realizar tarefas (por exemplo, extensões, funções, repositórios de dados e plug-ins).
- Identificar serviços relevantes do Google Cloud e APIs de IA pré-configuradas para ferramentas de agentes (por exemplo, Cloud Storage, bancos de dados, Cloud Functions, Cloud Run, Agent Platform, API Speech-to-Text, API Text-to-Speech, API Translation, API Document Translation, API Document AI, API Cloud Vision, API Cloud Video Intelligence, API Natural Language, biblioteca de APIs do Google Cloud).
- Determinar quando usar o Agent Studio e o Google AI Studio.

Seção 3: técnicas para melhorar a resposta do modelo de IA generativa (cerca de 20% do exame)

3.1 Descrever como superar de forma proativa as limitações do modelo de fundação. As considerações incluem:

- Identificar as limitações comuns dos modelos de fundação (por exemplo, dependência de dados, limite de conhecimento, viés, imparcialidade, alucinações, casos extremos).
- Descrever as práticas recomendadas pelo Google Cloud para lidar com limitações, por exemplo, embasamento, geração aumentada por recuperação (RAG), engenharia de comando, ajuste fino, human in the loop (HITL).
- Reconhecer as práticas recomendadas pelo Google para monitoramento e avaliação contínuos de modelos de IA generativa (por exemplo, atualizações automáticas de modelos, indicadores principais de desempenho, patches e atualizações de segurança, controle de versões, acompanhamento de performance, monitoramento de degradação, Agent Platform Feature Store).

3.2 Descrever técnicas de engenharia de comando e como elas geram resultados melhores. As considerações incluem:

- Definir a engenharia de comando e descrever a importância dela na interação com modelos de linguagem grandes (LLMs).
- Identificar técnicas de criação de comandos e casos de uso (por exemplo, zero-shot, one-shot, few-shot, criação de comandos baseados em papel, encadeamento de comandos).
- Identificar técnicas avançadas de criação de comandos e quando utilizá-las (por exemplo, comandos com linha de raciocínio, comandos ReAct).

3.3 Identificar técnicas de embasamento e casos de uso. As considerações incluem:

- Descrever o conceito de embasamento em LLMs e diferenciar entre embasamento com dados próprios da empresa, dados de terceiros e dados globais.
- Descrever como a geração aumentada por recuperação (RAG) pode afetar a resposta gerada pelos seus modelos de IA generativa.
- Soluções de embasamento do Google Cloud:
 - a. RAG pré-configurada com Agent Platform para Pesquisa
 - b. APIs RAG
 - c. Embasamento com a Pesquisa Google
- Identificar como os parâmetros e as configurações de amostragem são usados para controlar o comportamento de modelos de IA generativa, por exemplo, contagem de tokens, temperatura, top-p (amostragem de núcleo), configurações de segurança e comprimento da resposta.

Seção 4: estratégias de negócios para uma solução de IA generativa bem-sucedida (cerca de 15% do exame)

4.1 Descrever as etapas recomendadas pelo Google Cloud para implementar uma solução de IA generativa transformadora. As considerações incluem:

- Reconhecer os diferentes tipos de soluções de IA generativa (por exemplo, geração de texto, de imagens e de código, necessidades personalizadas do usuário).
- Identificar os principais fatores que influenciam as necessidades de IA generativa (por exemplo, requisitos comerciais, restrições técnicas).
- Descrever como escolher a solução de IA generativa adequada para uma necessidade comercial específica.
- Identificar as etapas para integrar a IA generativa a uma organização.
- Identificar as técnicas para avaliar o impacto das iniciativas de IA generativa.

4.2 Definir o que torna a IA segura e a importância disso na proteção de sistemas de IA contra ataques mal-intencionados e uso indevido. As considerações incluem:

- Explicar como é a segurança ao longo de todo o ciclo de vida do ML.
- Identificar a finalidade e os benefícios do framework de IA segura (SAIF) do Google.
- Reconhecer as ferramentas de segurança do Google Cloud e a finalidade delas (por exemplo, infraestrutura com segurança incorporada ao design, Identity and Access Management (IAM), Security Command Center e ferramentas de monitoramento de cargas de trabalho).

4.3 Descrever a importância da IA responsável nas empresas. As considerações incluem:

- Explicar a importância da IA responsável e da transparência.
- Descrever as considerações de privacidade (por exemplo, riscos à privacidade, anonimização e pseudonimização de dados).
- Descrever as implicações da qualidade dos dados, do viés e da imparcialidade.
- Descrever a importância da responsabilização e da explicabilidade em sistemas de IA.