Google Cloud

Generative AI Leader

Certification exam guide

A Google Cloud Certified Generative AI Leader is a visionary professional with comprehensive knowledge of how generative AI (gen AI) can transform and be used within a business. This individual has business-level knowledge of Google Cloud's gen AI products and services. They recognize how Google's AI-first approach and cutting-edge products and solutions can lead their organizations toward innovative and responsible AI adoption. They can engage in meaningful conversations with both technical and non-technical teams, fostering collaboration and influencing gen AI-powered initiatives. They can identify potential use cases for gen AI across various business functions and industries, using their knowledge of Google Cloud's enterprise-ready offerings to accelerate innovation. Their expertise is in strategic leadership and influence, not technical implementation, though they have a conceptual understanding of gen AI concepts and technology.

Section 1: Fundamentals of gen AI (~30% of the exam)

1.1 Describe core generative AI (gen AI) concepts and use cases. Considerations include:

- Defining core gen AI concepts (e.g., artificial intelligence, natural language processing, machine learning, generative AI, foundation models, multimodal foundation models, diffusion models, prompt tuning, prompt engineering, large language models).
- Describing the machine learning approaches (e.g., supervised, unsupervised, reinforcement).
- Identifying the stages of the machine learning lifecycle; data ingestion, data preparation, model training, model deployment, and model management; and the Google Cloud tools for each stage.
- Identifying how to choose the appropriate foundation model for a business use case (e.g., modality, context window, security, availability and reliability, cost, performance, fine-tuning, and customization).
- Identifying business use cases where gen AI can create, summarize, discover, and automate (e.g., text generation, image generation, code generation, video generation, data analysis, and personalized user experience).
- Describing how various data types are used in gen AI and the business implications.
- Explaining the characteristics and importance of data quality and data accessibility in AI (e.g., completeness, consistency, relevance, availability, cost, format).
- Identifying the differences between structured and unstructured data, and identifying real-world examples of each type.
- Identifying the differences between labeled and unlabeled data.

1.2 Describe how various data types are used in gen AI and the business implications. Considerations include:

- Explaining the characteristics and importance of data quality and data accessibility in AI (e.g., completeness, consistency, relevance, availability, cost, format).
- Identifying the differences between structured and unstructured data, and identifying real-world examples of each type.
- Identifying the differences between labeled and unlabeled data.

1.3 Identify the core layers of the gen AI landscape and the business implications. Considerations include:

- Infrastructure
- Models
- Platforms
- Agents
- Applications

1.4 Identify the use cases and strengths of Google's foundation models. Considerations include:

- Gemini
- Gemma
- Imagen
- Veo

Section 2: Google Cloud's gen AI offerings (~35% of the exam)

2.1 Describe Google Cloud's strengths in the field of gen AI. Considerations include:

- Describing how Google's AI-first approach and commitment to future innovation translate into cutting-edge gen AI solutions.
- Describing how Google Cloud has an enterprise-ready AI platform (e.g., responsible, secure, private, reliable, scalable).
- Recognizing the advantages of Google's comprehensive AI ecosystem (e.g., integration of gen AI across Google products and services).
- Describing the benefits of Google Cloud's open approach.
- Identifying the essential components of Google Cloud's AI-optimized infrastructure and its benefits (e.g., hypercomputer, Google's custom-designed TPUs, GPUs, data centers, cloud computing).

- Explaining how Google Cloud's AI platform provides users with control over their data (e.g., security, privacy, governance, open and leading first party models, pre-built and customizable solutions, agents).
- Describing how Google Cloud's AI platform democratizes AI development (e.g., low-code and no-code tools, pre-trained models, APIs).

2.2 Describe how Google Cloud's prebuilt gen AI offerings enable AI powered work. Considerations include:

- Recognizing the functionality, use cases, and business value of the Gemini app and Gemini Advanced (e.g., Gems).
- Recognizing the functionality, use cases, and business value of Gemini Enterprise (e.g., Gemini Notebook API, multimodal search, and custom agent capabilities).
- Recognizing the functionality, use cases, and business value of Gemini for Google Workspace.

2.3 Describe how Google Cloud's gen AI offerings improve the customer experience. Considerations include:

- Recognizing the functionality, use cases, and business benefits of Google Cloud's external search offerings (e.g., Agent Search on Gemini Enterprise Agent Platform , Google Search).
- Recognizing the functionality, use cases, and business value of Google's Customer Engagement Suite (e.g., Conversational Agents, Agent Assist, Conversational Insights, Google Cloud Contact Center as a Service).

2.4 Describe how Google Cloud empowers developers to build with AI. Considerations include:

- Recognizing the functionality, use cases, and business value of Agent Platform (e.g., Model Garden, Agent Search, Agent Platform AutoML).
- Recognizing the functionality, use cases, and business value of Google Cloud's RAG offerings (e.g., prebuilt RAG with Agent Search, RAG APIs).
- Recognizing the functionality, use cases, and business value of using Agent Platform to build custom agents.

2.5 Define the purpose and types of tooling for gen AI agents. Considerations include:

- Identifying how agents use tools to interact with the external environment and achieve tasks (e.g., extensions, functions, data stores, and plugins).
- Identifying relevant Google Cloud services and pre-built AI APIs for agent tooling (e.g., Cloud Storage, databases, Cloud Functions, Cloud Run, Agent Platform, 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 API, Google Cloud API Library).
- Determining when to use Agent Studio and Google AI Studio.

Section 3: Techniques to improve gen AI model output (~20% of the exam)

3.1 Describe how to proactively overcome foundation model limitations. Considerations include:

- Identifying common limitations of foundation models (e.g., data dependency, the knowledge cutoff, bias, fairness, hallucinations, edge cases).
- Describing the Google Cloud-recommended practices to address limitations (e.g., grounding, retrieval-augmented generation [RAG], prompt engineering, fine-tuning, human in the loop [HITL]).
- Recognizing Google-recommended practices for continuous monitoring and evaluation of gen AI models (e.g., automatic model upgrades, key performance indicators, security patches and updates, versioning, performance tracking, drift monitoring, Agent Platform Feature Store).

3.2 Describe prompt engineering techniques and how they drive better results. Considerations include:

- Defining prompt engineering and describing its significance in interacting with large language models (LLMs).
- Identifying prompting techniques and use cases (e.g., zero-shot, one-shot, few-shot, role prompting, prompt chaining).
- Identifying advanced prompting techniques and when to use them (e.g., chain-of-thought prompting, ReAct prompting).

3.3 Identify grounding techniques and their use cases. Considerations include:

- Describing the concept of grounding in LLMs and differentiating between grounding with first-party enterprise data, third-party data, and world data.
- Describing how retrieval-augmented generation (RAG) can affect the generated output from your gen AI models.
- Google Cloud grounding offerings:
    a. Pre-built RAG with Agent Search
    b. RAG APIs
    c. Grounding with Google Search
- Identifying how sampling parameters and settings are used to control the behavior of gen AI models (e.g., token count, temperature, top-p [nucleus sampling], safety settings, and output length).

Section 4: Business strategies for a successful gen AI solution (~15% of the exam)

4.1 Describe the Google Cloud-recommended steps to successfully implement a transformational gen AI solution. Considerations include:

- Recognizing the different types of gen AI solutions (e.g., text generation, image generation, code generation, personalized user needs).
- Identifying the key factors that influence gen AI needs (e.g., business requirements, technical constraints).
- Describing how to choose the right gen AI solution for a specific business need.
- Identifying the steps to integrate gen AI into an organization.
- Identifying techniques to measure the impact of gen AI initiatives.

4.2 Define secure AI and its importance in protecting AI systems from malicious attacks and misuse. Considerations include:

- Explaining security throughout the ML lifecycle.
- Identifying the purpose and benefits of Google's Secure AI Framework (SAIF).
- Recognizing Google Cloud security tools and their purpose (e.g., secure-by-design infrastructure, Identity and Access Management (IAM), Security Command Center, and workload monitoring tools).

4.3 Describe the importance of responsible AI in business. Considerations include:

- Explaining the importance of responsible AI and transparency.
- Describing privacy considerations (e.g., privacy risks, data anonymization and pseudonymization).
- Describing the implications of data quality, bias, and fairness.
- Describing the importance of accountability and explainability in AI systems.