

Microsoft

AI-103

Developing AI Apps and Agents on Azure

Questions & Answers (Demo)

Version: 4.0

Topic 1, Case Study Contoso, Ltd

Overview

Contoso, Ltd is a multinational retail company that builds, deploys, and manages generative AI and agent-based solutions by using Microsoft Foundry.

Identity Environment:

Contoso uses Microsoft Entra ID for identity management, authentication, and authorization capabilities that enable agents to access organizational resources and services.

Contoso recently formed a new AI engineering team named Agent1Dev Team to optimize and maintain existing AI solutions.

The team collaborates with solution architects, DevOps engineers, and security engineers to design, implement, monitor, and secure AI applications.

Contoso also has a team named Agent1Test Team that is responsible for validating AI solutions before the solution deployments.

Generative Environment:

Contoso has a Microsoft Foundry deployment that contains two projects named Project1 and Project2.

Project1

Project1 contains a customer support agent named Agent1 that assists customers with product inquiries and troubleshooting requests.

Agent1 has the following configurations:

Agent1 uses a base model deployment.

A safety evaluation pipeline is NOT enabled.

Tool invocation approval workflows are NOT enabled.

Conversation memory constraints are NOT configured.

Agent1 interacts with customers by using digital support channels and answers general questions about Contoso products.

Project1 is deployed to an Azure region located in the European Union (EU).

Agent1Dev Team will use Project1 to optimize and maintain Agent1.

Project2

Project2 contains a deployed video generation model. The marketing department at Contoso has access to Project2 and plans to use the model to develop a video creation solution.

Development of the solution is incomplete.

Data Environment:

Contoso stores product-related information in Azure resources that support AI applications.

The Azure environment contains an Azure Blob Storage account named storage1 that stores product detail sheets for all the Contoso products.

The product sheets include specifications, feature descriptions, and product support information that Agent1 can use to answer customer questions. The product sheets are stored in the PDF format.

Problem Statement:

Contoso identifies the following issues:

Agent1 has only general knowledge of the Contoso products.

A recent chat interaction with Agent1 was analyzed for sentiment. The results of the analysis have NOT been processed yet.

Agent1 does NOT use the detailed product information in the product sheets stored in storage1 when responding to customer questions.

The finance department at Contoso reports that vendor invoices must be reviewed manually to ensure that the invoices match the terms defined in the vendor contracts. The invoices contain tables, logos, and varied layouts that make the documents difficult to process consistently.

Requirement:**Planned Changes:**

Contoso plans to implement the following changes:

Implement a solution for Project1 that analyzes the vendor invoices by evaluating both the visual layout and the textual content of the invoices, so that the invoice details can be verified against the vendor contract terms.

Update the base model deployment used by Agent1 and standardize the model version to ensure continuity and consistent responses.

Enable Agent1 to retrieve and use the detailed product information from the product sheets stored in storage1.

Implement an indexing solution for the product sheets that Agent1 can use to answer customer questions.

Complete the development of the video creation solution.

Technical Requirements:

Contoso identifies the following technical requirements:

The model deployment used by Agent1 must support scalable, high-throughput generative AI workloads and dynamically scale to handle variable customer support traffic, without requiring reserved throughput capacity.

The product sheets must be processed by using an indexing pipeline that enables semantic and vector search, so that Agent1 can retrieve the relevant product information.

Responses generated by using the product sheet information must be relevant, complete, and accurate.

Agent1 must be able to use the product sheets to answer natural language questions about product details.

The model version used by Agent1 must remain consistent to ensure stable responses.

The data processed by the model must remain within the EU.

Safety and Compliance Requirements:

Contoso identifies the following security and compliance requirements:

API keys must NOT be used to access Foundry-deployed models.

Access to the Azure resources must follow the principle of least privilege.

The developers at Contoso must authenticate to Microsoft Foundry resources by using Microsoft Entra authentication.

Access to Project1 must be assigned to the members of Agent1Dev Team by using a security group named SC_Agent1_Dev.

Access to Project1 must be assigned to the members of Agent1Test Team by using a security group named SC_Agent1_Test.

Agent1 must never reveal customer information, even if a document that contains customer data is added erroneously to the product sheet repository in storage1.

The product sheets might contain images that include embedded text. Agent1 must be protected from malicious instructions potentially hidden within the images.

Business Information:

Contoso identifies the following business requirements:

Users that interact with Agent1 must have a personalized experience in future interactions, including the ability for Agent1 to retain conversation context and recall relevant information from previous interactions.

Agent1 must answer questions only about the products sold by Contoso.

Question: 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

Explanation:

The correct answer is B. Modify the system message instructions. The case study states that Agent1 answers general questions about Contoso products and that the business requirement is for Agent1 to answer questions only about the products sold by Contoso. This is a behavioral boundary for the agent, so it should be implemented in the highest-priority instructions that define the agent's role, allowed scope, and refusal behavior.

Microsoft Foundry guidance states that a system message is used to steer model behavior, define the assistant's role and boundaries, and add safety or quality constraints for the scenario. The system message should instruct Agent1 to answer only when the question concerns Contoso products, use the configured Contoso product documentation as grounding, and politely refuse or redirect questions about non-Contoso products.

Top-p sampling and temperature control randomness, not business-domain scope. Increasing temperature would make responses less deterministic. Few-shot examples can support desired behavior, but examples alone are weaker than explicit system-level instructions for defining operating boundaries. Reference topics: system message design, prompt engineering, agent instructions, response constraints, and grounded generative AI behavior.

Question: 2

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

Explanation:

The correct answer is B. Modify the system message instructions. The case study states that Agent1 answers general questions about Contoso products and that a business requirement is for Agent1 to answer questions only about products sold by Contoso. This requirement defines the agent's allowed domain and refusal boundary, so it must be expressed in the agent's system-level instructions. Microsoft Foundry guidance states that system messages steer Azure OpenAI chat model behavior and are used to define the assistant's role, boundaries, output format, and safety or quality constraints.

The system message should instruct Agent1 to answer only Contoso-product questions, use Contoso product documentation when available, and decline questions about non-Contoso products. This directly enforces the intended business scope at the highest instruction level. Few-shot examples can reinforce desired behavior but are not the primary control for defining mandatory operating boundaries. Top-p sampling and temperature are decoding controls; they influence randomness and diversity, not whether the agent restricts answers to a specific product domain. Increasing temperature would likely reduce consistency. Reference topics: Microsoft Foundry agent instructions, system message design, prompt engineering, response boundaries, and grounded generative AI behavior.

Question: 3

HOTSPOT

You need to ensure that Agent1Dev Team can access Agent1. The solution must meet the security and compliance requirements.

How should you complete the Python code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```

from azure.identity import DefaultAzureCredential
from azure.ai.projects import AIProjectClient
from azure.core.credentials import AzureKeyCredential
myEndpoint = "https://contoso.services.ai.azure.com/api/projects/project1"
project_client = AIProjectClient(
    endpoint=myEndpoint,
    credential=  ,
)
myAgent = "Agent1"
agent = project_client.agents.  (agent_name=myAgent)
print(f"Retrieved agent: {agent.name}")

```

Answer:

Explanation:

```

credential = DefaultAzureCredential()
agent = project_client.agents.get(agent_name=myAgent)

```

The correct authentication option is DefaultAzureCredential() because the case study states that API keys must not be used to access Foundry-deployed models and that Contoso developers must authenticate to Microsoft Foundry resources by using Microsoft Entra authentication. It also states that access to Project1 must be assigned to Agent1Dev Team by using the security group SC_Agent1_Dev. Microsoft Foundry authentication guidance recommends Microsoft Entra ID for production workloads because it supports least-privilege RBAC, per-principal auditing, and keyless authentication. AzureKeyCredential() would violate the no-API-key requirement, and None would not provide a valid credential.

The correct agent operation is get because the task is to access an existing agent named Agent1, not

create a new version or retrieve a specific published version. Microsoft Foundry SDK examples show `AIProjectClient` created with `DefaultAzureCredential()` and then using project agent operations to create, retrieve, or interact with agents by name. To meet the compliance requirement, the group `SC_Agent1_Dev` must also be granted the appropriate project-scoped Foundry role, such as Foundry User, for Project1. Reference topics: Microsoft Entra authentication, Foundry RBAC, `AIProjectClient`, and project agent access.

Question: 4

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

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 extract OCR, layout, and tables, but Content Understanding is the better end-to-end Foundry capability for combining visual and textual understanding with structured extraction for downstream verification. Image Analysis focuses on image-level visual features and is insufficient for invoice field and table review. Reference topics: Content Understanding, document analyzers, multimodal extraction, invoice processing, tables, layout, and structured JSON output.

Question: 5

You need to recommend a solution to support the planned changes and technical requirements for Agent1 to use the product information stored in storage1.

What should you include in the recommendation?

- A. Azure AI Search
- B. Azure Translator in Foundry Tools
- C. Azure Document Intelligence in Foundry Tools
- D. Grounding with Bing Search

Answer: A

Explanation:

The correct recommendation is Azure AI Search. The case study states that the product detail sheets are stored as PDFs in storage1, and that Agent1 must be enabled to retrieve and use detailed product information from those sheets. It also specifies that the indexing pipeline must enable semantic and vector search, and that Agent1 must answer natural language questions about product details by using the product sheet information. Azure AI Search is the Azure service designed to ingest content from sources such as Azure Blob Storage, create searchable indexes, and support keyword, semantic, hybrid, and vector retrieval for Retrieval Augmented Generation (RAG) solutions.

Microsoft's Azure AI Search guidance states that integrated vectorization can chunk content and generate embeddings during indexing, enabling vector search over source documents. It also states

that Azure AI Search supports text and vector queries and can improve raw content for search-related scenarios through enrichment pipelines. Azure Translator is unrelated to retrieval. Document Intelligence can extract document structure, but it is not the retrieval index for Agent1. Grounding with Bing Search retrieves public web content, not Contoso's private PDFs in storage1. Reference topics: Azure AI Search, RAG, semantic search, vector search, Azure Blob Storage indexing, and agent grounding.