ATMIC NETWORKS
YOUR GUIDE TO GLOBAL CERTIFIATION

"Please note that these files may not be up to date. However, the questions will help you understand the exam format and typical question patterns."

www.atmicnetworks .com

Warning: Keep connected with our support team for latest updates

Question: 1

A logistics company wants to use a generative AI (gen AI) agent to automatically check real-time inventory levels across its warehouses and adjust delivery schedules. The gen AI agent needs access to internal inventory dat

a. They want the most cost-effective solution. What should the organization do?

A. Build a custom API instead of using the gen AI agent.
B. Use pre-built gen AI chatbots for inventory questions.
C. Use Vertex AI Studio to fine-tune a model with sample inventory data.
D. Use Google Cloud databases and Vertex AI for the agent to get live data.

Answer: D

Explanation:
To achieve real-time inventory checks and adjust delivery schedules, the generative AI agent needs live access to the company's internal inventory data. Google Cloud databases provide the structured storage for this data, and Vertex AI offers the platform to build, deploy, and manage the AI agent, including connecting it to these live data sources. This approach allows the agent to make informed decisions based on current information. Building a custom API for every interaction might be less cost-effective in the long run for dynamic inventory data. Pre-built chatbots might not have the direct integration needed for real-time adjustments, and fine-tuning with sample data wouldn't provide the live data access required.

Question: 2

A pharmaceutical company's research and development department spends significant time manually reviewing new scientific papers to identify potential drug targets. They need a solution that can answer questions about these documents and provide summarized insights to researchers without requiring extensive coding expertise. What should the organization do?

A. Use Gemini for Google Workspace to facilitate collaborative document review.
B. Use Vertex AI Search to index the papers and enable keyword-based searches.
C. Use Vertex AI AutoML to train a model that classifies papers into predefined research areas.
D. Use Vertex AI Agent Builder to create a custom AI agent.

Answer: D

Explanation:
The requirement is to answer questions about the documents and provide summarized insights without requiring extensive coding expertise. Vertex AI Agent Builder is designed precisely for creating custom AI agents, often with low-code or no-code capabilities, that can interact with and process large volumes of information like scientific papers. While Vertex AI Search could index papers for keyword searches, it doesn't directly answer questions or provide summarized insights in the same way a generative AI agent built with Agent Builder could. Gemini for Google Workspace is for collaborative work, not specifically for building custom AI agents for document analysis. Vertex AI AutoML is for training classification models, which is

different from answering questions and summarizing.

Question: 3

The office of the CISO wants to use generative AI (gen AI) to help automate tasks like summarizing case information, researching threats, and taking actions like creating detection rules. What agent should they use?

A. Security agent
B. Data agent
C. Code agent
D. Customer service agent

Answer: A

Explanation:
Given the tasks involve researching threats and creating detection rules, the most appropriate and specialized agent would be a Security agent. This type of agent would be pre-configured or easily adaptable to understand security-specific contexts, data, and actions within a CISO's domain.

Question: 4

A development team is configuring a generative AI model for a customer-facing application and wants to ensure the generated content is appropriate and harmless. What is the primary function of the safety settings parameter in a generative AI model?

A. To limit the maximum text length that the model generates by ensuring concise responses.
B. To determine the number of tokens the model can process at once by influencing the complexity and length of inputs and outputs.
C. To filter out potentially harmful or inappropriate content from the model's output based on the desired level of filtering.
D. To control the creativity and randomness of the model's output by adjusting the diversity of word choices.

Answer: C

Explanation:
Safety settings in generative AI models are specifically designed to prevent the generation of content that could be harmful, offensive, or inappropriate. This includes filtering for categories like hate speech, sexually explicit content, self-harm, and violence, based on predefined thresholds. Options A, B, and D refer to other parameters like max_output_tokens or temperature, which control output length, input/output processing, and creativity, respectively, not safety.

Question: 5

What is a characteristic of Google Cloud as a generative AI company?

A. Google Cloud provides fully autonomous AI agents that require zero configuration or management overhead.
B. Google Cloud has an AI-first focus that enables innovation, with continuous updates and broad integration across its platform.
C. Google Cloud ensures that all generative AI models and data are completely secured and isolated from external networks.
D. Google Cloud relies on proprietary, closed-source AI technologies for maximum security benefits.

Answer: B

Explanation:
Google Cloud emphasizes an AI-first approach, integrating AI capabilities across its services and consistently innovating with new models and features. While security is a high priority, fully autonomous AI agents requiring zero configuration are generally not the norm, and "completely secured and isolated from external networks" is an oversimplification of cloud security models. Google also contributes to and supports open-source AI initiatives, not solely relying on proprietary closed-source technologies.

Question: 6

A company wants a generative AI platform that provides the infrastructure, tools, and pre-trained models needed to build, deploy, and manage its generative AI solutions. Which Google Cloud offering should the company use?

A. BigQuery
B. Vertex AI
C. Google Kubernetes Engine (GKE)
D. Google Cloud Storage

Answer: B

Explanation:
Vertex AI is Google Cloud's unified machine learning platform that provides end-to-end support for the ML lifecycle, including access to pre-trained models (foundation models), tools for fine-tuning, deployment, and management of generative AI solutions. BigQuery is a data warehouse, GKE is for container orchestration, and Cloud Storage is for object storage; while they might be components used with Vertex AI, they are not the comprehensive generative AI platform themselves.

Question: 7

A highly regulated financial institution wants to use Gemini as the core decision engine for a loan approval system that will deterministically approve or reject loan applications based on a strict set of predefined

criteri

a. Why is this an inappropriate use case for Gemini?

A. Gemini cannot integrate with required financial databases.

B. Gemini is not equipped to handle structured numerical data for financial assessments.

C. Gemini is designed for flexible content generation and inference, not rigid rule-based decisions.

D. Gemini deployment for this scenario would be too expensive and complex.

Answer: C

Explanation:

Gemini, as a large language model, excels at flexible content generation, summarization, understanding, and inference. However, it is not designed for deterministic, rule-based decisionmaking that requires absolute consistency and adherence to strict, predefined criteria, as is common in highly regulated financial systems like loan approvals. Such systems typically require traditional programming logic or specific rule engines for auditable and consistent outcomes.

Question: 8

A company wants to adopt generative AI and is concerned about vendor lock-in. They want to maintain flexibility in their technology stack. What Google Cloud strength would ease their concerns?

A. Google Cloud's AI solutions have an open approach that supports customer choice across offerings.

B. Google Cloud's AI solutions are pre-packaged for easy deployment, eliminating the need for customization and integration efforts.

C. Google Cloud's strict adherence to proprietary technologies ensures the highest level of security and performance.

D. Google Cloud's focus on automation aims to replace human jobs with AI systems, potentially leading to significant workforce reductions.

Answer: A

Explanation:

Google Cloud promotes an open and flexible approach to its AI offerings, supporting open standards, open-source initiatives (like TensorFlow, Kubernetes, and Gemma), and providing various integration options. This helps alleviate vendor lock-in concerns by giving customers choice and control over their technology stack.

Question: 9

What will Google Cloud's Agent Assist help a company achieve?

A. The infrastructure to provide an enterprise-grade contact center solution with omnichannel support, routing, and integration with CRM systems.

B. The ability to analyze conversational data to identify customer sentiment, common topics of discussion, and insights into agent performance and customer experience.

C. The ability to provide real-time assistance and recommended responses to live customer service agents during their interactions.

D. The ability to build and deploy deterministic and generative chatbot agents for automated customer support.

Answer: C

Explanation:
Google Cloud's Agent Assist is specifically designed to augment human customer service agents. It provides real-time suggestions, retrieves relevant information, and offers recommended responses to agents during live interactions, improving their efficiency and consistency.

Question: 10

A software development team wants to use generative AI (gen AI) to code faster so they can launch their software prototype quicker. What should the team do?

A. Use gen AI to refactor and optimize existing code.
B. Use gen AI to suggest code snippets and complete functions.
C. Use gen AI to automatically generate comprehensive documentation for their code.
D. Use gen AI to identify potential bugs and security vulnerabilities in their code.

Answer: B

Explanation:
While generative AI can assist with all the options listed (refactoring, documentation, bug identification), its most direct and significant impact on coding faster for a prototype is through code generation. Suggesting code snippets and completing functions directly accelerates the writing of new code, enabling quicker prototyping.

Question: 11

What does Vertex AI Search enable companies to do?

A. To index and retrieve information from the entire public web, providing a comprehensive view of publicly available data.
B. To surface the most popular and frequently accessed content based on global user search patterns and trends.
C. To compare products from numerous online retailers, allowing users to find the best deals and product options across the internet.
D. To ground LLM responses with first-party data, third-party data, and Google's knowledge graph.

Answer: D

Explanation:

Vertex AI Search is designed to enable powerful search experiences over an organization's own data (first-party), external data (third-party), and can leverage Google's knowledge graph to provide more relevant and accurate responses, especially when grounding Large Language Models (LLMs). It does not index the entire public web like Google Search.

Question: 12

A large e-commerce company with a substantial product catalog and many support documents has customers struggling to find information on their website. This leads to high support costs and poor user experience. The company wants a Google Cloud solution to improve website search and reduce support costs while improving customer satisfaction. What Google Cloud product should the company use?

A. Vertex AI Search
B. Vertex AI Platform
C. Google Shopping
D. Google Search

Answer: A

Explanation:
Vertex AI Search is ideal for this scenario. It allows companies to build sophisticated search experiences over their own product catalogs and support documents. This improves accuracy and helps customers find what they need, directly addressing high support costs and poor user experience. Vertex AI Platform is broader for general ML development, Google Shopping is for consumers, and Google Search is for the public web.

Question: 13

A financial services company receives a high volume of loan applications daily submitted as scanned documents and PDFs with varying layouts. The manual process of extracting key information is timeconsuming and prone to errors. This causes delays in loan processing and impacts customer

satisfaction. The company wants to automate the extraction of this critical data to improve efficiency and accuracy. Which Google Cloud tool should they use?

A. Natural Language API
B. Dataflow
C. Vision AI
D. Document AI API

Answer: D

Explanation:
Document AI API is specifically designed for intelligent document processing. It uses machine learning to extract structured data from unstructured documents like scanned forms and PDFs, even with varying layouts.

This directly addresses the challenge of automating data extraction from loan applications. Natural Language API focuses on text understanding, Vision AI on image analysis (not structured extraction from documents), and Dataflow is for data processing pipelines.

Question: 14

A company is defining their generative AI strategy. They want to follow Google-recommended practices to increase their chances of success. Which strategy should they use?

A. Rapid implementation strategy
B. Bottom-up strategy
C. Multi-directional strategy
D. Top-down strategy

Answer: D

Explanation:
Google Cloud often recommends a "top-down" approach for generative AI strategy. This means starting with clear business objectives and leadership alignment on how generative AI can solve critical business problems, rather than simply experimenting from the bottom up without a clear strategic direction.

Question: 15

A company wants to use generative AI to create a chatbot that can answer customer questions about their products and services. They need to ensure that the chatbot only uses information from the company's official documentation. What should the company do?

A. Use role prompting.
B. Adjust the temperature parameter.
C. Use prompt chaining.
D. Use grounding.

Answer: D

Explanation:
Grounding is the technique of "grounding" the LLM's responses in specific, authoritative data sources (like the company's official documentation). This prevents the model from "hallucinating" or providing information outside of the approved knowledge base, ensuring accuracy and relevance to the company's specific products and services.

Question: 16

A company is developing an AI character for a video game. The AI character needs to learn how to navigate a complex environment and make decisions to achieve certain objectives within the game. When the AI takes actions that lead to positive outcomes, like finding a reward or overcoming an obstacle, it receives a positive

score. When it takes actions that lead to negative outcomes, like hitting a wall or losing progress, it receives a negative score. Through this process of trial and error, the AI gradually improves the character's ability to play the game effectively. What machine learning should the company use?

A. Reinforcement learning
B. Unsupervised learning
C. Supervised learning
D. Deep learning

Answer: A

Explanation:
This scenario perfectly describes reinforcement learning. In reinforcement learning, an agent learns to make decisions by interacting with an environment, receiving1 rewards for desirable actions and penalties for undesirable ones,2 and iteratively improving its behavior through trial and error to maximize cumulative reward.

Question: 17

A retail company with a large online catalog wants to improve customer experience and drive sales by implementing multimodal search capabilities (image, voice, and text). What is a primary business benefit of this capability?

A. Improved customer engagement and product discovery leading to increased satisfaction and potential sales.
B. Reduced dependency on keyword optimization for product listings and improved search engine rankings.
C. Lowered operational costs associated with managing and updating product information across different platforms and channels.
D. Streamlined inventory management processes and more accurate demand forecasting for popular items.

Answer: A

Explanation:
Multimodal search directly enhances the customer experience by allowing them to find products using various intuitive methods (images, voice, text). This leads to easier product discovery, higher engagement, and ultimately increased customer satisfaction and potential sales, which is a primary business benefit.

Question: 18

A company wants to create an AI-powered educational solution that provides personalized learning experiences for students. This platform will assess a student's knowledge, recommend relevant learning materials, and generate personalized exercises. The application would provide the structure for lessons and track progress. What type of AI solution should they use?

A. An AI-powered recommendation system for learning resources
B. A large language model fine-tuned on educational content
C. A learning management system (LMS)
D. A customized learning agent

Answer: D

Explanation:
The request goes beyond just recommendations or content generation. It involves assessing knowledge, recommending materials, generating personalized exercises, providing lesson structure, and tracking progress. This implies a more comprehensive, intelligent system that acts as an assistant or tutor for the student, which is best described as a customized learning agent. This agent would likely leverage LLMs and recommendation systems as components, but the overall solution is an agent.

Question: 19

What is a primary benefit of using a multi-agent system?

A. To simplify the most basic and repetitive rule-based tasks.
B. To consolidate all unique AI functions into a single, undifferentiated model.
C. To serve as a platform for hosting traditional, non-AI applications.
D. To manage complex tasks that demand coordinated AI functions.

Answer: D

Explanation:
Multi-agent systems are designed to tackle complex problems by breaking them down into sub-tasks,

where each agent specializes in a specific function. These agents then coordinate and collaborate to achieve a larger, more intricate goal that a single, monolithic AI model might struggle with.

Question: 20

An organization needs an AI tool to analyze and summarize lengthy customer feedback text transcripts. You need to choose a Google foundation model with a large context window. What foundation model should the organization choose?

A. Gemini
B. CodeGemma
C. Imagen
D. Chirp

Answer: A

Explanation:
Gemini models are known for their large context windows, making them highly suitable for processing and

summarizing lengthy texts like customer feedback transcripts. CodeGemma is specialized for code, Imagen for image generation, and Chirp for speech.

Question: 21

A software developer needs a highly efficient, open-source large language model that can be finetuned on a local machine for rapid prototyping of a chatbot application. They require a model that offers strong performance in natural language understanding and generation, while being lightweight enough to run on limited hardware. Which Google-developed family of models should they use?

A. Veo
B. Gemini
C. Gemma
D. Imagen

Answer: C

Explanation:
Gemma is Google's family of lightweight, state-of-the-art open models, built from the same research and technology used to create the Gemini3 models. They are designed for developers to build innovative AI applications on their local machines or in the cloud, offering a balance of performance and efficiency suitable for limited hardware and rapid prototyping. Veo is for video generation, Gemini is typically larger and more general-purpose, and Imagen is for image generation.

Question: 22

An organization wants granular control over who can use and see their generative AI models and related resources on Google Cloud. Which Google Cloud security offering is specifically for this purpose?

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

Answer: A

Explanation:
Identity and Access Management (IAM) is the fundamental Google Cloud service that allows you to define who has what access to which resources. It provides granular control over permissions for users, groups, and service accounts, including access to generative AI models and related data.

Question: 23

A company is developing a conversational AI chatbot. They need to ensure the chatbot can engage in human-

like conversations and provide accurate information. What should they do to enhance the chatbot's ability to understand and respond effectively to user prompts?

A. Use prompt engineering techniques, like few-shot prompting, to provide the chatbot with examples of successful interactions.
B. Limit the chatbot's training data to prevent it from learning irrelevant information.
C. Use strict keyword matching to ensure that the chatbot only responds to specific commands.
D. Lower model temperature setting to produce more consistent and predictable responses.

Answer: A

Explanation:
Prompt engineering, especially techniques like few-shot prompting (providing examples of desired input-output pairs), is crucial for guiding a generative AI model to understand context and generate relevant, human-like responses. Limiting data or using strict keyword matching would severely restrict the chatbot's conversational ability, and lowering temperature makes responses less creative, not necessarily more understanding.

Question: 24

A human resources team is implementing a new generative AI application to assist the department in screening a large volume of job applications. They want to ensure fairness and build trust with potential candidates. What should the team prioritize?

A. Integrating the AI application with various job boards to maximize candidate reach.
B. Focusing on minimizing the processing time for each application to improve efficiency.
C. Ensuring AI operates transparently, especially regarding application evaluation and data usage. D. Ensuring that the AI application can automatically rank all candidates without requiring human review.

Answer: C

Explanation:
To ensure fairness and build trust, especially in sensitive areas like job applications, transparency in how AI evaluates applications and uses data is paramount. This involves understanding potential biases, explaining decisions (where possible), and ensuring human oversight.

Question: 25

A company's large learning model (LLM) is producing hallucinations that are a result of the Knowledge cutoff. How does retrieval-augmented generation (RAG) overcome this limitation?

A. RAG fine-tunes the LLM on specific customer query patterns to improve the speed and efficiency of response generation.
B. RAG enhances the creative writing capabilities of the LLM to generate more engaging and informative responses.

C. RAG enables the LLM to retrieve relevant and up-to-date information from knowledge sources.
D. RAG uses human oversight to ensure accuracy before presenting information to the customer.

Answer: C

Explanation:
The primary purpose of RAG is to address the "knowledge cutoff" and hallucination issues of LLMs. It does this by retrieving relevant, up-to-date information from external knowledge sources (like databases or documents) at inference time and then using this retrieved information to ground the LLM's generation, ensuring factual accuracy and relevance to the specific query.

Question: 26

An organization with a team of live customer service agents wants to improve agent efficiency and customer satisfaction during support interactions. They are looking for a tool that can provide realtime guidance to agents, suggest helpful information, and streamline the support process without fully automating customer conversations. Which component of Google's Customer Engagement Suite should they use?

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

Answer: A

Explanation:
As previously mentioned, Agent Assist is specifically designed for real-time support to human agents, providing them with suggestions and relevant information during live customer interactions.
Conversational Agents (chatbots) automate interactions, Conversational Insights analyze conversations after they occur, and Contact Center as a Service is the broader infrastructure.

Question: 27

A financial institution uses generative AI (gen AI) to approve and reject loan applications, but gives no reasons for rejection. Customers are starting to file complaints. The company needs to implement a solution to reduce the complaints. What should the company do?

A. Collect a larger and more diverse dataset for the gen AI model.
B. Implement explainable gen AI policies.
C. Fine-tune the gen AI model.
D. Develop fairness assessments for the gen AI model.

Answer: B

Explanation:
The core problem is the lack of reasons for rejection, leading to customer complaints. This falls under the

domain of explainable AI (XAI). Implementing explainable gen AI policies or mechanisms would allow the institution to provide transparency into how the AI made its decision, addressing the customer complaints directly. While other options might improve the model, they don't directly solve the transparency issue.

Question: 28

A large multinational corporation with geographically dispersed teams struggles with knowledge silos and inconsistent access to crucial internal information. What is a key business benefit of using Google Agentspace in this scenario?

A. Improved IT infrastructure management across offices.
B. Seamless knowledge sharing and collaboration across internal systems.
C. Enhanced data encryption and compliance for internal communications.
D. Automation of employee performance reviews using AI.

Answer: B

Explanation:
Google Agentspace (or similar agent-based frameworks) aims to connect and orchestrate various AI capabilities and data sources. In a scenario with knowledge silos, a key benefit would be to enable seamless knowledge sharing and collaboration by allowing agents to access, process, and

disseminate information across different internal systems and teams.

Question: 29

What is the definition of generative AI?

A. A type of artificial intelligence that enables a system to autonomously learn and improve using neural networks and deep learning.4
B. A type of artificial intelligence that can create new content and ideas, including text, images, music, and code.
C. A type of machine learning algorithm inspired by the human brain that is made up of interconnected nodes.
D. A type of predictive model that estimates a relationship by fitting a line to the observed data.

Answer: B

Explanation:
The defining characteristic of generative AI is its ability to create new, original content that resembles its training data. This includes various modalities like text, images, music, and code, rather than just classifying, predicting, or analyzing existing data.

Question: 30

A marketing team wants to use a foundation model to create social media and advertising campaigns. They want to create written articles and images from text. They lack deep AI expertise and need a versatile solution. Which Google foundation model should they use?

A. Gemma
B. Imagen
C. Gemini
D. Veo

Answer: C

Explanation:
Gemini is Google's most advanced and multimodal foundation model, capable of understanding and generating various forms of content, including text and images, from a single prompt. Its versatility makes it suitable for marketing teams that need to create diverse campaign materials without deep AI expertise. Imagen is specifically for image generation, Gemma is a family of smaller, open models, and Veo is for video generation.

Question: 31

A company has a machine learning project that involves diverse data types like streaming data and structured databases. How does Google Cloud support data gathering for this project?

A. Google Cloud provides tools such as Pub/Sub, Cloud Storage, and Cloud SQL.
B. The Gemini app is the primary Google Cloud tool for directly collecting data.
C. Google Cloud's strengths are in the data analysis tools such as BigQuery.
D. Google Cloud relies on Vertex AI to connect to external data.

Answer: A

Explanation:
Google Cloud offers a comprehensive suite of services for data ingestion and storage. Pub/Sub is for streaming data, Cloud Storage for various file types (including unstructured), and Cloud SQL for relational structured databases. These are fundamental for gathering diverse data. Gemini is a model, BigQuery is for analysis, and Vertex AI is for ML platform, not primary data collection tools themselves.

Question: 32

What is an example of unsupervised machine learning?

A. Analyzing customer purchase patterns to identify natural groupings.
B. Training a system to recognize product images using labeled categories.

C. Predicting subscription renewal based on past renewal status data.
D. Forecasting sales figures using historical sales and marketing spend.

Answer: A

Explanation:
Unsupervised learning deals with unlabeled data. Identifying "natural groupings" or clusters in customer purchase patterns (e.g., segmenting customers into different buying behaviors without predefined labels) is a classic example of unsupervised learning (clustering). Options B, C, and D are examples of supervised learning, as they involve labeled data for training (product categories, renewal status, sales figures).

Question: 33

A large e-commerce company with a vast and frequently updated product catalog finds that customers struggle to find products on their website, and support agents spend too much time finding detailed product information. The company wants to improve search accuracy and efficiency for both customers and support.
What Google Cloud solution should they use?

A. Vertex AI Conversation
B. Vertex AI Natural Language API
C. Pre-built RAG with Vertex AI Search
D. Vertex AI Model Garden

Answer: C

Explanation:
This scenario strongly points to the need for accurate and up-to-date information retrieval from a product catalog. Pre-built RAG (Retrieval-Augmented Generation) combined with Vertex AI Search is the ideal solution. Vertex AI Search can index the product catalog, and RAG can then use this indexed data to ground the responses of a generative AI model, ensuring that both customer searches and support agent queries retrieve precise and relevant product information.

Question: 34

A company's development team is eager to start building generative AI solutions with Google Cloud, but has limited experience in AI development. They need to launch their gen AI solution quickly.
What Google Cloud benefit would help the company achieve their goal?

A. Google Cloud's pre-trained models and low- and no-code AI tools and services.
B. Google Cloud's collaborative AI community and support forums connect developers with AI experts.
C. Google Cloud's comprehensive training materials and tutorials to help developers.
D. Google Cloud's focus on continuous improvement provides access to the latest AI tools, features, and best practices.

Answer: A

Explanation:
For a team with limited AI experience needing to launch quickly, leveraging pre-trained models (foundation models) and low-code/no-code tools significantly reduces the development burden and accelerates time to market. This allows them to build and deploy generative AI solutions without requiring deep expertise from scratch. While other options are helpful, this directly addresses the need for quick launch with limited experience.

Question: 35

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

Explanation:
The primary benefit of RAG (Retrieval-Augmented Generation) in this context is its ability to ensure the chatbot provides accurate and up-to-date information. By retrieving relevant and recent policy documents from Cloud Storage in real-time and then grounding the LLM's response with this information, the chatbot avoids hallucinating or providing outdated answers, which is crucial for an internal knowledge base.

Question: 36

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

Explanation:

Security Command Center is Google Cloud's comprehensive security management and data risk platform. It provides centralized visibility into security posture, identifies vulnerabilities, detects threats, and helps manage compliance across the entire Google Cloud environment, including services and deployments like generative AI.

Question: 37

A company wants to use an AI agent to automate some tasks. They want everyone to understand the different functions of an AI agent. What is the function of an AI agent in the context of gen AI?

A. To provide the computational resources needed to train and run gen AI models.
B. To store and manage large datasets used for training and running gen AI models.
C. To provide a user-friendly interface for interacting with gen AI models.
D. To analyze situations, use multiple tools, and make informed decisions without requiring constant human input.

Answer: D

Explanation:

An AI agent, especially in the context of generative AI, is designed to be more autonomous and capable than a simple model. Its function is to understand a goal, analyze a situation, leverage various tools (including other generative AI models or external APIs), and make decisions or take actions to achieve that goal, often with minimal human intervention.

Question: 38

A research team has collected a large dataset of sensor readings from various industrial machines. This dataset includes measurements like temperature, pressure, vibration levels, and electrical current, recorded at regular intervals. The team has not yet assigned any labels or categories to these readings and wants to identify potential anomalies, malfunctions, or natural groupings of machine behavior based on the sensor data alone. What type of machine learning should they use?

A. Reinforcement learning
B. Unsupervised learning
C. Deep learning
D. Supervised learning

Answer: B

Explanation:

Since the team has not yet assigned any labels or categories to the sensor readings and wants to identify "anomalies, malfunctions, or natural groupings" based on the data alone, this is a classic unsupervised learning problem. Unsupervised learning techniques like clustering or anomaly detection are used to find hidden

patterns or structures in unlabeled data.

Question: 39

A social media platform uses a generative AI model to automatically generate summaries of user- submitted posts to provide quick overviews for other users. While the summaries are generally accurate for factual posts, the model occasionally misinterprets sarcasm, satire, or nuanced opinions, leading to summaries that misrepresent the original intent and potentially cause misunderstandings or offense among users. What should the platform do to overcome this limitation of the AI- generated summaries?

A. Implement stricter safety settings to filter out potentially misinterpreted content altogether.
B. Increase the temperature parameter of the model to encourage more varied and less literal interpretations.
C. Decrease the output length of the summaries to make them more concise.
D. Incorporate a human-in-the-loop (HITL) review process to refine the summaries.

Answer: D

Explanation:
When AI struggles with nuances like sarcasm or satire, human oversight is often the most effective solution. A human-in-the-loop (HITL) process allows human reviewers to check, correct, and refine AI-generated content before it is published, ensuring accuracy and appropriateness, especially for sensitive or complex language.

Question: 40

A global news agency is developing a generative AI tool to quickly summarize breaking news articles as they emerge online. The goal is to provide their audience with rapid updates on fast-developing stories from various global sources. What Google Cloud solution should they use?

A. Document AI
B. BigQuery
C. Vertex AI Natural Language API
D. Grounding with Google Search

Answer: D

Explanation:
For summarizing breaking news articles as they emerge online from various global sources, the generative AI model needs access to current, broad, and rapidly updating information. Grounding with Google Search allows the LLM to pull in the latest information from the web, ensuring the summaries are current and comprehensive. While Vertex AI Natural Language API can summarize text, it wouldn't inherently have access to the latest breaking news unless explicitly fed.

Question: 41

A data science team needs a centralized and organized location to store its various model versions, track their metadata, and easily deploy them to the respective applications. What Google Cloud service should they use?

A. Cloud Storage
B. Model Registry
C. BigQuery
D. Vertex AI Pipelines

Answer: B

Explanation:
A Model Registry (specifically part of Vertex AI Model Registry) is designed precisely for managing the lifecycle of machine learning models. It provides a centralized repository for storing, versioning, tracking metadata, and facilitating the deployment of models, which is essential for MLOps. Cloud Storage is for raw data, BigQuery for data warehousing, and Vertex AI Pipelines for workflow orchestration.

Question: 42

An organization wants to use generative AI to create a marketing campaign. They need to ensure that the AI model generates text that is appropriate for the target audience. What should the organization do?

A. Use role prompting.
B. Use prompt chaining.
C. Use few-shot prompting.
D. Adjust the temperature parameter.

Answer: A

Explanation:
Role prompting is a technique where you instruct the generative AI model to "act as" a specific persona or character. By assigning the model a role (e.g., "Act as a marketing expert writing for a young, tech-savvy audience"), you can guide its tone, style, and content to be appropriate for the target audience of the marketing campaign.

Question: 43

An organization wants to use generative AI to create a chatbot that can answer customer questions about their account balances. They need to ensure that the chatbot can access previous portions of the conversation with the customer. Which prompting technique should they use?

A. Use zero-shot prompting.
B. Use role prompting.
C. Use few-shot prompting.

D. Use prompt chaining.

Answer: D

Explanation:
Prompt chaining (or conversational memory/context management) is the technique used to maintain the conversational context. It involves feeding previous turns of a conversation (or a summary of them) back into the model along with the current user query, allowing the chatbot to "remember" and reference past interactions for coherent and contextually relevant responses, especially crucial for tasks like checking account balances that span multiple turns.

Question: 44

A company is exploring Google 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 Agentspace?

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

Answer: C

Explanation:
Google Agentspace (or similar agent platforms) is designed to empower employees with AI-powered assistants that can navigate and interact with enterprise systems, analyze documents, and automate tasks. This directly leads to improved employee productivity and more efficient data interaction by leveraging AI to streamline workflows and provide faster access to information.

Question: 45

A marketing team wants to use a generative AI model to create product descriptions for their new line of eco-friendly water bottles. They provide a brief prompt stating, "Write a product description for our new water bottle." The model generates a generic, lackluster description that is factually accurate but lacks engaging language and doesn't highlight the environmental benefits that are key to their brand. What should the marketing team do to overcome this limitation of the generated product description?

A. Train the model on a dataset of marketing materials from other eco-friendly brands.
B. Add details to the prompt about the audience, tone, and keywords.
C. Increase the token count for the model to allow for longer descriptions.
D. Lower the temperature setting of the model to produce more consistent results.

Answer: B

Explanation:

The core problem described is a lackluster and generic output that fails to capture the desired tone and key information (environmental benefits). This is a classic limitation of zero-shot prompting (a brief, un-detailed prompt), where the generative AI model relies solely on its general training data and lacks the necessary context to produce a highly relevant and engaging response. The solution is to improve the quality of the prompt itself, a process known as Prompt Engineering.

Option A, training the model, is an expensive and time-consuming process (fine-tuning) that is usually unnecessary for stylistic or content-specific guidance that can be achieved with a good prompt. Options C and D control the length and creativity, respectively, but don't inject the missing information or brand requirements.

Adding details to the prompt is the most immediate and effective technique to guide the model. By specifying the target audience (e.g., eco-conscious consumers), the desired tone (e.g., enthusiastic, persuasive), and mandatory keywords (e.g., "sustainable," "BPA-free," "ocean-friendly"), the marketing team is effectively providing the model with the necessary constraints and context to produce a description that is tailored to their brand and marketing goals. This technique is

fundamental to improving the output of generative AI models without resorting to model customization.

Question: 46

A learning and development team wants to quickly create a new hire training video with a custom avatar and voiceover that matches their company's branding and key messaging. They did not receive any money to spend on the production. What should they do?

A. Generate the video frames with Imagen.
B. Prompt the Gemini app to create a video.
C. Train a model with Vertex AI and produce a video.
D. Create a video with Google Vids.

Answer: D

Explanation:

The scenario requires quick creation of a training video using a custom avatar and voiceover while adhering to zero cost for production.

Google Vids is an AI-powered video creation app (part of Google Workspace/Gemini features) designed to make video creation accessible for teams without the overhead of traditional production. It specifically offers features like AI avatars and voiceovers for content such as trainings, demos, and onboarding videos. This directly addresses the need for a low-cost, fast solution for a new hire training video with custom branding elements (custom avatars and voiceovers are a key feature of the tool).

Option A, Imagen, is a Google foundation model specialized for image generation, not the creation of structured, narrated training videos with avatars. Option B, using the Gemini app, is primarily for text, code, and multimodal chat/generation, and is not the dedicated Google application for video production. Option C, training a model with Vertex AI, is a highly technical, time-consuming, and expensive endeavor that violates the need for a quick and zero-cost solution. Therefore, using the purpose-built, gen AI-enabled Google Vids application is the correct and most efficient choice.

Question: 47

A user asks a generative AI model about the scientific accuracy of a popular science fiction movie.
The model confidently states that humans can indeed travel faster than light, referencing specific but entirely fictional theories and providing made-up explanations of how this is achieved according to the movie's "established science." The model presents this information as factual, without indicating that it originates from a fictional work. What type of model limitation is this?

A. Bias
B. Knowledge cutoff
C. Data dependency
D. Hallucination

Answer: D

Explanation:

The limitation described is the AI model generating a false or misleading response (humans traveling faster than light is scientifically impossible/unproven) and presenting it as fact (confidently stating a fictional theory is real) without the ability to indicate its uncertainty or the source's fictional nature. This is the definition of a Hallucination in generative AI.
AI Hallucinations occur when a Large Language Model (LLM) generates outputs that are factually incorrect, irrelevant, or nonsensical, despite being linguistically fluent and seemingly plausible. They arise because the model is designed to predict the most statistically probable next word or token based on its training data, even when it lacks information or when its training data contains a mixture of fact and fiction. The model is overconfident in its generated response, a behavior that diminishes user trust and reliability, especially in applications where factual accuracy is critical. While a knowledge cutoff (B) is a common cause of hallucinations when an LLM is asked about recent events, the core limitation of fabricating facts from its own hardwired knowledge is the hallucination itself. Data dependency (A) relates to the model's reliance on the quality and completeness of its training data, and while flawed training data can be a cause, the error mode of inventing facts is the Hallucination.

Question: 48

A company wants to use generative AI to create a chatbot that can answer customer questions about their products and services. They need to ensure that the chatbot only uses information from the company's official documentation. What should the company do?

A. Use role prompting.
B. Adjust the temperature parameter.
C. Use prompt chaining.
D. Use grounding.

Answer: D

Explanation:

The core requirement is to guarantee that the chatbot only uses information from the company's official documentation and does not rely on its general knowledge base. This is crucial for ensuring factual accuracy, relevance to the company's specific products, and preventing the generation of fabricated or incorrect information (hallucinations).

The specific technique designed to address this challenge is Grounding. Grounding is the process of connecting the Large Language Model's (LLM's) responses to a trusted, verifiable source of information, such as an organization's internal documents, databases, or live data feeds. When an LLM is grounded, it is forced to base its answers only on the provided context, effectively preventing

it from drawing on its broad, generalized training data. Grounding is often implemented using a method called Retrieval-Augmented Generation (RAG), particularly with tools like Google Cloud's Vertex AI Search, which indexes the official documentation and feeds the relevant snippets to the model.

Options A, B, and C address different aspects of model output: Role prompting sets the model's persona, adjusting temperature controls creativity, and prompt chaining manages conversation history, but none of these techniques restrict the model's source of truth to the official documentation. Therefore, Grounding is the correct and most effective technique for this requirement.

Question: 49

A national bank is overwhelmed by customer inquiries across multiple channels and needs an AI- powered solution to provide seamless, consistent support, empower customer support agents, and improve service quality. What Google Cloud product should the bank use?

A. Vertex AI Search
B. Gemini for Google Workspace
C. Google Contact Center as a Service
D. Gemini for Google Cloud

Answer: C

Explanation:

The bank's requirement is for a solution that provides seamless, consistent support across multiple channels and helps to empower customer support agents and improve service quality. This describes the need for a comprehensive, end-to-end customer service infrastructure.

Google Contact Center as a Service (CCaaS) is the full, cloud-native contact center solution offered by Google Cloud (part of the Customer Engagement Suite). It is specifically designed to unify customer interactions across various channels (phone, chat, web messaging) and provides the necessary infrastructure for routing, managing agent workflows, and ensuring a consistent and secure customer experience at scale. This solution goes beyond simply automating a chatbot.

While Vertex AI Search (A) can be used as a component within the solution to ground answers in an internal knowledge base, and Gemini for Google Workspace (B) can boost individual agent productivity, neither provides the comprehensive multi-channel contact center infrastructure that the scenario demands. The scale

and nature of the problem—unifying overwhelmed support across channels and empowering agents—requires an enterprise-grade platform, which is precisely the function of Google Contact Center as a Service.

Question: 50

A company is developing a generative AI application to analyze customer feedback collected through online surveys. Stakeholders are concerned about potential privacy risks associated with this data, as the feedback contains personally identifiable information (PII). They need to mitigate these risks before using the data to train the AI model. What action should the company prioritize?

A. Focusing on collecting only quantitative feedback data in future surveys.
B. Ensuring that the AI model is trained on a large and diverse dataset.
C. Implementing strong access controls to limit which teams can view the raw survey data.
D. Applying data anonymization techniques to remove or obscure sensitive data.

Answer: D

Explanation:

The problem is the existence of Personally Identifiable Information (PII) within the customer feedback data, which introduces privacy risks for the development and training of the generative AI model. The goal is to mitigate these risks before using the data to train the AI model.

According to Google's Responsible AI and data handling best practices, when sensitive data like PII is present in a dataset intended for model training, the most critical step to prioritize is data minimization and privacy protection at the source. This is often achieved through anonymization or de-identification.

Applying data anonymization techniques (D) directly addresses the risk by removing or obscuring the sensitive data elements. This prevents the PII from being embedded into the model's parameters during training, thereby eliminating the risk of data leakage or privacy violations in the AI application's outputs. This is a crucial early step in the ML lifecycle for datasets containing sensitive information.

Option C, implementing access controls, is a necessary security measure but is a reactive control that protects the raw data; it does not remove the PII risk from the derived model itself. Option A is a long-term change to data collection but doesn't solve the problem for the existing data. Option B relates to bias and accuracy, not specifically PII risk mitigation.

(Reference: Google Cloud's Secure AI Framework (SAIF) and Responsible AI principles emphasize protecting sensitive data at all stages of the ML lifecycle, with de-identification being the primary method before training.)

Question: 51

According to Google-recommended practices, when should generative AI be used to automate tasks?

A. When tasks are highly creative and require original thought.
B. When tasks involve sensitive information or require human oversight
C. When tasks are repetitive and rule-based.
D. When tasks are complex and require strategic decision-making.

Answer: C

Explanation:

The strategic value of Generative AI (Gen AI) in a business context, as taught in Google's courses, is primarily to enhance efficiency and productivity by taking over tasks that consume significant employee time.

Gen AI excels in automating tasks that:

Are repetitive and time-consuming, such as drafting initial emails, summarizing long documents, or generating code snippets. Automating these routine tasks (C) frees employees to focus on higher- value activities (like building customer relationships or strategic planning).

Involve the generation of new content based on patterns learned from large datasets (e.g., text, images, code).

Options A and D represent high-value, strategic work—highly creative or complex strategic decision- making—where human judgment and oversight remain paramount. While Gen AI can assist with these (e.g., brainstorming creative ideas or providing data-backed insights), it is generally not recommended for full automation. Option B explicitly requires human oversight due to its sensitive nature. Therefore, the best fit for full or augmented automation for efficiency is the handling of routine, repeatable, and non-complex tasks.

(Reference: Google Cloud documentation on Gen AI adoption and efficiency states that Gen AI transforms work by automating repetitive and time-consuming tasks to free up time for strategic thinking and creativity.)

Question: 52

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

Explanation:

The core problem is the model's hallucination—it invented a factual detail—in a context (news reporting) where factual accuracy is non-negotiable. To correct a factual error in a generative summary, the model must be constrained to speak only based on verifiable facts from a reliable source.

The most effective technique to combat hallucinations and ensure factual adherence is Grounding (D). Grounding connects the Large Language Model's (LLM's) output to a specific, trusted, and verifiable source of information. This is often implemented using Retrieval-Augmented Generation (RAG). In this scenario, grounding the summary model on the original source articles ensures that every generated statement is

directly entailed by the provided facts (the source article content). Option B, fine-tuning, is expensive and only updates the model's general knowledge and style; it does not prevent the model from guessing or fabricating details when retrieving information. Option C, increasing temperature, would make the output less consistent and more diverse, likely increasing the chance of hallucination, which is the opposite of the desired effect. Option A is unrelated to factual accuracy. Therefore, Grounding is the necessary step to anchor the model's responses to the

true content of the source articles.
(Reference: Google Cloud documentation on RAG/Grounding emphasizes that its primary purpose is to address the “knowledge cutoff” and hallucination issues of LLMs by retrieving relevant, up-to-date information from external knowledge sources and using this retrieved information to ground the LLM's generation, ensuring factual accuracy.)

Question: 53

What does Model Garden enable a company to do?

A. Discover, customize, and deploy existing models from Google and its partners.
B. Evaluate the performance of different models using various metrics.
C. Manage different versions of a model, including the code, data, and parameters used to train it.
D. Train new models from scratch using large datasets.

Answer: A

Explanation:
Model Garden is a key component of the Vertex AI Platform on Google Cloud, positioned as an AI/ML model library. Its core function is to provide a central, organized place for users to find and utilize a wide variety of machine learning assets.
Specifically, Model Garden enables customers to:

Discover a curated collection of models, including Google's latest Foundation Models (like Gemini and Imagen), specialized models, and enterprise-ready models from Google partners and the opensource community (e.g., Gemma).
Test and customize these models, often with tools like Vertex AI Studio for prompt tuning or finetuning with custom data.
Deploy the selected and customized models directly to applications with a consistent deployment pattern.
Options B and C describe features of other MLOps tools within Vertex AI (Model Evaluation and Model Registry/Metadata Management). Option D describes the Custom Training service within Vertex AI. Model Garden's unique value proposition is acting as the starting point: a marketplace or repository to discover and immediately deploy or customize existing, pre-trained models.
(Reference: Google Cloud documentation states that Model Garden on Vertex AI is a place to discover, test, customize, and deploy a wide variety of models from Google and Google partners, including first-party and open-source models.)

Question: 54

An order fulfillment team has an agent that automatically processes orders, updates inventory, sends shipping notifications, and handles returns. What type of agent is this?

A. A workflow agent
B. An employee productivity agent
C. A customer service agent
D. A conversational agent

Answer: A

Explanation:

Generative AI agents are typically categorized based on the goal they are designed to achieve. The agent described is performing a sequence of distinct, interconnected, operational tasks (processes orders, updates inventory, sends notifications, handles returns). These steps are typical components of a business workflow or process automation.

A Workflow Agent is an AI agent whose purpose is to automate and manage an entire business process or a complex multi-step sequence of operations that traditionally required manual handoffs between different systems or teams. It uses its large language model brain, coupled with tools (such as APIs to a CRM, Inventory database, or shipping system), to observe the state of a process (e.g., a new order), reason about the next step, and execute the necessary actions to move the process forward toward completion.

Customer Service Agents (C) and Conversational Agents (D) are focused on user interaction (chatbots, virtual assistants) rather than back-end transactional automation.

Employee Productivity Agents (B) typically focus on individual tasks like drafting emails, summarizing meetings, or internal search, not automating an end-to-end operational flow like order fulfillment. Therefore, an agent designed to automate a complete, multi-step business process like order fulfillment is correctly classified as a Workflow Agent.

(Reference: Google Cloud Generative AI training materials categorize agents based on function, with Workflow Agents being those designed to automate multi-step business processes and operational sequences.)

Question: 55

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

Explanation:

The problem centers on unexpected rejections and potential bias in a high-stakes, regulated domain (lending). In such a context, the central tenet of Responsible AI is transparency and fairness.

While all options are valid goals, the priority when facing bias concerns and customer complaints due to rejection is to provide accountability and verify the fairness of the automated decision. This is achieved through Explainable AI (XAI).

Ensuring AI decision-making is explainable (B) means building mechanisms that allow developers, regulators, and affected customers to understand why a specific decision (rejection) was made. Explainability is crucial for:

Auditing for bias: If the reasons for rejection can be traced (e.g., system rejects based on loan-to- value ratio, not race), bias can be identified and corrected.

Compliance: Financial services are heavily regulated, and the ability to explain a lending decision is often a legal or regulatory requirement.

Customer Trust: Providing a clear reason for rejection (even if the news is bad) reduces complaints and fosters confidence, directly addressing the core issue of unexpected rejections.

Options A, C, and D address security, speed, and accuracy, respectively, but Explainability is the direct mechanism for proving fairness and ensuring accountability, making it the most critical priority in this scenario.

(Reference: Google's Responsible AI principles and training materials highlight that in high-stakes domains like finance, explainability is essential for establishing trust, identifying and mitigating bias, and meeting regulatory compliance.)

Question: 56

A team is using a generative AI model to automatically generate short summaries of customer feedback. They need to ensure that these summaries are concise and easy to digest. What model setting should they adjust?

A. Top-p (nucleus sampling)
B. Safety settings
C. Temperature
D. Output length

Answer: D

Explanation:

The objective is to make the generated summaries concise—that is, to control their length.

In the configuration of a generative AI model, particularly a large language model (LLM), the parameter used to directly control the maximum size of the response is the Output Length parameter (often referred to as max_output_tokens or max_tokens). By setting a low limit on this parameter, the team can ensure that the model is forced to terminate its response once that limit is reached, resulting in a shorter, more concise summary that is "easy to digest," as requested.

The other parameters control different aspects of the output quality:

Temperature (C) controls the creativity or randomness of the output. Lowering it makes the output more predictable; raising it makes it more diverse. It does not control length.

Top-p (A) is a decoding method related to temperature that also controls the model's creativity by limiting the vocabulary from which it can choose the next token. It does not control length.

Safety settings (B) are used to filter and block the generation of harmful, illegal, or inappropriate content. They do not affect the length or conciseness of the output.

(Reference: Google Cloud's Generative AI documentation on model parameters explicitly lists max_output_tokens or Output Length as the setting used to determine the maximum size of a model's generated response.)

Question: 57

A sales manager wants to responsibly use generative AI (gen AI) to increase efficiency with their existing tasks. They want to allow the sales team to focus on building customer relationships and closing deals. How should the sales team use gen AI?

A. To replace the sales team's CRM system with a more intuitive and user-friendly interface.
B. To analyze customer interactions on social media and automatically generate sales pitches tailored to their public profiles.
C. To draft emails and provide real-time insights about customer needs.
D. To automate creative content like blog posts and social media updates to attract new leads.

Answer: C

Explanation:

The strategic goal is to boost sales efficiency by shifting the team's focus to high-value activities (relationships and closing deals) by automating repetitive administrative tasks.

Option C directly addresses this goal by leveraging Gen AI's core capabilities for text generation and summarization/analysis:

Drafting emails automates a major time sink for sales reps (a common, repetitive task).

Providing real-time insights automates the labor-intensive research and manual data analysis required to understand customer needs, giving the rep instant, actionable context.

Options A and D are less direct solutions for improving sales efficiency: Option A is an expensive, high-risk platform replacement, not an efficiency use case. Option D describes marketing tasks, which, while related, are not the primary, day-to-day tasks that sales reps perform to clear their schedules for relationship building.

Therefore, Gen AI's most effective role in sales is as a productivity assistant for drafting and quick research.

(Reference: Google Cloud documentation on sales enablement use cases emphasizes that Gen AI's role is to automate administrative and time-consuming tasks like drafting outreach messages and synthesizing customer information to enhance seller productivity, allowing them to focus on revenue-generating activities.)

Question: 58

A company is using a language model to solve complex customer service inquiries. For a particular

issue, the prompt includes the following instructions:
"To address this customer's problem, we should first identify the core issue they are experiencing. Then, we need to check if there are any known solutions or workarounds in our knowledge base. If a solution exists, we should clearly explain it to the customer. If not, we might need to escalate the issue to a specialist. Following these steps will help us provide a comprehensive and helpful response. Now, given the customer's message: 'My order hasn't arrived, and the tracking number shows no updates for a week,' what should be the next step in resolving this?" What type of prompting is this?

A. Zero-shot
B. Few-shot
C. Role-based
D. Chain-of-thought

Answer: D

Explanation:

The prompt explicitly instructs the Large Language Model (LLM) to perform a step-by-step reasoning process before arriving at the final answer. The instructions lay out a sequential series of intermediate steps: "first identify," "then check," "if a solution exists, explain," "if not, escalate." This technique is known as Chain-of-Thought (CoT) Prompting. CoT is a powerful prompt engineering technique where the user or developer explicitly includes intermediate reasoning steps in the prompt. This guides the model to break down a complex, multi-step problem into smaller, manageable, logical steps, significantly improving its reasoning ability and the accuracy of its final output for complex queries like customer service troubleshooting or multi-step analysis.

Zero-shot (A) would be the raw question without any structure.
Few-shot (B) would involve providing examples of successfully solved problems.
Role-based (C) would involve assigning a persona (e.g., "Act as a customer service expert") but would not explicitly mandate the sequential process.
The inclusion of the explicit steps ("first identify," "then check," etc.) is the defining characteristic of Chain-of-Thought prompting.
(Reference: Google's courses on Prompt Engineering classify Chain-of-Thought prompting as the technique that improves reasoning by explicitly giving the model a series of sequential, intermediate steps to follow to arrive at a better answer for complex tasks.)

Question: 59

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

Explanation:

Gemma is a family of lightweight, open-source Large Language Models (LLMs) from Google that are based on the same research and technology as the Gemini models. As an LLM, its core strength lies in language-based tasks, particularly the generation and summarization of text.

The problem that Gemma, or any pure LLM, can most efficiently address is:

Generating text: creating new content quickly (Option D).

Summarizing text: condensing long communications or documents (Option D).

Option D, producing high-quality written summaries and initial drafts, is a natural language generation task that aligns perfectly with the core function of an LLM like Gemma. It is a key productivity booster for analysts needing to draft reports or emails quickly.

Option B (Analyzing large datasets/predicting performance) requires traditional machine learning (ML) models or analytical tools like BigQuery ML, as LLMs are not specialized for numerical predictive modeling.

Option C (Extracting key financial figures from documents) is a task for a highly specialized tool like Google's Document AI.

Option A (Building internal knowledge bases for Q&A) is a broader use case that is best solved with a platform solution using RAG, such as Vertex AI Search, not just a base model.

(Reference: Google's description of the Gemma model family emphasizes its role as a flexible, open LLM that excels at language fundamentals, making it ideal for content creation, summarization, and other text generation tasks.)

Question: 60

A large company is creating their generative AI (gen AI) solution by using Google Cloud's offerings. They want to ensure that their mid-level managers contribute to a successful gen AI rollout by following Google-recommended practices. What should the mid-level managers do?

A. Perform continuous testing, measurement, and refinement based on user feedback and real- world performance data.

B. Create a robust data strategy to ensure teams can access high-quality, relevant data that is appropriate for training and fine-tuning gen AI models.

C. Drive gen AI adoption by identifying high-impact, feasible solutions that address specific challenges within their workflows.

D. Secure funding and resources for AI initiatives by demonstrating the potential return on investment to the chief financial officer (CFO).

Answer: C

Explanation:
Google's recommended strategy for a successful generative AI rollout involves a combination of topdown strategic alignment and bottom-up adoption. In this structure, the role of the mid-level manager is critical for driving tangible value within their specific domain.

Securing funding (D) is typically the responsibility of senior leadership or the steering committee. Creating a robust data strategy (B) is the domain of data governance teams and data scientists. Continuous testing and refinement (A) is the job of MLOps/engineering teams and end-users. The primary role of the mid-level manager is to act as the bridge between high-level strategy and daily operations. They possess the domain knowledge to pinpoint pain points. Therefore, their most impactful contribution is to identify specific, high-impact, and feasible use cases (C) for their teams— such as automating report summaries or drafting internal communications—that directly address operational challenges and demonstrate quick wins. This action fuels successful adoption and validates the AI strategy from the ground up.
(Reference: Google Cloud's guidance on Gen AI strategy emphasizes that successful adoption requires strong top-down vision (like defining goals/funding) combined with bottom-up discovery, where functional leaders (mid-level managers) identify and prioritize high-value, feasible solutions within their specific workflows to drive adoption.)

Question: 61

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

Explanation:

The requirement is clearly focused on analytics and business intelligence derived from existing conversational data, specifically to understand trends and sentiment.
Conversational Insights is the dedicated component within Google's Customer Engagement Suite (which includes Contact Center AI) whose primary function is to analyze large volumes of interaction data (transcripts from chat, calls, etc.). It uses AI and Natural Language Processing (NLP) to extract valuable patterns, identify root causes of issues, and measure customer sentiment and agent performance. This analysis generates the actionable insights necessary for strategic planning and overall customer experience improvement.

Google Cloud Contact Center as a Service (CCaaS) (A) is the full platform for managing all channels and agents, but it's the system, not the analytical tool.
Agent Assist (B) is a real-time tool used by live agents for suggestions during a conversation; it is a

productivity tool, not a retrospective analytics tool.
Conversational Agents (C) are the chatbots or virtual assistants used for automation, not the tool for analyzing their performance and the raw data.
(Reference: Google Cloud documentation on the Customer Engagement Suite states that Conversational Insights is the tool used for conversational analytics to surface business intelligence from historical customer interaction data, including sentiment and trend analysis.)

Question: 62

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

Explanation:

The core reason for the model's failure is that the training data itself was flawed (disproportionate demographic representation and outdated language). This flaw directly leads to the observed bias and poor performance on underrepresented groups and modern communication styles.
This is a classic example of Data Dependency, a fundamental limitation of all machine learning models, including generative AI. Data dependency refers to the absolute reliance of an AI model on the quality, completeness, and fairness of the data on which it was trained. Since the model essentially only mimics the patterns it learned from its dataset, if the dataset contains societal, demographic, or linguistic biases, the model will faithfully reproduce and amplify those biases in its output, leading to unfair classification for certain groups.

Hallucination (C) is the invention of facts or data.
Overfitting (D) is poor generalization because the model memorized the training data too well, typically resulting in very poor performance across all unseen data, not just specific demographics. Bias is the result of the data dependency, not the fundamental limitation itself.
(Reference: Google's training on Generative AI Limitations identifies Data Dependency as the fundamental limitation where the model is limited by the scope and quality of its training data, directly leading to issues of bias when the data is not diverse or representative.)

Question: 63

An organization wants to quickly experiment with different Gemini models and parameters for content creation without a complex setup. What service should the organization use for this initial

exploration?

A. Google AI Studio
B. Vertex AI Prediction
C. Vertex AI Studio
D. Gemini for Google Workspace

Answer: C

Explanation:

The requirement is for a tool that facilitates quick experimentation with Gemini models and parameters without requiring significant technical setup, specifically targeting content creation (prompting/tuning) within the enterprise environment.
Vertex AI Studio (C) is the low-code, web-based UI component of Google Cloud's unified ML platform (Vertex AI). It is explicitly designed for non-technical users, developers, and data scientists to:

Quickly prototype and test different Foundation Models (including Gemini, Imagen, and Codey). Experiment with model parameters (like Temperature, Top-P, and Max Output Tokens) through a user-friendly interface.
Refine prompts and set up initial tuning or grounding configurations before moving to large-scale production deployment.
Google AI Studio (A) is a very similar tool, but it's generally associated with non-enterprise/public prototyping for Google's models, whereas Vertex AI Studio is the enterprise-ready environment for Gen AI development on Google Cloud, which is the context of the exam.
Vertex AI Prediction (B) is the service for deploying and serving models for inference, not for initial experimentation.
Gemini for Google Workspace (D) is an application that uses Gen AI to boost productivity within apps like Docs and Gmail, but it does not provide the interface needed to experiment with models and tune parameters.
(Reference: Google Cloud documentation positions Vertex AI Studio as the low-code/no-code interface for rapidly prototyping, testing, and customizing Google's Foundation Models (like Gemini) before full production deployment.)

Question: 64

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

Explanation:
A Diffusion Model (or Denoising Diffusion Probabilistic Model) is a specific class of generative AI model that is best known for its ability to create highly realistic images (e.g., Google's Imagen and Stable Diffusion are based on this architecture).

The core mechanism of a diffusion model is a two-step process:

Forward Diffusion (Adding Noise): It learns how to gradually corrupt data (like an image) by adding random noise until the original content is completely indistinguishable.
Reverse Diffusion (Denoising): It then learns to reverse this process—to gradually remove the noise—starting from a random noise pattern and iteratively refining it, guided by a text prompt, until a clear, coherent, and high-quality piece of content (an image or video) is generated.
Option D accurately captures this mechanism: the model starts with pure noise and generates the final structured data (the image) by refining that noise.

Option A describes predictive AI (forecasting models).
Option C describes a database or storage service.
Option B describes a workflow agent or optimization AI.
(Reference: Google's training materials on Foundation Models define Diffusion Models as generative models that operate by gradually converting a state of random noise into a structured, meaningful output, most commonly for the generation of high-quality images and video.)

Question: 65

A customer service team wants to use generative AI to improve the quality and consistency of their email responses to customer inquiries. They need a solution that can guide the AI to adopt a helpful, empathetic tone while adhering to company policies. Which prompting technique should they use?

A. Prompt chaining that engages the AI in a conversation to gather the necessary information before generating the email response.
B. Role prompting that instructs the AI to act as an experienced customer service representative with corporate knowledge.
C. One-shot prompting that provides a single example of a good customer service email.
D. Few-shot prompting that provides examples of good and bad customer service emails.

Answer: B

Explanation:

The most direct and effective way to influence the style, personality, and knowledge context of an AI's response is through Role Prompting.
Role Prompting involves instructing the model to assume a specific persona (a "role") before responding. By assigning the AI the role of an "experienced customer service representative" (B), the model is implicitly directed to adopt a professional, helpful, and empathetic tone. Furthermore, specifying "with corporate knowledge" directs the model to prioritize responses consistent with internal company policies. This technique is a foundational element of prompt engineering, often used in conjunction with other methods (like grounding, if specific policy documents were needed) to dramatically shift the output style and relevance.
While Few-shot prompting (D) could provide examples to influence style, it's less efficient than a clear role instruction and still requires the model to infer the persona. Prompt Chaining (A) is used to manage multi-turn conversation memory, not to set the tone or persona. Therefore, defining the Role is the core technique for establishing both the desired tone and the necessary professional context in a single instruction.

(Reference: Google's documentation on prompt engineering for customer service shows examples where users begin the prompt with "I am a customer service representative" to set the tone and persona for the generated response, confirming Role Prompting as the technique for ensuring style and consistency.)

Question: 66

A company collects customer feedback through open-ended survey questions where customers can write detailed responses in their own words, such as "The product was easy to use, and the customer support was excellent, but the delivery took longer than expected." What type of data is this?

A. Unstructured data
B. Structured data
C. Labeled data
D. Quantitative data

Answer: A

Explanation:

Data is typically classified into two main types: structured and unstructured.

Structured data is highly organized, formatted for a predefined data model, and easily searchable in tabular form (e.g., columns and rows in a database, like customer names, order IDs, or star ratings). Unstructured data lacks a pre-defined format or organization.
The customer feedback described is a detailed, free-text response written in the customer's own words. This qualitative data, whether it is an email, an essay, or a long-form survey response, does not fit into fixed fields and requires advanced Natural Language Processing (NLP) or Generative AI techniques to extract meaning.
Since the text is non-tabular and has no inherent structure enforced by the collection method, it is correctly classified as Unstructured Data.
Quantitative data (D) refers to numerical values that can be counted or measured. Labeled data (C) is data that has been tagged with a meaningful output category, which this raw feedback has not yet received.
(Reference: Google's Generative AI Study Guides define Unstructured Data as data that does not have a predefined structure or data model, such as text documents, images, audio, and video. Free- text responses in a survey are a primary example of unstructured data.)

Question: 67

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

Explanation:

According to Google's principles for successful AI adoption, organizations should adopt a "problem- first" approach to ensure their investments deliver measurable value. The strategic choice of a use case should always be motivated by a clear business imperative.

Determining the specific business problems and desired outcomes (B) is the foundational step in any successful Gen AI strategy. Without a well-defined problem (e.g., "reduce customer response time by 30%") and a measurable desired outcome (e.g., "increase customer satisfaction scores"), any AI solution runs the risk of being a technology in search of a purpose, leading to limited adoption or failure to deliver meaningful ROI.

Options A, C, and D are considerations secondary to the initial strategic alignment:

Availability of models (C) only dictates the technical feasibility, not the business value.

Training employees (A) is a resource requirement, not the goal itself.

Model updates (D) is a technical concern related to model longevity, not the primary strategic driver for use case selection.

The priority is always to align the AI solution with high-value business objectives.

(Reference: Google Cloud Generative AI strategy guidelines state: "A fundamental principle for successful AI adoption, including generative AI, is to start with clear business problems and desired outcomes. Without a well-defined problem, the AI solution might not deliver meaningful value, regardless of the technology used. This 'problem-first' approach is crucial for impactful AI strategy.")

Question: 68

A company is trying to decide which platform to use to optimize its generative AI (gen AI) solutions.

Why should the company use Vertex AI Platform?

A. It provides a mechanism for efficient analysis and exploration of large datasets used in machine learning.
B. It provides gen AI coding assistance with enterprise security and privacy protection.
C. It provides scalable and cost-effective object storage for data used in machine learning workflows.
D. It provides a unified platform of tools for building, deploying, and managing machine learning.

Answer: D

Explanation:

Vertex AI is Google Cloud's core, end-to-end Machine Learning Operations (MLOps) platform, designed to cover the entire ML lifecycle.

The key benefit of Vertex AI, particularly for generative AI, is that it provides a unified platform (D) where all stages of AI development—from accessing foundation models in Model Garden, testing in

Vertex AI Studio, training and tuning (via tools like Reinforcement Learning from Human Feedback), to deploying, and monitoring models in production—can be managed from a single service. This significantly reduces complexity, improves collaboration between teams (data scientists, engineers, business leaders), and ensures enterprise-grade governance and scalability necessary for production Gen AI solutions.

Option A describes BigQuery.
Option B describes Gemini Code Assist.
Option C describes Cloud Storage.
Vertex AI is the overarching platform that integrates all these tools to deliver a streamlined MLOps workflow.
(Reference: Google Cloud documentation states that Vertex AI is the unified AI development platform that brings together Google Cloud services for building, deploying, and managing machine learning models and generative AI solutions.)

Question: 69

A company wants to build a model to classify customer reviews as positive, negative, or neutral.
They have collected a dataset of thousands of customer reviews, and each review has been manually tagged with the corresponding sentiment: positive, negative, or neutral. What machine learning should the company use?

A. Deep learning
B. Unsupervised learning
C. Reinforcement learning
D. Supervised learning

Answer: D

Explanation:

The machine learning approach is determined by the nature of the data available and the desired output.

Data Available: Customer reviews (input) that are manually tagged with a sentiment category (output/label).
Desired Output: A model that can classify new, untagged reviews into one of the predefined categories (positive, negative, or neutral).
This scenario perfectly aligns with the definition of Supervised Learning (D). Supervised learning is the machine learning paradigm where the model is trained on a labeled dataset—a dataset where the input data is explicitly paired with the correct output label. The model learns a function that maps the input (the review text) to the output (the sentiment tag) and is then used to predict the label for unseen data.

Unsupervised Learning (B) is used for unlabeled data to find hidden patterns or groupings (clustering), which is not the goal here.
Reinforcement Learning (C) is used for training an agent through trial and error using a system of rewards and penalties.

Deep Learning (A) is a type of model (using deep neural networks) that can be used for supervised learning, but the learning approach required here is definitively supervised.
(Reference: Google's training materials on Machine Learning Approaches define Supervised Learning as training a model using labeled data to make predictions or classifications for new, unseen inputs.
Sentiment analysis is a canonical example of a supervised learning classification task.)

Question: 70

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

Explanation:
TPUs (Tensor Processing Units) are custom-designed hardware accelerators developed by Google specifically for high-performance machine learning tasks. Their advantage lies in their architecture, which is optimized for the massively parallel matrix multiplication operations that form the mathematical backbone of deep learning and large language models (LLMs).

TPUs excel at parallel processing (C) for training and running machine learning workloads, allowing computations to be performed simultaneously across numerous cores. This makes them significantly faster and more efficient than traditional CPUs or even general-purpose GPUs for tasks like training massive generative models (e.g., Gemini).

TPUs are a core component of the Infrastructure Layer in the Generative AI landscape, providing the foundational compute resources.

While Google offers very small, specialized TPUs for the edge (like Edge TPU), the primary, large- scale advantage is in the cloud for accelerating training and inference for complex ML models. Options A describes the Edge TPU or Gemini Nano deployment strategy, not the general, key advantage. Options B and D misrepresent the function, as TPUs are compute hardware, not storage accelerators or general-purpose CPU replacements.

(Reference: Google's training materials on the Generative AI Infrastructure Layer explicitly list TPUs and GPUs as the physical hardware components providing the core computing resources needed for generative AI, with TPUs being specialized for accelerating ML workloads and parallel processing.)

Question: 71

An organization is collecting data to train a generative AI model for customer service. They want to ensure security throughout the ML lifecycle. What is a critical consideration at this stage?

A. Implementing access controls and protecting sensitive information within the training data.
B. Applying the latest software patches to the AI model on a regular basis.
C. Establishing ethical guidelines for AI model responses to ensure fairness and avoid harm.
D. Monitoring the AI model's performance for unexpected outputs and potential errors.

Answer: A

Explanation:

The stage mentioned is Data Collection/Training Data Preparation. In the machine learning lifecycle, this

initial stage is where raw data is ingested and processed. If the model is being trained for customer service, the data (e.g., customer transcripts) is highly likely to contain sensitive information (like Personally Identifiable Information or PII).
Therefore, the most critical security and privacy consideration at this stage is protecting the integrity and confidentiality of the data itself.
Implementing strong access controls and protecting sensitive information (A) is the essential first step in a secure AI pipeline, aligning with Google's Secure AI Framework (SAIF). If data access is not controlled and sensitive data is not de-identified or redacted before it is used for training, the resulting model could leak that sensitive information to users.

Options B, C, and D are all important controls, but they occur at later stages of the ML lifecycle:
B (Software patches/latest versions) is part of deployment and management.
C (Ethical guidelines/fairness) is a Responsible AI goal implemented via guardrails and testing (later stages).
D (Monitoring) is an MLOps step that happens after deployment.
The critical consideration at the data collection stage is ensuring the data's security and privacy before it influences the model.
(Reference: Google Cloud guidance on securing generative AI emphasizes that one of the most significant risks is data leakage, making safeguarding training data and implementing identity and access control the foundational steps in the data ingestion and preparation phases.)

Question: 72

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

Explanation:

The Generative AI landscape is often broken down into several functional layers: Applications, Agents, Platforms, Models, and Infrastructure.
The Infrastructure Layer is the foundation, providing the physical and virtual computing resources necessary to run and train the large models. These resources include servers, storage, networking, and most importantly, the specialized hardware accelerators required for high-volume, parallel computation.
The core hardware components are the Graphics Processing Units (GPUs) and the custom-designed Tensor Processing Units (TPUs) (A). These accelerators are optimized for the massive matrix operations fundamental to deep learning and Gen AI model training and inference.

Options B (User interfaces) and D (Tools and services) refer to the Application and Platform layers, respectively.
Option C (Pre-trained models) refers to the Model layer.
The physical hardware underpinning these abstract layers are the TPUs and GPUs.
(Reference: Google Cloud Generative AI Study Guides state that the Infrastructure Layer provides the core

computing resources needed for generative AI, including the physical hardware (like servers, GPUs, and TPUs) and the essential software needed to train, store, and run AI models.)

Question: 73

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

Explanation:

The central problem is the Large Language Model's (LLM's) knowledge cutoff, where it cannot answer questions about information that appeared after its training data was collected (e.g., recently updated product details).

Retrieval-Augmented Generation (RAG) is specifically designed to overcome this limitation. The process involves:

Retrieval: When a question is asked, the RAG system first searches an external, up-to-date knowledge source (like a vector database of current product docs).

Augmentation: It retrieves the most relevant, recent text snippets (the context).

Generation: This retrieved context is added to the user's prompt (augmentation) and sent to the LLM, forcing the model to ground its response in the current facts.

The key benefit is thus to enable the chatbot to access and utilize external, up-to-date knowledge sources (D). This ensures the answers are accurate and relevant to the most current product information, directly addressing the knowledge cutoff issue without requiring expensive model retraining.

Option B is the function of the Temperature setting, not RAG.

Option C describes an unproven and unscalable model update mechanism (fine-tuning is a separate process).

RAG is a process enhancement that prioritizes accuracy and relevance over merely reducing computation (A).

(Reference: Google Cloud documentation on RAG states that its primary purpose is to address the "knowledge cutoff" and hallucination issues of LLMs by retrieving relevant and up-to-date information from external knowledge sources at inference time and using this retrieved information to ground the LLM's generation, ensuring factual accuracy.)

Question: 74

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

Explanation:

The requirements are for a prebuilt tool that is designed for:

Analyzing uploaded private documents (lengthy PDFs).
Providing summarized insights (extracting KPIs and trends).
Offering citations (grounding the answers to the source material).
NotebookLM (C) is the Google tool explicitly designed for this use case. It is a generative AI powered notebook/research assistant that allows users to upload source documents (including PDFs), then ask questions and generate summaries or insights that are grounded in and cited back to the source documents. This makes it an ideal prebuilt solution for an analyst who needs to process complex, lengthy financial reports and verify the data with citations.

Gemini Advanced (A) and Gemini app (B) are general-purpose conversational tools that are not primarily focused on deep, grounded analysis of uploaded documents that require source citations for research integrity.
Gemini for Google Workspace (D) is limited to data already in Workspace apps (Docs, Gmail, Drive) and the manual copy/paste process would be inefficient for "several lengthy PDF documents." (Reference: Google's Generative AI Leader training materials highlight NotebookLM as the specific generative AI application built for research and information synthesis from uploaded documents, offering key features like grounding and citations back to the source material.)