

Microsoft

Exam Questions AI-103

Developing AI Apps and Agents on Azure



NEW QUESTION 1

- (Topic 1)

You need to configure Agent1 to answer customer questions about only the Contoso products. The solution must meet the business requirements. What should you do?

- A. Apply top-p sampling.
- B. Modify the system message instructions.
- C. Add few-shot examples.
- D. Increase the value of the temperature parameter.

Answer: B

NEW QUESTION 2

- (Topic 1)

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

Answer: A

NEW QUESTION 3

- (Topic 2)

You have a Microsoft Foundry agent that grounds responses from an Azure AI Search index containing:

? Searchable text fields for product names and product codes.

? A vector field containing embeddings for product descriptions.

You need users to query by exact product names or codes and by natural-language product descriptions.

- A. Semantic search only
- B. Hybrid search
- C. Keyword search only
- D. Vector search only

Answer: B

NEW QUESTION 4

HOTSPOT - (Topic 2)

You have a Microsoft Foundry project that contains an internal Q&A agent. Users report the following issues when they ask the agent questions:

- An increase in the following response: ??No relevant information found??
- Periodic HTTP 429 rate limit exceeded errors during peak hours

You need to identify whether each issue is caused by model unavailability, resource limits, or inference failures.

What should you do? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Metrics to enable:

Model Availability Rate and Provisioned Utilization
Only Tokens Cache Match Rate
Only Total Requests filtered to status code 200
Time To Response and Total Tokens

Diagnostic log to collect:

AllMetrics
audit
RequestResponse
trace

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Metrics to enable:

- Model Availability Rate and Provisioned Utilization
- Only Tokens Cache Match Rate
- Only Total Requests filtered to status code 200
- Time To Response and Total Tokens

Diagnostic log to collect:

- AllMetrics
- audit
- RequestResponse
- trace

NEW QUESTION 5

- (Topic 2)

You have a Microsoft Foundry project that contains a prompt agent used by a customer support web app.

The agent is invoked from a Python service that does NOT run in the Foundry portal.

You need to implement end-to-end tracing to capture latency breakdowns and exceptions across agent runs.

Which two components can you use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. a Log Analytics workspace
- B. OpenTelemetry
- C. Application Insights
- D. the Azure Monitor Agent
- E. Microsoft Sentinel

Answer: BC

NEW QUESTION 6

- (Topic 2)

You have an application that processes scanned PDF invoices. The invoices have varied layouts and include multipage tables.

You have a pipeline that uses optical character recognition (OCR) and extracts totals and invoice numbers. The results are often incorrect because the document structure is ignored.

You need to implement a solution that provides OCR, layout analysis, and template- generalizing field extraction. The solution must NOT require training a custom model. The solution must minimize administrative effort. What should you include in the solution?

- A. an Azure Machine Learning model
- B. Azure Content Understanding in Foundry Tools
- C. Azure Language in Foundry Tools

Answer: B

NEW QUESTION 7

- (Topic 2)

You have a Microsoft Foundry project that contains an agent and an image generation model deployment.

The agent generates original images from user-supplied product photos.

You need to ensure that the generated images maintain the product identity and visual characteristics of the provided photo.

What should you do?

- A. Set the input_fidelity parameter to high.
- B. Apply a groundedness detection filter.
- C. Include a prompt and input image in the request.
- D. Decrease the value of the temperature parameter.

Answer: A

NEW QUESTION 8

HOTSPOT - (Topic 2)

You have an Azure AI Search resource named Search1 that is used by multiple apps hosted in Azure. You need to secure Search1. The solution must meet the following requirements:

- Prevent access to Search1 from the internet.
- Limit the access of each app to query specific indexes.

What should you do? To answer, select the appropriate options in the answer area. NOTE: Each correct answer is worth one point.

Answer Area

To prevent access from the internet:

▼

Configure an IP firewall.
Create a private endpoint
Use Azure roles.

To limit access to query specific indexes:

▼

Create a private endpoint.
Use Azure roles.
User key authentication

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

To prevent access from the internet:

▼

Configure an IP firewall.
Create a private endpoint
Use Azure roles.

To limit access to query specific indexes:

▼

Create a private endpoint.
Use Azure roles.
User key authentication

NEW QUESTION 9

- (Topic 2)

You have a customer support agent that uses the Microsoft Foundry Agent Service.

Sometimes, customers return to a session days later to continue the same support case, and the agent must resume with the full historical context. The agent must provide the following:

- Multi-turn continuity within the session
- Cross-session continuity for the same case
- Access to the full interaction history, including user messages, agent messages, tool calls, and tool outputs

You need to ensure that the agent automatically reloads the complete history on each new turn.

What should you do?

- A. Persist only the final model response stored in the client application and prepend the response to future prompts.
B. Enable memory summarization on the agent definition to persist the context automatically.
C. Create and reuse a conversation by storing the conversation??s ID and supplying the ID on subsequent requests.

Answer: C

NEW QUESTION 10

- (Topic 2)

You plan to configure an evaluation in Microsoft Foundry for a Retrieval Augmented Generation (RAG) chat app. You need to provide scores for groundedness,

relevance, and harmful-content categories. Which two evaluation categories can you use? Each correct answer presents part of the solution.

- A. risk and safety metrics
- B. AI quality (NLP) metrics
- C. AI quality (AI assisted) metrics
- D. fluency evaluator
- E. similarity evaluators

Answer: AC

NEW QUESTION 10

- (Topic 2)

You have a Microsoft Foundry project. You need to deploy a model from the model catalog to support real-time inference. The solution must meet the following requirements:

- ? Use key-based authentication
- ? Support real-time REST API access
- ? Not consume the vCPU quota of the virtual machines in the Azure subscription

Which type of deployment should you use?

- A. serverless API
- B. self-hosted container
- C. standard
- D. batch

Answer: A

NEW QUESTION 13

- (Topic 2)

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

Answer: D

NEW QUESTION 18

HOTSPOT - (Topic 2)

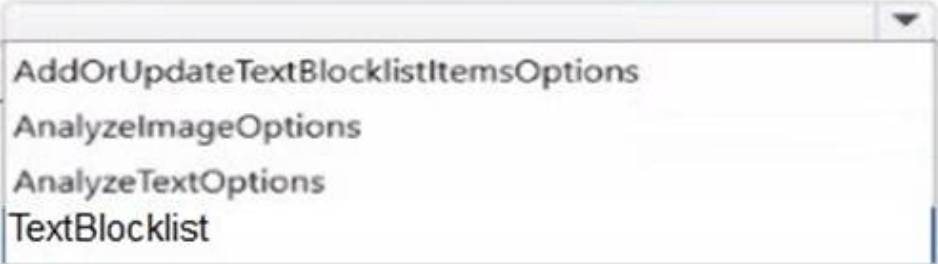
You have a configurable guardrail in a Microsoft Foundry project.

You have a Python application. A text blocklist named ConfidentialTerms already contains terms that must be detected in user messages. The application stores each incoming user message in a variable named input_text.

You plan to moderate input_text before the text is sent to a language model. You need to analyze input_text by using the existing blocklist.

Answer Area

```
...
blocklist_name = "ConfidentialTerms"
client = ContentSafetyClient(endpoint, credential)
request =
    text=input_text
    ...
```



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```
...
blocklist_name = "ConfidentialTerms"
client = ContentSafetyClient(endpoint, credential)
request =
    text=Input
    ...
    (
        AddOrUpdateTextBlocklistItemsOptions
        AnalyzeImageOptions
        AnalyzeTextOptions
        TextBlocklist
    )
```

NEW QUESTION 23

- (Topic 2)

You are building a speech processing solution in Microsoft Foundry for a customer support platform.

The platform will transcribe live phone calls, so that supervisors at your company can view call transcripts and detect issues while the calls are in progress. The call audio will arrive as a continuous stream from the telephony system.

You need to ensure that the call transcripts appear within only a few seconds of the audio stream.

What should you do?

- A. Run a batch transcription job on recorded audio files.
- B. Use real-time speech to text to process streaming audio input.
- C. Use speech translation to generate the transcripts into multiple languages.
- D. Use text to speech by using a custom neural voice.

Answer: B

NEW QUESTION 25

- (Topic 2)

In Microsoft Foundry, you use the Chat playground with the GPT-35 Turbo model. You have a prompt that contains the following code.

```
function F(n)
{
    var f = [0, 1];
    for (var i = 2; i < n; i++) f[i] = f[i-1] + f[i-2];
    return f;
}
```

You need the model to create an explanation of the code. The solution must minimize costs. What should you do?

- A. Add function F (explanation) to the prompt.
- B. Change the model to GPT-4-32k.
- C. Add// what does function F do? to the prompt.
- D. Set the temperature parameter to 1.

Answer: C

NEW QUESTION 30

- (Topic 2)

You are building an app that will share user images. You need to configure the app to:

? Categorize each image as a photograph or drawing.

? Generate a caption for the image.

? Minimize development effort.

Which two services should you include?

- A. Image type detection in Azure Vision in Foundry Tools
- B. Object detection in Azure Vision in Foundry Tools
- C. Content tags in Azure Vision in Foundry Tools
- D. Image classification in Azure Custom Vision
- E. Image descriptions in Azure Vision in Foundry Tools

Answer: AE

NEW QUESTION 34

- (Topic 2)
 You have a Microsoft Foundry project named Project1 that contains the following:

- An OpenAPI tool that calls an external API
- A project connection named Connection1 that stores the API key of the external API When an agent calls the OpenAPI tool, the API returns a 401 unauthorized error, and traces show that the API key header is NOT being sent.

You need to ensure that the OpenAPI tool automatically includes the API key from Connection1 on all requests.
 What should you do?

A. Configure the tool to use the default connection of Project1.
 B. Connect the tool to Connection1.
 C. Enable identity passthrough so that the tool uses the Microsoft Entra token of the caller.
 D. Add the API key header manually to the OpenAPI specification.

Answer: B

NEW QUESTION 37

- (Topic 2)
 You have an agent named Agent1 that uses Model Context Protocol (MCP) calls to retrieve external data. You need to implement guardrails to ensure that Agent1 cannot send any content tagged as Violence to the MCP server. Which intervention point should you use?

A. output
 B. user input
 C. tool call
 D. tool response

Answer: C

NEW QUESTION 42

HOTSPOT - (Topic 2)
 You have a Microsoft Foundry project that contains an agent.
 The agent uses tools to retrieve internal content and call external APIs. The agent is configured to let the model decide when to call the tools.
 You need to publish the agent for a compliance workflow. The solution must meet the following requirements:

- Each workflow run must include a retrieval step before generating a response.
- Tool calls must authenticate by using the published agent's own identity.
- Tool access must use an identity isolated from other project resources.
- Tool access must support audit tracing.

What should you do? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Set tool_choice to:

auto

none

required

Configure the tool to authenticate by:

Storing API keys in prompts

Using the shared project agent identity

Using a distinct agent identity bound to the client application

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Set tool_choice to:

auto

none

required

Configure the tool to authenticate by:

Storing API keys in prompts

Using the shared project agent identity

Using a distinct agent identity bound to the client application

NEW QUESTION 45

- (Topic 2)

You are deploying a support agent that enables users to upload photos.

You need to automatically classify uploaded images for harmful content. The solution must block content based on severity levels.

What should you do?

- A. Implement image moderation.
- B. Enable prompt shields.
- C. Apply keyword scanning to optical character recognition (OCR) output by using Azure Vision in Foundry Tools.
- D. Use blocklists.

Answer: A

NEW QUESTION 46

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

AI-103 Practice Exam Features:

- * AI-103 Questions and Answers Updated Frequently
- * AI-103 Practice Questions Verified by Expert Senior Certified Staff
- * AI-103 Most Realistic Questions that Guarantee you a Pass on Your First Try
- * AI-103 Practice Test Questions in Multiple Choice Formats and Updates for 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The AI-103 Practice Test Here](#)