Exam

GENERATIVE AI LEADER

Google Cloud Generative AI Leader

Free demo	16 sample questions
Full exam	76 questions with answers and explanations
Contents	Questions with answers and explanations

Exam domains

Domain	Exam weight	Questions
D1 Fundamentals of gen AI		4
D2 Google Cloud's gen AI offerings		8
D3 Techniques to improve gen AI model output		2
D4 Business strategies for a successful gen AI solution		2

This is a free demo from Dumpsgate: 16 of the exam's 76 questions. **Get the full GENERATIVE AI LEADER exam** with every question, answer and explanation.

Question 1

Single choice · Domain: D1 Fundamentals of gen AI › Define an AI agent and distinguish it from static application components, human operators, and user interfaces

What is a definition of an AI agent?

A. A static component of a gen AI application that provides information to the user.
B. A human expert who oversees and manages the operation of a gen AI application.
C. An application that learns how to achieve a goal based on inputs and tools available to it.
D. A type of user interface that allows users to interact with a gen AI application.

Answer: C. An application that learns how to achieve a goal based on inputs and tools available to it.

Explanation

This question tests understanding of the core characteristics of an AI agent: pursuing a goal, using available inputs and tools, and adapting its actions rather than merely displaying information or providing an interface.

✗ **A** A static component of a gen AI application that provides information to the user.
Incorrect. A static information component does not independently learn, reason, or act toward a goal, which are central characteristics of an AI agent.

https://cloud.google.com/discover/what-are-ai-agents

✗ **B** A human expert who oversees and manages the operation of a gen AI application.
Incorrect. A human expert may supervise an AI system, but is not itself an AI agent. An AI agent is a software system that can perform tasks and take actions toward a goal.

https://cloud.google.com/discover/what-are-ai-agents

✓ **C** An application that learns how to achieve a goal based on inputs and tools available to it.
Correct. This describes an AI agent as an application that uses inputs and available tools to determine and execute actions in pursuit of a goal.

https://cloud.google.com/discover/what-are-ai-agents

✗ **D** A type of user interface that allows users to interact with a gen AI application.
Incorrect. A user interface enables interaction with an AI application but is not the agent itself. The agent performs reasoning and actions to accomplish a goal.

https://cloud.google.com/discover/what-are-ai-agents

Free demo · full exam: 76 questions

Question 2

Single choice · Domain: D1 Fundamentals of gen AI › Identify the hardware components that form the infrastructure layer of a generative AI solution

What are core hardware components of the infrastructure layer in the generative AI landscape?

A. TPUs and GPUs
B. User interfaces
C. Pre-trained models
D. Tools and services for building AI models

Answer: A. TPUs and GPUs

Explanation

This question tests understanding of the generative AI technology stack and the role of specialized compute hardware in its infrastructure layer. TPUs and GPUs provide the accelerated processing used to train and run generative AI models.

✓ **A** TPUs and GPUs
Correct. Tensor Processing Units and Graphics Processing Units are specialized hardware accelerators used for generative AI model training and inference, making them core components of the infrastructure layer. and.

https://cloud.google.com/tpu/docs/intro
https://cloud.google.com/compute/docs/gpus

✗ **B** User interfaces
Incorrect. User interfaces are application or experience-layer components that allow people to interact with AI systems; they are not core infrastructure hardware.

https://cloud.google.com/architecture/ai-ml

✗ **C** Pre-trained models
Incorrect. Pre-trained models are software assets used by generative AI applications, not physical infrastructure components.

https://cloud.google.com/vertex-ai/generative-ai/docs/learn/models

✗ **D** Tools and services for building AI models
Incorrect. Tools and services for building AI models belong to the platform or development layer, whereas the infrastructure layer consists of compute, storage, networking, and specialized accelerators such as TPUs and GPUs.

https://cloud.google.com/vertex-ai

Dumpsgate

Question 3

Single choice · Domain: D1 Fundamentals of gen AI › Identify the function of diffusion models in generative AI

What does a diffusion model do?

A. Analyzes data and predicts future trends and patterns.
B. Optimizes business processes and resource allocation.
C. Facilitates the storage and management of structured data.
D. Generates high-quality content by refining noise into structured data.

Answer: D. Generates high-quality content by refining noise into structured data.

Explanation

This question tests understanding of how diffusion models generate content by progressively transforming noise into coherent, structured outputs.

✗ **A** Analyzes data and predicts future trends and patterns.
Incorrect. This describes predictive analytics or forecasting rather than the generative process of a diffusion model.

https://cloud.google.com/vertex-ai/generative-ai/docs/image/overview

✗ **B** Optimizes business processes and resource allocation.
Incorrect. Business-process optimization is an application of AI, not the defining function of a diffusion model.

https://cloud.google.com/vertex-ai/generative-ai/docs/image/overview

✗ **C** Facilitates the storage and management of structured data.
Incorrect. Storage and management of structured data are database functions, not generative-model functions.

https://cloud.google.com/vertex-ai/generative-ai/docs/image/overview

✓ **D** Generates high-quality content by refining noise into structured data.
Correct. Diffusion models learn to remove noise through a denoising process, transforming random noise into structured outputs such as images. Google Cloud describes Imagen as a generative image model available through Vertex AI.

https://cloud.google.com/vertex-ai/generative-ai/docs/image/overview

Question 4

Single choice · Domain: D1 Fundamentals of gen AI › Identify data dependency as a generative AI model limitation

A company trains a generative AI model designed to classify customer feedback as positive, negative, or neutral. However, the training dataset disproportionately includes feedback from a specific demographic and uses outdated language norms that don't reflect current customer communication styles. When the model is deployed, it shows a strong bias in its sentiment analysis for new customer feedback, misclassifying reviews from underrepresented demographics and struggling to understand current slang or phrasing. What type of model limitation is this?

A. Data dependency
B. Edge case
C. Hallucination
D. Overfitting

Answer: A. Data dependency

Explanation

This question tests the ability to recognize that model performance depends on the quality, representativeness, and currency of its training data. The demographic imbalance and outdated language in the dataset cause biased and poorly generalized sentiment classifications.

✓ **A** Data dependency
Correct. The model reflects limitations in its training data: the dataset is demographically unbalanced and uses outdated language, causing biased predictions and difficulty handling current phrasing. Google discusses the importance of data quality and representativeness for responsible AI.
https://cloud.google.com/vertex-ai/generative-ai/docs/learn/responsible-ai

✗ **B** Edge case
Incorrect. An edge case is an unusual or rare input scenario. This problem is broader: systematic demographic underrepresentation and outdated training data affect the model's general performance.
https://cloud.google.com/vertex-ai/generative-ai/docs/learn/responsible-ai

✗ **C** Hallucination
Incorrect. Hallucination occurs when a generative model produces inaccurate or fabricated information, whereas this question describes biased and outdated sentiment classification caused by the training data.
https://cloud.google.com/vertex-ai/generative-ai/docs/learn/responsible-ai

✗ **D** Overfitting
Incorrect. Overfitting occurs when a model learns the training data too specifically and performs poorly on unseen data. Although the model performs poorly on new feedback, the stated limitation is specifically the biased and outdated training data, which is data dependency.

Question 5

Single choice · Domain: D2 Google Cloud's gen AI offerings › Identify the business value of Gemini Enterprise (Agentspace) for enterprise search, document analysis, and task automation

A company is exploring Gemini Enterprise (Agentspace) to improve how its employees search for information on their enterprise systems and automate certain tasks. What is the key business advantage of using Gemini Enterprise (Agentspace)?

A. More granular control over support team access and permissions for sensitive data.
B. Enhanced real-time communication and collaboration among team members.
C. Improved productivity and data interaction using AI assistants and advanced document analysis.
D. Greater interoperability with legacy software systems and databases.

Answer: C. Improved productivity and data interaction using AI assistants and advanced document analysis.

Explanation

This question tests recognition of how Gemini Enterprise helps employees find and interact with enterprise information through AI assistants, while also supporting productivity and task automation.

✗ **A** More granular control over support team access and permissions for sensitive data.
Incorrect. Access controls and data governance are important capabilities, but they are not the primary business advantage described in the question. Gemini Enterprise focuses on helping employees search, analyze, and act on enterprise information.

https://cloud.google.com/gemini-enterprise/docs/overview

✗ **B** Enhanced real-time communication and collaboration among team members.
Incorrect. Gemini Enterprise is not primarily a real-time team communication or collaboration platform. Its central value is AI-powered enterprise search, assistance, and task support.

https://cloud.google.com/gemini-enterprise/docs/overview

✓ **C** Improved productivity and data interaction using AI assistants and advanced document analysis.
Correct. Gemini Enterprise provides AI assistants and enterprise search capabilities that help employees discover information, interact with organizational data, analyze documents, and automate tasks, improving productivity.

https://cloud.google.com/gemini-enterprise/docs/overview

✗ **D** Greater interoperability with legacy software systems and databases.
Incorrect. Gemini Enterprise can connect to enterprise data sources, but interoperability with legacy systems is not the key business advantage highlighted here. The primary advantage is enabling productive interaction with enterprise information through AI assistants and search.

https://cloud.google.com/gemini-enterprise/docs/overview

Free demo · full exam: 76 questions

Question 6

Single choice · Domain: D2 Google Cloud's gen AI offerings › Select the appropriate Google Cloud generative AI offering to automate CRM-based sales research and data-entry workflows

A company's sales team spends a significant amount of time researching potential leads and manually entering data into their customer relationship management (CRM) tool. They want to improve the team's efficiency and enable them to focus on building relationships and closing deals.

What should the organization do?

A. Integrate the CRM with a popular sales intelligence platform to automatically enrich lead profiles with valuable data.

B. Develop a custom AI solution using Google Cloud's AutoML Natural Language to analyze lead communications and automatically update the CRM.

C. Implement Gemini Enterprise "unified enterprise search" including a CRM agent to automate lead research and data entry.

D. Implement Google Cloud's Contact Center AI to qualify leads and route them to the appropriate sales representatives.

Answer: C. Implement Gemini Enterprise "unified enterprise search" including a CRM agent to automate lead research and data entry.

Explanation

This question tests the ability to match a business productivity use case with Gemini Enterprise capabilities, including unified enterprise search and enterprise agents that can work with business applications.

✗ **A** Integrate the CRM with a popular sales intelligence platform to automatically enrich lead profiles with valuable data.
Incorrect. A sales intelligence integration may enrich lead data, but it does not directly address the broader need to use generative AI to automate research and CRM data-entry tasks. Gemini Enterprise is designed to provide enterprise search and agent-based assistance across business information and workflows.

https://cloud.google.com/gemini-enterprise

✗ **B** Develop a custom AI solution using Google Cloud's AutoML Natural Language to analyze lead communications and automatically update the CRM.
Incorrect. Natural-language analysis alone would require substantial custom development for workflow orchestration, CRM integration, and reliable updates. A managed Gemini Enterprise agent is a more appropriate fit for automating enterprise search and business workflows.

https://cloud.google.com/vertex-ai/docs/generative-ai/overview

✓ **C** Implement Gemini Enterprise "unified enterprise search" including a CRM agent to automate lead research and data entry.
Correct. Gemini Enterprise provides unified enterprise search and agent capabilities that can help employees find information across enterprise sources and automate repetitive business tasks, such as researching leads and entering information into a CRM.

https://cloud.google.com/gemini-enterprise

✗ **D** Implement Google Cloud's Contact Center AI to qualify leads and route them to the appropriate sales representatives.
Incorrect. Contact Center AI is focused primarily on customer-service and contact-center interactions, such as assisting agents and routing customer contacts. It is not the best fit for automating sales-team

lead research and CRM data entry.
https://cloud.google.com/contact-center/ccai-platform

Question 7

Single choice · Domain: D2 Google Cloud's gen AI offerings › Select the appropriate Google generative AI tool for analyzing source documents with grounded summaries and citations

A research company needs to analyze several lengthy PDF documents containing financial reports and identify key performance indicators (KPIs) and their trends over the past year. They want a Google Cloud prebuilt generative AI tool that can process these documents and provide summarized insights directly from the source material with citations. What should the analyst do?

A. Create a custom Gem in Gemini Advanced with predefined KPIs to look across different financial reports.
B. Use the Gemini app to ask general financial trend questions.
C. Use NotebookLM to upload and analyze the documents.
D. Use Gemini for Google Workspace within Google Docs to copy and paste sections of the reports for summary and analysis.

Answer: C. Use NotebookLM to upload and analyze the documents.

Explanation

This question tests the ability to distinguish among Google generative AI offerings and select NotebookLM for analyzing uploaded documents, generating grounded insights, and citing the source material.

✗ **A** Create a custom Gem in Gemini Advanced with predefined KPIs to look across different financial reports.
Incorrect. Gems can provide customized instructions and assistance, but this option does not identify the purpose-built document-analysis experience that grounds responses in uploaded sources with citations.

✗ **B** Use the Gemini app to ask general financial trend questions.
Incorrect. General questions in the Gemini app are not the most appropriate choice for analyzing a defined set of lengthy source documents and returning source-grounded insights with citations.

✓ **C** Use NotebookLM to upload and analyze the documents.
Correct. NotebookLM is designed to work with user-provided sources, answer questions about them, summarize their content, and provide citations that link responses back to the source material.

✗ **D** Use Gemini for Google Workspace within Google Docs to copy and paste sections of the reports for summary and analysis.
Incorrect. Gemini in Google Docs can help summarize and work with content in a document, but manually copying sections is less suitable than NotebookLM for analyzing multiple lengthy reports as a source set with citations.

Question 8

Single choice · Domain: D2 Google Cloud's gen AI offerings › Identify the role and performance advantage of Google Cloud TPUs in machine learning workloads

What is a key advantage of using Google's custom-designed TPUs?

A. TPUs are lightweight processors intended for deployment on edge devices.
B. TPUs increase the storage capacity and data retrieval speeds within Google Cloud data centers.
C. TPUs are specialized AI processors that excel at parallel processing for machine learning workloads.
D. TPUs are primarily designed to improve the general processing speed of virtual machines in the cloud.

Answer: C. TPUs are specialized AI processors that excel at parallel processing for machine learning workloads.

Explanation

This question tests recognition that TPUs are Google's specialized processors for accelerating machine learning computations, particularly through highly parallel processing. This is a Google Cloud infrastructure offering relevant to generative AI workloads.

✗ **A** TPUs are lightweight processors intended for deployment on edge devices.
Incorrect. TPUs are purpose-built accelerators used primarily for large-scale machine learning workloads in data centers, not lightweight processors primarily intended for edge-device deployment.
https://cloud.google.com/tpu/docs/intro

✗ **B** TPUs increase the storage capacity and data retrieval speeds within Google Cloud data centers.
Incorrect. TPUs accelerate machine learning computations; they are not storage or data-retrieval components.
https://cloud.google.com/tpu/docs/intro

✓ **C** TPUs are specialized AI processors that excel at parallel processing for machine learning workloads.
Correct. Tensor Processing Units are Google's custom-designed ASICs optimized for machine learning workloads, where parallel matrix and tensor operations are central to training and inference.
https://cloud.google.com/tpu/docs/intro

✗ **D** TPUs are primarily designed to improve the general processing speed of virtual machines in the cloud.
Incorrect. TPUs are specialized accelerators for machine learning rather than general-purpose processors for improving the overall speed of cloud virtual machines.
https://cloud.google.com/tpu/docs/intro

Question 9

Single choice · Domain: D2 Google Cloud's gen AI offerings › Identify Conversational Insights as the Customer Engagement Suite component for analyzing conversational data and extracting customer experience intelligence

An organization wants to understand trends in customer interactions, identify common issues, gauge customer sentiment, and improve the overall customer experience across both their automated chatbot interactions and live agent support. They need a tool that can analyze their existing conversational data to gain actionable business intelligence. What component of Google's Customer Engagement Suite best addresses this need?

A. Google Cloud Contact Center as a Service
B. Agent Assist
C. Conversational Agents
D. Conversational Insights

Answer: D. Conversational Insights

Explanation

This question tests the ability to distinguish among Google Cloud customer engagement products and select the offering designed to analyze conversation data, identify trends and issues, measure sentiment, and generate actionable insights.

✗ **A** Google Cloud Contact Center as a Service
Incorrect. Contact Center as a Service provides the broader contact-center platform and capabilities, but it is not the component specifically focused on analyzing existing conversational data for trends, issues, sentiment, and business insights.

https://cloud.google.com/contact-center/ccai-platform

✗ **B** Agent Assist
Incorrect. Agent Assist supports live agents during customer interactions by providing real-time recommendations and assistance; it is not the primary analytics component for discovering trends across chatbot and agent conversations.

https://cloud.google.com/agent-assist/docs/overview

✗ **C** Conversational Agents
Incorrect. Conversational Agents are used to build and operate automated conversational experiences, such as chatbots and virtual agents. They do not primarily provide the cross-conversation analytics described in the question.

https://cloud.google.com/dialogflow/docs/concept/agent

✓ **D** Conversational Insights
Correct. Conversational Insights analyzes customer conversations to identify topics, trends, recurring issues, sentiment, and opportunities to improve the customer experience across automated and human-assisted interactions.

https://cloud.google.com/contact-center/insights/docs/overview

Question 10

Single choice · Domain: D2 Google Cloud's gen AI offerings › Select an appropriate business use case for Gemma

A finance team wants to use Gemma to help with daily tasks so that the financial analysts can focus on other work. Which business problem can Gemma most efficiently address?

A. The complexity of building and deploying sophisticated internal knowledge bases to answer employees' finance-related questions with accurate and up-to-date information.

B. The difficulty in analyzing large datasets of financial transactions and market data to identify anomalies and predict future financial performance.

C. The struggle to accurately extract key financial figures and insights from a variety of document formats, such as balance sheets and income statements, for quick reporting.

D. The challenge of efficiently producing high-quality written summaries and initial drafts of financial communications.

Answer: D. The challenge of efficiently producing high-quality written summaries and initial drafts of financial communications.

Explanation

This question tests the ability to match Gemma, Google's open family of generative AI models, to a business task where text generation is the primary requirement. Producing summaries and first drafts is a direct and efficient use of a text-generation model.

✗ **A** The complexity of building and deploying sophisticated internal knowledge bases to answer employees' finance-related questions with accurate and up-to-date information.
Incorrect. Gemma can generate responses, but building an accurate and current internal knowledge base generally requires additional retrieval, data integration, and grounding components rather than Gemma alone.

✗ **B** The difficulty in analyzing large datasets of financial transactions and market data to identify anomalies and predict future financial performance.
Incorrect. Anomaly detection and financial forecasting are primarily data analytics and machine learning tasks that require specialized statistical or predictive models and data pipelines. Gemma is a generative language model and is not the most efficient primary solution for this problem.

✗ **C** The struggle to accurately extract key financial figures and insights from a variety of document formats, such as balance sheets and income statements, for quick reporting.
Incorrect. Document extraction across varied formats generally requires document-processing, OCR, or multimodal capabilities and validation workflows. Gemma may support parts of a broader workflow, but it is not the most direct solution described here.

https://cloud.google.com/vertex-ai/generative-ai/docs/open-models/use-gemma

✓ **D** The challenge of efficiently producing high-quality written summaries and initial drafts of financial communications.
Correct. Summarization and drafting are direct text-generation use cases for Gemma, allowing analysts to automate routine writing while retaining human review for accuracy, compliance, and financial judgment.

Question 11

Single choice · Domain: D2 Google Cloud's gen AI offerings › Identify Security Command Center as the centralized platform for monitoring and managing security risks across Google Cloud environments

A security team needs a centralized platform to gain a comprehensive overview of their organization's security health across their entire Google Cloud environment, including potential threats to their generative AI deployments. Which Google Cloud security offering is specifically for this purpose?

A. Workload monitoring tools
B. Security Command Center
C. Identity and Access Management
D. Secure-by-design infrastructure

Answer: B. Security Command Center

Explanation

This question tests recognition of the Google Cloud security offering that provides centralized visibility into security posture, vulnerabilities, and threats, including risks relevant to generative AI workloads.

✗ **A** Workload monitoring tools
Incorrect. Workload monitoring tools can provide operational metrics and performance information, but they are not the centralized security posture and threat-management platform described in the question.

https://cloud.google.com/monitoring/docs

✓ **B** Security Command Center
Correct. Security Command Center is Google Cloud's centralized security management platform for gaining visibility into security posture, vulnerabilities, threats, and risks across cloud resources and workloads.

https://cloud.google.com/security-command-center

✗ **C** Identity and Access Management
Incorrect. Identity and Access Management controls authentication and authorization, but it does not provide the comprehensive, organization-wide security health and threat overview described in the question.

https://cloud.google.com/iam/docs/overview

✗ **D** Secure-by-design infrastructure
Incorrect. Secure-by-design infrastructure is a security design approach rather than a centralized platform for viewing security posture and threats across a Google Cloud environment.

https://cloud.google.com/security/infrastructure-design

Dumpsgate

Question 12

Single choice · Domain: D2 Google Cloud's gen AI offerings › Identify how Google Cloud RAG APIs ground generative AI responses in current enterprise data

A development team is building an internal knowledge base chatbot to answer employee questions about company policies and procedures. This information is stored across various documents in Google Cloud Storage and is updated regularly by different departments. What is the primary benefit of using Google Cloud's RAG APIs in this scenario?

A. They provide a pre-built user interface for the chatbot, simplifying the front-end development process.

B. They allow the development team to train a single foundation model on all company documents.

C. They enable the generative AI model to retrieve the most up-to-date and relevant information from the policy documents in real-time.

D. They automatically create summaries of all company policies, which are then presented to employees as quick answers.

Answer: C. They enable the generative AI model to retrieve the most up-to-date and relevant information from the policy documents in real-time.

Explanation

This question tests knowledge of the primary value of retrieval-augmented generation APIs for enterprise chatbots: retrieving relevant information from connected data sources at query time so responses reflect current documents.

✗ **A** They provide a pre-built user interface for the chatbot, simplifying the front-end development process.
Incorrect. RAG APIs provide retrieval and grounding capabilities, not a complete pre-built chatbot user interface.

https://cloud.google.com/vertex-ai/generative-ai/docs/rag-engine/rag-engine-overview

✗ **B** They allow the development team to train a single foundation model on all company documents.
Incorrect. RAG retrieves relevant source information at query time; it does not require retraining a foundation model on all company documents.

https://cloud.google.com/vertex-ai/generative-ai/docs/rag-engine/rag-engine-overview

✓ **C** They enable the generative AI model to retrieve the most up-to-date and relevant information from the policy documents in real-time.
Correct. Retrieval-augmented generation connects the model to external data, allowing it to retrieve relevant information from current policy documents and use that information to ground its responses.

https://cloud.google.com/vertex-ai/generative-ai/docs/rag-engine/rag-engine-overview

✗ **D** They automatically create summaries of all company policies, which are then presented to employees as quick answers.
Incorrect. RAG APIs retrieve relevant content for a user query; they do not automatically summarize every policy and publish those summaries as the primary function.

https://cloud.google.com/vertex-ai/generative-ai/docs/rag-engine/rag-engine-overview

Question 13

Single choice · Domain: D3 Techniques to improve gen AI model output › Identify how retrieval-augmented generation uses current external information to improve response accuracy and relevance

A company is developing a generative AI-powered customer support chatbot. They want to ensure the chatbot can answer a wide range of customer questions accurately, even those related to recently updated product information not present in the model's original training dat

a. What is a key benefit of implementing retrieval-augmented generation (RAG) in this chatbot?

A. RAG will significantly reduce the computational resources required to run the generative AI model.

B. RAG will primarily help the chatbot generate more creative and engaging conversational responses.

C. RAG will enable the chatbot to fine-tune its underlying language model on the fly based on customer interactions.

D. RAG will enable the chatbot to access and utilize external, up-to-date knowledge sources to provide more accurate and relevant answers.

Answer: D. RAG will enable the chatbot to access and utilize external, up-to-date knowledge sources to provide more accurate and relevant answers.

Explanation

This question tests understanding of RAG as a grounding technique that retrieves information from external data sources and provides it to a generative model at inference time. This is particularly useful for answering questions about information that was not included in the model's original training data or has changed recently.

✗ **A** RAG will significantly reduce the computational resources required to run the generative AI model.
Incorrect. RAG adds a retrieval step and does not primarily reduce the computational resources required to run the underlying generative AI model.

https://cloud.google.com/vertex-ai/generative-ai/docs/grounding/overview

✗ **B** RAG will primarily help the chatbot generate more creative and engaging conversational responses.
Incorrect. RAG is primarily intended to improve factual grounding, accuracy, and relevance by supplying retrieved information, rather than directly increasing creativity or conversational engagement.

https://cloud.google.com/vertex-ai/generative-ai/docs/grounding/overview

✗ **C** RAG will enable the chatbot to fine-tune its underlying language model on the fly based on customer interactions.
Incorrect. RAG retrieves and provides external information at response-generation time; it does not fine-tune the underlying language model on the fly.

https://cloud.google.com/vertex-ai/generative-ai/docs/grounding/overview

✓ **D** RAG will enable the chatbot to access and utilize external, up-to-date knowledge sources to provide more accurate and relevant answers.
Correct. Retrieval-augmented generation grounds model responses in information retrieved from external sources, including current product documentation, which can improve answer accuracy and relevance for knowledge not present in the model's original training data.

https://cloud.google.com/vertex-ai/generative-ai/docs/grounding/overview

Free demo · full exam: 76 questions

Question 14

Single choice · Domain: D3 Techniques to improve gen AI model output › Select grounding to reduce hallucinations by anchoring generated content in authoritative source information

A global news company is using a large language model to automatically generate summaries of news articles for their website. The model's summary of an international summit was accurate until it hallucinated by stating a detail that did not occur. How should the company overcome this hallucination?

A. Implement stricter safety settings to filter out potentially controversial topics.
B. Fine-tune the model on a larger dataset of news articles.
C. Increase the temperature setting of the model to encourage more diverse outputs.
D. Use grounding to base the model output on the source articles.

Answer: D. Use grounding to base the model output on the source articles.

Explanation

This question tests the ability to choose an output-grounding technique when a generative AI model produces an inaccurate factual claim. Grounding connects the model's response to relevant source content, such as the original news article.

✗ **A** Implement stricter safety settings to filter out potentially controversial topics.
Incorrect. Safety settings are intended to address harmful or disallowed content, not to ensure that factual summaries remain supported by the source article.

https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/configure-safety-attributes

✗ **B** Fine-tune the model on a larger dataset of news articles.
Incorrect. Fine-tuning can adapt a model to specific tasks or styles, but simply using a larger dataset does not directly ensure that each generated summary is factually grounded in the specific source article.

https://cloud.google.com/vertex-ai/generative-ai/docs/models/tune-models

✗ **C** Increase the temperature setting of the model to encourage more diverse outputs.
Incorrect. Increasing temperature generally makes outputs more variable and can increase the risk of unsupported statements; it does not anchor the summary to the article's facts.

https://cloud.google.com/vertex-ai/generative-ai/docs/multimodal/content-generation-parameters

✓ **D** Use grounding to base the model output on the source articles.
Correct. Grounding connects the model's response to verifiable information from a specified source, such as the original news article, which helps reduce hallucinations and improve factual accuracy.

https://cloud.google.com/vertex-ai/generative-ai/docs/grounding/overview

Question 15

Single choice · Domain: D4 Business strategies for a successful gen AI solution › Identify the business problem and desired outcomes before selecting a generative AI use case

A company wants to choose a generative AI (gen AI) use case that will be successful and have the most impact. What key factor should they determine first according to Google Cloud-recommended practices?

A. The number of employees who will be trained to use the new gen AI tools.
B. The specific business problems the company aims to solve and the desired outcomes.
C. The availability of pre-trained models that are offered on various cloud computing platforms.
D. The frequency of updates to the underlying foundation models used by different gen AI platforms.

Answer: B. The specific business problems the company aims to solve and the desired outcomes.

Explanation

This question tests the ability to prioritize business value by defining the problems to solve and the outcomes to achieve before evaluating gen AI tools, models, or implementation requirements.

✗ **A** The number of employees who will be trained to use the new gen AI tools.
Incorrect. Employee training is important for adoption, but it should follow the identification of the business problem and desired outcomes.

https://cloud.google.com/transform/ai

✓ **B** The specific business problems the company aims to solve and the desired outcomes.
Correct. Google Cloud-recommended practices emphasize starting with a clear business problem and measurable desired outcomes so that the gen AI use case can be evaluated based on business impact.

https://cloud.google.com/transform/ai

✗ **C** The availability of pre-trained models that are offered on various cloud computing platforms.
Incorrect. Available models are implementation considerations; they should be assessed after defining the business need and expected value.

https://cloud.google.com/transform/ai

✗ **D** The frequency of updates to the underlying foundation models used by different gen AI platforms.
Incorrect. Foundation-model update frequency may affect platform selection and lifecycle management, but it is not the first factor in choosing a high-impact use case.

https://cloud.google.com/transform/ai

Question 16

Single choice · Domain: D4 Business strategies for a successful gen AI solution › Identify explainability as a responsible AI control for fair and accountable lending decisions

A home loan company is deploying a generative AI system to automate initial loan application reviews. Several applicants have been unexpectedly rejected, leading to customer complaints and potential bias concerns. They need to ensure responsible and fair lending practices. What aspect of the AI system should they prioritize?

A. Implementing stricter data security measures to protect applicants' financial information from unauthorized access.
B. Ensuring AI decision-making is explainable to understand decision reasons and establish accountability.
C. Increasing the speed at which the AI system processes loan applications to handle the high volume.
D. Regularly updating the AI model with more financial data to improve its accuracy over time.

Answer: B. Ensuring AI decision-making is explainable to understand decision reasons and establish accountability.

Explanation

This question tests the ability to prioritize responsible AI practices when automated decisions may produce unexpected outcomes or create bias concerns. Explainability helps organizations understand decision reasons, investigate potential unfairness, and establish accountability.

✗ **A** Implementing stricter data security measures to protect applicants' financial information from unauthorized access.
Incorrect. Data security is essential for protecting sensitive financial information, but it does not directly address whether loan decisions are biased, explainable, or accountable.

https://cloud.google.com/responsible-ai

✓ **B** Ensuring AI decision-making is explainable to understand decision reasons and establish accountability.
Correct. Explainability enables the company to understand why applications were rejected, investigate possible bias, communicate decision reasons, and establish accountability for automated lending decisions.

https://cloud.google.com/vertex-ai/docs/explainable-ai/overview

✗ **C** Increasing the speed at which the AI system processes loan applications to handle the high volume.
Incorrect. Increasing processing speed may improve operational efficiency, but it does not address unexpected rejections, fairness, transparency, or accountability.

https://cloud.google.com/responsible-ai

✗ **D** Regularly updating the AI model with more financial data to improve its accuracy over time.
Incorrect. More data may improve accuracy, but additional data alone does not guarantee fair outcomes and could reinforce existing bias if the data or labels are problematic. Responsible AI practices require evaluating fairness and understanding model behavior, not only improving accuracy.

https://cloud.google.com/responsible-ai