

AI-103 Pass Test Guide | AI-103 Valid Braindumps Book



With the AI-103 certification exam you can climb up the corporate ladder faster and achieve your professional career objectives. Do you plan to enroll in the Microsoft AI-103 certification exam? Looking for a simple and quick way to crack the AI-103 test? If your answer is yes then you need to start Microsoft AI-103 Test Preparation with Microsoft AI-103 PDF Questions and practice tests. With the NewPassLeader Developing AI Apps and Agents on Azure AI-103 practice test questions you can prepare yourself shortly for the final Microsoft AI-103 exam.

As we all know, HR from many companies hold the view that candidates who own a AI-103 professional certification are preferred, because they are more likely to solve potential problems during work. And the AI-103 certification vividly demonstrates the fact that they are better learners. Concentrated all our energies on the study AI-103 learning guide we never change the goal of helping candidates pass the exam. Our AI-103 test questions' quality is guaranteed by our experts' hard work. So what are you waiting for? Just choose our AI-103 exam materials, and you won't be regret.

>> AI-103 Pass Test Guide <<

AI-103 Valid Braindumps Book & AI-103 Exam Demo

Passing the exam just one time is a good wish of every candidate. If you choose us, we can help you pass your exam in your first attempt. AI-103 exam braindumps are high quality, and you can improve your efficiency during the preparation. Furthermore, AI-103 exam dumps are cover most of the knowledge points for the exam, you can have a good command of the knowledge points during practicing. We have online and offline service for AI-103 Exam Materials, if you any questions bother you, you can just have a conversation with us or you can clarify the problem through email, and we will give you reply as quickly as we can.

Microsoft Developing AI Apps and Agents on Azure Sample Questions (Q48-Q53):

NEW QUESTION # 48

You have a Microsoft Foundry project that contains an agent named PaymentAgent.

PaymentAgent includes a function tool that issues customer refunds by using an external API.

You are creating a workflow in YAML.

You need to ensure that the workflow pauses for human approval and continues with the refund step only after approval is granted.

How should you complete the workflow definition? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

```

steps:
  - id: propose_refund
    type: agent
    agent: PaymentAgent
  - id: approval
    type:
      ask_question
      basic_chat
      data_transformation
  - id: execute_refund
    type: agent
    agent: PaymentAgent
    condition:
      approval == "approved"
      propose_refund.output != null
      true

```

Answer:

Explanation:

```

steps:
  - id: propose_refund
    type: agent
    agent: PaymentAgent
  - id: approval
    type:
      ask_question
      basic_chat
      data_transformation
  - id: execute_refund
    type: agent
    agent: PaymentAgent
    condition:
      approval == "approved"
      propose_refund.output != null
      true

```

Explanation:

type: ask_question

condition: approval == "approved "

The approval step must use type: ask_question because the workflow must pause and wait for a human response before the refund execution proceeds. Microsoft Foundry workflows support human-in-the-loop patterns where the workflow asks the user a question and awaits input before continuing; this pattern is explicitly intended for approval requests and clarifying questions. The workflow guidance also identifies workflows as declarative sequences that orchestrate agents and business logic, including branching

logic and human-in-the-loop steps.

The refund execution step must use condition: approval == "approved" so that the second invocation of PaymentAgent runs only when the approval response matches the required approval value. Using true would always execute the refund, bypassing the approval control. Using propose_refund.output != null would only confirm that the first agent step produced output; it would not prove that a human approved the refund.

data_transformation is also incorrect for the approval node because it sets or parses values rather than pausing for user input.

Reference topics: Microsoft Foundry workflows, human-in-the-loop workflow pattern, YAML workflow editing, agent orchestration, conditional execution, and workflow approval gates.

NEW QUESTION # 49

You have a Microsoft Foundry project that contains a high-traffic agent.

After a recent update, operational costs increase significantly.

Monitoring confirms that the volume of user traffic to the agent remains unchanged.

You suspect that changes to the request or response characteristics are causing the increase.

You need to identify whether the additional costs are driven by the model input size, the model output size, or expanded tool usage.

Which observability capability should you use?

- A. evaluation metrics
- **B. token usage**
- C. run success rate
- D. latency

Answer: B

Explanation:

The correct capability is token usage. In Microsoft Foundry observability, token consumption is the primary signal for diagnosing model-cost changes when request volume is unchanged. Token usage lets you distinguish whether costs increased because prompts became larger, retrieved or tool-provided context expanded, responses became longer, or agent execution added more model calls. Microsoft Foundry monitoring dashboards track operational metrics such as token consumption, latency, error rates, and quality scores, and the agent monitoring dashboard is specifically intended to help analyze token usage, latency, success rates, and evaluation outcomes for production traffic.

This directly matches the scenario because the issue is not more traffic, but changed request or response characteristics. Input tokens reveal whether the prompt, chat history, grounding data, or tool outputs being sent to the model increased. Output tokens reveal whether the model is generating longer completions.

Expanded tool usage can also increase cost indirectly by adding more tool results, intermediate calls, and context into subsequent model requests; Foundry tracing and observability capture tool usage and token consumption for agent runs.

Evaluation metrics assess response quality and safety, not cost drivers. Latency identifies performance delays, and run success rate measures reliability. Reference topics: Microsoft Foundry observability, agent monitoring dashboard, token consumption, cost analysis, tool usage, and production monitoring.

NEW QUESTION # 50

You have a Microsoft Foundry project that serves a high-volume chat app.

Most requests are simple FAQs, but some require advanced reasoning.

You need to reduce costs and latency for common queries, without degrading the quality of the responses to complex questions.

What should you do?

- **A. Use a model cascade that routes the requests to different models.**
- B. Route all the requests to a smaller model.
- C. Route all the requests to the most capable model.
- D. Increase the value of the max_tokens parameter for all the requests.

Answer: A

Explanation:

The correct choice is to use a model cascade that routes the requests to different models. In Microsoft Foundry, this pattern aligns with model routing: simple, low-risk prompts can be handled by smaller, faster, lower-cost models, while complex prompts can be escalated to more capable or reasoning models. Microsoft's Foundry model router guidance states that the router optimizes cost and latency while maintaining comparable quality by using smaller, cheaper models when they are sufficient and larger or reasoning models when the task requires more advanced capability.

This directly matches the scenario: most traffic consists of simple FAQs, so routing those requests to efficient models reduces average latency and token-processing cost. Advanced reasoning requests still receive high-quality responses because they are routed to models with stronger reasoning capability. Microsoft's model router documentation also explains that routing decisions consider prompt difficulty, cost, quality, latency, and conversation context, making it suitable for diverse chat workloads. Increasing `max_tokens` for all requests would usually increase cost and latency. Sending all requests to a smaller model risks poor quality for complex questions, while sending all requests to the most capable model wastes cost and latency on simple FAQs. Reference topics: Microsoft Foundry model routing, model selection, generative AI optimization, latency management, and cost-aware AI application design.

NEW QUESTION # 51

You have a Microsoft Foundry project named Project1.

Project1 contains an application that processes PDF vendor invoices.

You need to configure Azure Document Intelligence in Foundry Tools to generate a Markdown output that preserves the sections and table structure of the PDFs. The solution must minimize development effort.

What should you do?

- A. Configure `output=figures` when you analyze the PDF.
- B. Configure `content=markdown` when you analyze the document.
- C. Increase the confidence threshold.
- D. Set the `output_content_format=ContentFormat.MARKDOWN` value.

Answer: D

Explanation:

The correct answer is D. Set the `output_content_format=ContentFormat.MARKDOWN` value. Azure Document Intelligence Layout API can return extracted document content in Markdown format, preserving semantic structure such as headings, paragraphs, sections, tables, and other layout elements. Microsoft's Document Intelligence layout guidance shows the Python SDK pattern for analyzing a document with the prebuilt-layout model and setting `output_content_format=ContentFormat.MARKDOWN` in the `begin_analyze_document` call. The Markdown output is returned in the top-level `content` section of the analysis result. This minimizes development effort because the service produces structure-preserving Markdown directly, rather than requiring custom post-processing to reconstruct sections and table formatting from raw OCR spans. Microsoft's Markdown output documentation states that specifying Markdown output produces semantically structured content that maintains paragraphs, headings, tables, and other document elements in their proper hierarchy. Option A only changes validation behavior and does not generate Markdown. Option B requests figures, not structured Markdown. Option C uses an incorrect parameter name; the documented SDK setting is `output_content_format`, not `content`. Reference topics: Azure Document Intelligence Layout API, Markdown output, PDF analysis, table extraction, and Foundry Tools document processing.

NEW QUESTION # 52

You need to recommend an invoice review solution that resolves the issue reported by the finance department.

What should you include in the recommendation?

- A. chat completions
- B. Azure Document Intelligence in Foundry Tools
- C. Azure Content Understanding in Foundry Tools
- D. Image Analysis

Answer: C

Explanation:

The correct recommendation is Azure Content Understanding in Foundry Tools. The case study states that Contoso's finance department must manually review vendor invoices to verify that invoice details match vendor contract terms, and that the invoices contain tables, logos, and varied layouts that make consistent processing difficult. It also states that the planned solution must evaluate both the visual layout and textual content of the invoices.

Azure Content Understanding is designed for this type of multimodal document-processing workload.

Microsoft describes Content Understanding as a Foundry Tool that processes unstructured and multimodal content, including documents and images, and transforms it into structured output for AI applications. It can use document analyzers to extract text, layout, tables, fields, and relationships from diverse document types.

Chat completions alone would not reliably extract structured invoice fields from complex layouts. Azure Document Intelligence can

myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, myportal.utt.edu.tt, mænnc832558.blgwiki.com, Disposable vapes