

EXAM AI-901

Microsoft Azure AI Fundamentals Practice Question Bank

Every question, answer, and explanation from the exam simulator

112

QUESTIONS

2

EXAM DOMAINS

700

PASS SCORE / 1000

Fact-checked. Every question and answer has been checked against current Microsoft Learn documentation, and every explanation links to the pages that support it.

Take the timed, scored, interactive version online: www.mightybs.com/certifications/ai-901/exam/

Study hub: www.mightybs.com/certifications/ai-901/

Updated September 29, 2026. Independent study material, not affiliated with or endorsed by Microsoft. All company names are fictitious.

Part 1: Questions

Answers are not shown here, so you can test yourself first.

- Every question is multiple choice. Choose one answer unless the question says "Choose two" or "Choose three".
- Check your answers in [Part 2](#). Each question number links to its answer. On the real exam each correct selection is worth one point, and 700 of 1000 passes.

1 Multiple choice · Domain 1

Lamna Healthcare Company uses an AI model to recommend follow-up appointments for patients.

An audit shows that rural patients receive later follow-up recommendations than urban patients with similar symptoms and risk factors.

You need to identify the responsible AI principle raised by the audit.

Which principle should you identify?

- ☐ A Inclusiveness
- ☐ B Transparency
- ☐ C Accountability
- ☐ D Fairness

2 Multiple choice · Domain 1

Tailspin Toys is building App1 to screen job applications for warehouse roles.

The project team wants to reduce the risk that biased historical hiring data causes discriminatory recommendations.

You need to recommend a responsible AI action for App1.

Which action should you recommend?

- ☐ A Add a text-only interface for applicants who cannot use audio.
- ☐ B Evaluate outcomes across relevant groups and adjust the training data to reduce bias.
- ☐ C Encrypt the training data and restrict access to the model registry.
- ☐ D Publish a user guide that explains the model capabilities and limitations.

3 Multiple choice · Domain 1

Adventure Works uses an AI-powered robot in a warehouse to identify packages before moving them. If the model is not confident that an object is a package, the robot must avoid causing damage or injury. You need to apply a responsible AI consideration for the robot.

Which action should you recommend?

- ☐ A Describe the training dataset to users before the robot starts.
- ☐ B Require a minimum confidence score before the robot moves the object.
- ☐ C Delete temporary images after the shift is completed.
- ☐ D Translate the user interface into additional languages.

4 Multiple choice · Domain 1

Proseware is building App1, a generative AI assistant for customer support. App1 must reduce the risk that responses include harmful, illegal, or offensive content. You need to choose a responsible AI control for App1.

Which capability should you use?

- ☐ A Model benchmarks
- ☐ B Model interpretability
- ☐ C Content filters
- ☐ D Personal data filtering

5 Multiple choice · Domain 1

Wide World Importers uses facial identification to grant temporary access to a secure loading dock. Visitor images must be deleted after access expires and must be visible only to staff members who need them.

You need to identify the responsible AI principle addressed by the requirement.

Which principle should you identify?

- ☐ A Transparency
- ☐ B Fairness
- ☐ C Privacy and security
- ☐ D Inclusiveness

6 Multiple choice, choose two · Domain 1

Woodgrove Bank is building a generative AI assistant that summarizes customer support notes. The assistant must reduce the risk that sensitive customer data is exposed in generated responses. You need to recommend responsible AI controls for the assistant.

Which two actions should you recommend?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Restrict access to resources and operations by user or group.
 - ☐ B Add text captions for speech interactions.
 - ☐ C Publish a page that explains model limitations.
 - ☐ D Evaluate approval rates across demographic groups.
 - ☐ E Enable the personal data filter for model outputs.
-

7 Multiple choice · Domain 1

Fourth Coffee is building a speech-based ordering kiosk for stores. Customers with hearing impairments must be able to use the kiosk without relying on spoken output. You need to apply the responsible AI principle of inclusiveness.

Which action should you recommend?

- ☐ A Test order predictions across income groups.
- ☐ B Track who approved each model deployment.
- ☐ C Display text captions for spoken responses.
- ☐ D Encrypt stored audio recordings.

8 Multiple choice, choose two · Domain 1

Margie's Travel is designing App1, an AI travel assistant for customers who use different devices and assistive technologies.

App1 must avoid excluding customers from using the assistant.

You need to recommend inclusive design actions for App1.

Which two actions should you recommend?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Hide the fact that the assistant uses an AI model.
 - ☐ B Provide a text-based path in addition to voice interaction.
 - ☐ C Limit the assistant to the newest browser version only.
 - ☐ D Test the experience with users who have accessibility needs.
 - ☐ E Collect additional passport data before each chat session.
-

9 Multiple choice · Domain 1

Woodgrove Bank uses an AI assistant to help customers understand loan application decisions.

Customers must know when AI is used and understand the main limitations of the assistant.

You need to apply the responsible AI principle of transparency.

Which action should you recommend?

- ☐ A Evaluate loan recommendations across demographic groups.
- ☐ B Disclose the AI role and describe the assistant's capabilities and limitations.
- ☐ C Add a confidence threshold before automated physical actions.
- ☐ D Restrict customer records to authorized employees only.

10 Multiple choice, choose two · Domain 1

Fabrikam deploys App1, a generative AI assistant that summarizes internal policy documents.

Employees must understand what App1 can and cannot do, and they must be able to check where grounded answers came from.

You need to recommend transparency measures for App1.

Which two actions should you recommend?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Evaluate output quality across employee demographic groups.
 - ☐ B Encrypt all stored policy documents with managed keys.
 - ☐ C Provide documentation that explains App1 capabilities and limitations.
 - ☐ D Show citations to the policy sources used in grounded answers.
 - ☐ E Block medium-severity harmful content in user prompts.
-

11 Multiple choice · Domain 1

Contoso maintains an AI model that affects customer eligibility decisions.

For every production deployment, Contoso records who approved the model, why the change was made, and when the model was deployed.

You need to identify the responsible AI principle supported by this practice.

Which principle should you identify?

- ☐ A Accountability
- ☐ B Transparency
- ☐ C Inclusiveness
- ☐ D Fairness

12 Multiple choice, choose two · Domain 1

Wide World Importers is preparing an AI solution for an internal governance review.

Reviewers need evidence about model lifecycle decisions and ongoing operational issues after deployment.

You need to recommend accountability practices for the solution.

Which two practices should you recommend?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Capture lineage metadata for model publishing and deployment.
- ☐ B Monitor deployed applications and alert on lifecycle events.
- ☐ C Tune the dataset to reduce demographic outcome differences.
- ☐ D Add captions for every speech-based response.
- ☐ E Use a personal data filter to block customer identifiers.

13 Multiple choice, choose three · Domain 1

Alpine Ski House is preparing to release App1, a generative AI assistant in Microsoft Foundry.

App1 must reduce the risk of harmful or unsafe responses before it is deployed.

You need to follow responsible AI practices for reliability and safety.

Which three actions should you perform?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Disable harmful content controls to reduce false positives for all users.
- ☐ B Wait until production incidents occur before evaluating safety risks.
- ☐ C Identify harms by using red team testing and stress testing.
- ☐ D Apply guardrails and then repeat measurement to test effectiveness.
- ☐ E Define metrics and test sets to measure harm frequency and severity.
- ☐ F Use customer prompts to train foundation models without permission.

14 Multiple choice · Domain 1

Relecloud plans to use Foundry Models sold by Azure for an internal chatbot.

Security reviewers ask how customer prompts, completions, and fine-tuned models are protected from other tenants and model providers.

You need to provide a correct privacy and security statement.

Which statement should you provide?

- ☐ A Customer prompts and completions are not available to other customers or model providers.
 - ☐ B Customer prompts and completions are always stored in model weights for future users.
 - ☐ C Customer fine-tuned models are available to every Azure tenant by default.
 - ☐ D Customer prompts and completions are shared with other tenants to improve accuracy.
-

15 Multiple choice · Domain 1

You are explaining how a large language model first represents the text that it learns from.

You need to describe what tokens can represent in the model vocabulary.

Which description should you use?

- ☐ A Words, subwords, punctuation, or other common character sequences
 - ☐ B Only complete words that appear in a dictionary
 - ☐ C Only full sentences that appear in the training data
 - ☐ D Only numeric labels that identify whole documents
-

16 Multiple choice · Domain 1

You are describing how a transformer learns which words in a sequence are most influential for another word.

You need to identify the technique that assigns influence weights to tokens in context.

Which technique should you identify?

- ☐ A Batch deployment
- ☐ B Attention
- ☐ C Keyless authentication
- ☐ D Content filtering

17 Multiple choice · Domain 1

A user submits the prompt, "When my dog was a" to a language model.

You need to explain how the model produces a completion.

What should you explain?

- ☐ A The model searches only for an exact sentence match in stored documents.
 - ☐ B The model permanently retrains its parameters for each user prompt.
 - ☐ C The model predicts the next token, appends it, and repeats the process.
 - ☐ D The model converts the prompt into an API key before responding.
-

18 Multiple choice · Domain 1

Contoso wants a generative AI assistant to answer questions from the latest employee handbook.

The solution must use the current handbook without retraining the language model.

You need to describe how retrieval-augmented generation addresses the requirement.

What should the application do?

- ☐ A Disable the user prompt so the model answers from training data only.
 - ☐ B Replace the model tokenizer with a tokenizer built from the handbook.
 - ☐ C Store the handbook only in the system message for every future question.
 - ☐ D Retrieve relevant handbook passages and add them to the prompt as grounding context.
-

19 Multiple choice, choose two · Domain 1

Fabrikam is preparing approved policy documents so a RAG application can retrieve relevant passages.

You need to identify the steps used to create a searchable vector index.

Which two steps should you identify?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Divide source text into focused chunks.
- ☐ B Remove metadata such as source and permission information.
- ☐ C Choose a Global Batch deployment for the index.
- ☐ D Increase the prompt temperature for every chunk.
- ☐ E Encode the chunks as embedding vectors.

20 Multiple choice · Domain 1

Tailspin Toys is adding an FAQ assistant to a lightweight mobile app.

The app needs concise answers for a narrow, well-defined set of questions and must keep latency low.

You need to choose an appropriate model size category.

Which model category should you choose?

- ☐ A An image generation model
- ☐ B A speech to text model
- ☐ C A video generation model
- ☐ D A small language model

21 Multiple choice · Domain 1

Adventure Works is building App1 to answer questions about uploaded product photos and related text notes.

You need to select a model capability that can use image input and return text answers.

Which capability should you select?

- ☐ A Image generation
- ☐ B Batch processing
- ☐ C Multimodal chat
- ☐ D Embeddings

22 Multiple choice · Domain 1

Woodgrove Bank needs an AI feature that performs multi-step analysis of complex contracts and explains differences between versions.

You need to select a model capability suited to complex problem solving and document comparison.

Which capability should you select?

- ☐ A Reasoning
- ☐ B Image generation
- ☐ C Speech synthesis
- ☐ D Embeddings only

23 Multiple choice, choose two · Domain 1

Northwind Traders is shortlisting models in the Foundry model catalog before creating a deployment. You need to compare candidates using catalog information that is intended for model selection.

Which two catalog capabilities should you use?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Model details such as capabilities and supported inference tasks
- ☐ B The prompt temperature setting in a deployed playground
- ☐ C API keys copied from an endpoint details page
- ☐ D A resource firewall rule for private networking
- ☐ E Benchmark metrics or model leaderboards

24 Multiple choice, choose three · Domain 1

Proseware plans three AI features: semantic retrieval over support articles, product image creation from text, and transcription of customer calls.

You need to identify model capabilities or inference tasks that match the planned features.

Which three capabilities should you identify?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Developer deployment
- ☐ B Content filtering
- ☐ C Audio
- ☐ D Image generation
- ☐ E Managed compute
- ☐ F Global Standard
- ☐ G Embeddings

25 Multiple choice · Domain 1

Litware is configuring a customer support chat app.

The assistant must consistently use a friendly tone, stay within support topics, and follow refusal guidance before responding to user requests.

You need to choose the configuration element that sets the assistant behavior and boundaries.

What should you configure?

- ☐ A A vector index chunk
- ☐ B A deployment SKU
- ☐ C A model benchmark
- ☐ D A system message

26 Multiple choice · Domain 1

Coho Winery is deploying a newly released chat model for an app that has no data residency or reserved-throughput requirement.

You need to choose the deployment type that Microsoft recommends as the starting point for most workloads.

Which deployment type should you choose?

- ☐ A Global Standard
- ☐ B Developer
- ☐ C Data Zone Standard
- ☐ D Regional Provisioned

27 Multiple choice · Domain 1

Wide World Importers has a high-volume production app with predictable traffic.

The app needs reserved model processing capacity and lower latency variance. There is no data residency restriction.

You need to choose an appropriate deployment type.

Which deployment type should you choose?

- ☐ A Developer
- ☐ B Global Provisioned
- ☐ C Standard
- ☐ D Global Batch

28 Multiple choice, choose two · Domain 1

Fourth Coffee is testing a chat model in the Foundry playground.

Responses should be less creative and should not exceed the length that the app can display.

You need to configure the request parameters for the requirement.

Which two parameters should you configure?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Max output tokens
- ☐ B Project name
- ☐ C Temperature
- ☐ D Deployment region
- ☐ E Model family

29 Multiple choice, choose two · Domain 1

Relecloud has two AI apps that must process prompts and responses only within a Microsoft-specified data zone.

App1 has variable interactive traffic. App2 needs reserved, predictable throughput.

You need to recommend one deployment type for each app.

Which two deployment types should you recommend?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Data Zone Standard
- ☐ B Regional Provisioned
- ☐ C Global Standard
- ☐ D Data Zone Provisioned
- ☐ E Managed compute
- ☐ F Developer

30 Multiple choice · Domain 1

You are designing App1 for Proseware. Users enter rough notes about a new product.

You need App1 to create original product descriptions from the notes.

Which AI workload should you use?

- ☐ A Object detection
- ☐ B Information extraction
- ☐ C Speech recognition
- ☐ D Generative AI

31 Multiple choice · Domain 1

You are building Agent1 for Woodgrove Bank. A customer can ask Agent1 to find a policy, schedule a callback, and send a confirmation email.

You need to identify the workload that can reason over prompts and use tools to perform actions.

Which workload should you identify?

- ☐ A Agentic AI
- ☐ B Image classification
- ☐ C Speech synthesis
- ☐ D Key phrase extraction

32 Multiple choice, choose three · Domain 1

You are reviewing requirements for a new solution at Wide World Importers.

You need to identify requirements that match common AI workloads.

Which three requirements should you identify?

Each correct answer presents a complete solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Generate a concise answer to a user's product question in natural language.
- ☐ B Transcribe recorded customer calls into text.
- ☐ C Apply role-based access control to storage.
- ☐ D Replicate a database to another region.
- ☐ E Create a virtual network for application servers.
- ☐ F Identify products that appear in uploaded shelf photos.

33 Multiple choice · Domain 1

Tailwind Traders receives photos of paper expense receipts from employees.

You need to populate structured fields such as vendor, date, and total from each receipt.

Which AI workload should you use?

- ☐ A Sentiment analysis
 - ☐ B Speech synthesis
 - ☐ C Information extraction
 - ☐ D Image generation
-

34 Multiple choice · Domain 1

You are preparing a set of support comments for text analysis at Fourth Coffee.

You need to break each comment into smaller units such as words before you count terms.

Which technique should you use?

- ☐ A Speech synthesis
 - ☐ B Image generation
 - ☐ C Tokenization
 - ☐ D Semantic segmentation
-

35 Multiple choice · Domain 1

You are analyzing news articles for Lucerne Publishing.

The output must list mentions of people, organizations, locations, dates, and quantities.

You need to identify the text analysis technique for the requirement.

Which technique should you identify?

- ☐ A Named entity recognition
- ☐ B Summarization
- ☐ C Key phrase extraction
- ☐ D Sentiment analysis

36 Multiple choice, choose two · Domain 1

You are designing a review analysis workflow for Coho Winery.

The workflow must create searchable tags for each review.

The workflow must produce a short version of long reviews.

You need to choose text analysis techniques that meet the requirements.

Which two techniques should you use?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Sentiment analysis
- ☐ B Key phrase extraction
- ☐ C Speech recognition
- ☐ D Summarization
- ☐ E Semantic segmentation

37 Multiple choice · Domain 1

You are comparing articles in a corpus for Northwind Traders.

Common words such as cloud appear in every article. You need to find terms that are frequent in one article but uncommon across the corpus.

Which statistical technique should you use?

- ☐ A Part-of-speech tagging
- ☐ B Term Frequency - Inverse Document Frequency (TF-IDF)
- ☐ C Stop word removal
- ☐ D Simple frequency analysis

38 Multiple choice · Domain 1

You are building App1 for Margie's Travel. Users upload audio recordings of tour guide briefings.

You need App1 to produce written transcripts of the recordings.

Which speech capability should you use?

- ☐ A Speech recognition
- ☐ B Image classification
- ☐ C Speech synthesis
- ☐ D Key phrase extraction

39 Multiple choice · Domain 1

You are building an accessibility feature for Alpine Ski House.

You need App1 to read written emergency alerts aloud to users.

Which speech capability should you use?

- ☐ A Speech synthesis
- ☐ B Speech recognition
- ☐ C Entity detection
- ☐ D Object detection

40 Multiple choice, choose two · Domain 1

You are designing a voice-driven customer service assistant for Litware.

Callers must speak requests by phone.

The assistant must answer with a natural-sounding voice.

You need to identify the speech capabilities required.

Which two capabilities should you identify?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Optical character recognition
- ☐ B Speech recognition
- ☐ C Speech synthesis
- ☐ D TF-IDF scoring
- ☐ E Object detection

41 Multiple choice · Domain 1

You are reviewing a speech synthesis design for Relecloud.

The system must pronounce raw text such as Dr., \$25.50, and 3:00 PM in natural spoken form.

You need to identify the speech synthesis stage that prepares this text for pronunciation.

Which stage should you identify?

- ☐ A Key phrase extraction
- ☐ B Object detection
- ☐ C Beam search decoding
- ☐ D Text normalization

42 Multiple choice · Domain 1

You are building a smart checkout app for Tailspin Toys. The app receives a single product photo at a time. You need the app to predict the main product category shown in each photo.

Which computer vision technique should you use?

- ☐ A Image classification
- ☐ B Speech recognition
- ☐ C Semantic segmentation
- ☐ D Object detection

43 Multiple choice, choose two · Domain 1

You are selecting computer vision techniques for a retail monitoring project at Contoso.

You need to identify techniques that return where objects appear in an image, rather than only the main image label.

Which two techniques should you identify?

Each correct answer presents a complete solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Object detection
- ☐ B Semantic segmentation
- ☐ C Image classification
- ☐ D Speech synthesis
- ☐ E Sentiment analysis

44 Multiple choice · Domain 1

You are designing a marketing tool for Graphic Design Institute.

You need the tool to create a new image from a natural language prompt.

Which model capability should you use?

- ☐ A Object detection
- ☐ B Text-to-image generation
- ☐ C Sentiment analysis
- ☐ D Optical character recognition

45 Multiple choice, choose three · Domain 1

You are reviewing proposed features for an app at Adventure Works.

You need to identify features that are computer vision or image-generation capabilities.

Which three features should you identify?

Each correct answer presents a complete solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Create a product concept image from a text prompt.
 - ☐ B Rank terms by TF-IDF scores across documents.
 - ☐ C Predict a label for the main subject in a photo.
 - ☐ D Detect whether review text is positive or negative.
 - ☐ E Convert a customer call recording to text.
 - ☐ F Find each bicycle in a warehouse image and return bounding boxes.
-

46 Multiple choice · Domain 1

You are building App1 for Lamna Healthcare Company. Users upload photos of printed intake forms.

You need to convert the visible text in each image into editable and searchable text.

Which technique should you use?

- ☐ A Optical character recognition
- ☐ B Sentiment analysis
- ☐ C Image generation
- ☐ D Speech synthesis

47 Multiple choice, choose two · Domain 1

You are designing an invoice processing workflow for Fabrikam.

The workflow must read text from scanned invoices.

The workflow must map values such as vendor name, invoice date, and total to structured fields.

You need to identify the two stages required for the workflow.

Which two stages should you identify?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Prosody generation
- ☐ B Field extraction and mapping
- ☐ C Image classification
- ☐ D Optical character recognition
- ☐ E Text-to-image generation

48 Multiple choice, choose two · Domain 1

You are evaluating information extraction scenarios for Northwind Traders.

You need to identify scenarios that extract structured information from audio or video content.

Which two scenarios should you identify?

Each correct answer presents a complete solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Generate a new logo from a text prompt.
- ☐ B Classify a product review as positive or negative.
- ☐ C Summarize a voicemail and extract the callback number left by the caller.
- ☐ D Analyze a recorded meeting to list speakers, assigned actions, and attendance counts.
- ☐ E Convert written policy text to spoken audio.

49 Multiple choice · Domain 2

Woodgrove Bank is building a chat assistant for employees who ask questions about internal policy. The assistant must use a formal tone, answer concisely, and decline requests that are outside policy support. You need to configure the prompt content that controls the assistant behavior across user requests.

What should you configure?

- ☐ A A user prompt
 - ☐ B A temperature value
 - ☐ C A max output tokens value
 - ☐ D A system prompt
-

50 Multiple choice · Domain 2

Fabrikam wants to use a deployed chat model to prepare a short summary for store managers. You need to write the most effective user prompt for the task.

Which prompt should you use?

- ☐ A Summarize the attached policy changes for store managers in five bullet points.
 - ☐ B Tell me about this policy.
 - ☐ C Set the answer to a higher temperature.
 - ☐ D Create content from the document.
-

51 Multiple choice, choose two · Domain 2

Northwind Traders is creating a chat experience in which users can ask follow-up questions. You need the model to understand each follow-up question in the context of the earlier conversation.

Which two actions should the application perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Include relevant earlier turns or a summary of them in the next prompt.
- ☐ B Remove conversation history after each model response.
- ☐ C Increase max output tokens for every follow-up question.
- ☐ D Replace the system prompt with the previous model response.
- ☐ E Include the new user question in the next prompt.

52 Multiple choice · Domain 2

Tailwind Traders uses retrieval-augmented generation to answer questions from approved product documents.

Some retrieved documents contain text that tries to tell the model to ignore its normal instructions. You need to design the augmented prompt to reduce the effect of those document instructions.

What should the prompt do?

- ☐ A Remove the trusted application instructions from the prompt.
- ☐ B Tell the model to follow any instructions that appear in the retrieved chunks.
- ☐ C Clearly label retrieved chunks as context and tell the model to treat them as data.
- ☐ D Place all retrieved chunks inside the system instructions without labels.

53 Multiple choice, choose three · Domain 2

Adventure Works wants a model to draft customer email replies that match the company style. You need to improve the user prompt for consistent results.

Which three prompt changes should you make?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Specify the desired format and maximum length.
- ☐ B State the audience and purpose of the email.
- ☐ C Set a deployment TPM allocation in the prompt.
- ☐ D Remove all business context from the prompt.
- ☐ E Ask for any answer length the model prefers.
- ☐ F Provide an example of the preferred style.

54 Multiple choice · Domain 2

Litware needs to choose a generative AI model for a customer support prototype in Microsoft Foundry. You need to review model capabilities, benchmarks, and supported tasks before deployment.

What should you use?

- ☐ A A Foundry IQ knowledge source
- ☐ B An Azure Key Vault secret
- ☐ C The Foundry model catalog
- ☐ D An Azure App Service plan

55 Multiple choice · Domain 2

Proseware deploys a model in the Foundry portal with the deployment name chat-prod. You need application requests to route to this deployment.

Which value should the application send in the model parameter?

- ☐ A The model catalog publisher name
- ☐ B The resource group name
- ☐ C The Azure subscription name
- ☐ D The deployment name

56 Multiple choice, choose two · Domain 2

Wide World Importers is using the Foundry portal to create a model deployment. You need to start from the portal model discovery experience and create a deployment.

Which two actions should you perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Select Deploy and choose default or custom deployment settings.
- ☐ B Create a hosted agent container before deploying the model.
- ☐ C Open Discover, select Models, and choose a model.
- ☐ D Create a Foundry IQ knowledge base before selecting a model.
- ☐ E Add an MCP tool before the model can be deployed.

57 Multiple choice · Domain 2

Coho Winery is testing a deployed chat model before adding it to an application. A developer wants to adjust temperature, max output tokens, and system instructions while trying prompts. You need to identify where the developer should work.

What should the developer use?

- ☐ A The Azure role assignments page
- ☐ B The Foundry Playgrounds
- ☐ C The Azure subscription cost analysis page
- ☐ D The Foundry IQ source configuration page

58 Multiple choice · Domain 2

Fourth Coffee tests a deployed model and sees throttling when prompts request very long responses. You need to reduce the token load sent to the deployment from the client or playground.

Which setting should you lower?

- ☐ A Max output tokens
- ☐ B The model catalog benchmark score
- ☐ C System instructions
- ☐ D Temperature

59 Multiple choice · Domain 2

Margie's Travel is building a lightweight Python chat client for a Foundry project. You need to obtain the OpenAI-compatible client from the Foundry project client. The client uses the Foundry SDK for Python.

```
from azure.identity import DefaultAzureCredential
from azure.ai.projects import AIProjectClient
project = AIProjectClient(endpoint=endpoint, credential=DefaultAzureCredential())
openai = project.TODO()
response = openai.responses.create(model=deployment, input="Plan a weekend trip.")
```

Which method should replace TODO?

- ☐ A conversations.create
- ☐ B deployments.list
- ☐ C get_openai_client
- ☐ D agents.create_version

60 Multiple choice, choose two · Domain 2

Alpine Ski House is configuring a Python chat client that connects to a Foundry project.
You need to configure the endpoint and authentication correctly.

Which two statements are correct?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Use DefaultAzureCredential to authenticate with Microsoft Entra ID.
- ☐ B Use the resource group name as the endpoint host name.
- ☐ C Use the project endpoint in the form `https://<resource>.services.ai.azure.com/api/projects/<project>`.
- ☐ D Use the Azure portal URL as the project endpoint.
- ☐ E Use only an API key because the Foundry SDK does not support Microsoft Entra ID.

61 Multiple choice · Domain 2

Lucerne Publishing is completing a Python chat client that sends a prompt to a deployed model.
You need the request to identify the model deployment to use.
The code calls the Responses API through the OpenAI-compatible client.

```
response = openai.responses.create(  
    TODO="gpt-5-mini",  
    input="Summarize this article in three bullets."  
)  
print(response.output_text)
```

Which parameter should replace TODO?

- ☐ A api_key
- ☐ B project
- ☐ C endpoint
- ☐ D model

62 Multiple choice · Domain 2

Graphic Design Institute wants a lightweight chat client instead of a complex custom application. You need to identify the main responsibilities of the lightweight client.

Which three responsibilities should you include?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Display the returned results.
- ☐ B Run GPU inference on the client device.
- ☐ C Collect user input.
- ☐ D Call a remote service or API.
- ☐ E Train the foundation model locally.
- ☐ F Replace the Foundry project endpoint.

63 Multiple choice · Domain 2

Contoso is deciding whether to write raw REST requests or use a language-specific helper for a small chat client.

You need to use language-specific code libraries that abstract the REST calls to the service.

What should Contoso use?

- ☐ A A resource group deployment script
- ☐ B A software development kit
- ☐ C A model benchmark table
- ☐ D A Foundry IQ knowledge source

64 Multiple choice · Domain 2

Lamna Healthcare Company is creating Agent1 in the Foundry portal.

Agent1 must answer in a calm tone, ask clarifying questions, and avoid giving medical advice.

You need to configure the part of the agent that defines its role and behavior.

What should you configure?

- ☐ A A knowledge source location
- ☐ B A quota limit
- ☐ C Instructions
- ☐ D A model benchmark

65 Multiple choice, choose two · Domain 2

Relecloud creates an expense policy knowledge base in Foundry IQ and wants Agent1 to use it in the Foundry portal.

You need Agent1 to answer policy questions with grounded responses and citations.

Which two actions should you perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Write instructions that tell the agent when to use the knowledge base and cite sources.
 - ☐ B Disable citations so the agent gives shorter answers.
 - ☐ C Add the knowledge base connection to the agent as an MCP tool.
 - ☐ D Create a separate model deployment for each policy document.
 - ☐ E Increase temperature so the agent uses more creative answers.
-

66 Multiple choice · Domain 2

Tailspin Toys connects Agent1 to a knowledge base and tests it in the Foundry portal playground.

You need a test set that can reveal grounding, permission, and missing-evidence issues.

Which set of prompts should you use?

- ☐ A Only creative writing prompts with high temperature settings
 - ☐ B Only greetings and small-talk prompts that do not require the knowledge base
 - ☐ C Only model benchmark prompts from the model catalog
 - ☐ D Realistic questions with clear answers, ambiguous cases, missing-answer cases, and permission variations
-

67 Multiple choice · Domain 2

Northwind Traders wants the fastest portal-first way to build a text agent without maintaining runtime code or containers.

You need to choose the agent type to create in Foundry Agent Service.

Which agent type should you choose?

- ☐ A A voice-based prompt agent
- ☐ B An ephemeral agent in application code
- ☐ C A prompt agent
- ☐ D A hosted agent

68 Multiple choice · Domain 2

Fabrikam is building a Python client that calls an existing prompt agent named Agent1.

You need to create an OpenAI-compatible client that is bound to Agent1.

The code uses AIProjectClient from the Foundry SDK.

```
from azure.ai.projects import AIProjectClient
from azure.identity import DefaultAzureCredential
project = AIProjectClient(
    endpoint=endpoint,
    credential=DefaultAzureCredential(),
    allow_preview=True,
)
openai = project.get_openai_client(TODO)
response = openai.responses.create(input="What can you help with?")
```

Which argument should replace TODO?

- ☐ A model="Agent1"
- ☐ B api_key="Agent1"
- ☐ C conversation="Agent1"
- ☐ D agent_name="Agent1"

69 Multiple choice, choose two · Domain 2

Proseware is building a lightweight client for Agent1. Users must be able to ask a question and then ask a related follow-up question.

You need the agent conversation to maintain history across both turns.

Which two actions should the client perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Pass conversation.id in each responses.create call for the thread.
- ☐ B Create a new agent version for every user message.
- ☐ C Put the follow-up question in the model parameter.
- ☐ D Remove the agent instructions after the first response.
- ☐ E Create a conversation by calling openai.conversations.create.

70 Multiple choice · Domain 2

Woodgrove Bank is finishing a Python client that invokes Agent1 through an agent-scoped OpenAI-compatible client.

You need to print the generated text returned by the agent.

The client sends a user message to the agent.

```
conversation = openai.conversations.create()
response = openai.responses.create(
    conversation=conversation.id,
    input="Explain the expense policy."
)
```

Which expression should you print?

- ☐ A response.id
- ☐ B response.output_text
- ☐ C project.endpoint
- ☐ D conversation.id

71 Multiple choice, choose two · Domain 2

Tailwind Traders wants to configure constants for a lightweight Python client that calls an existing prompt agent.

You need to configure the minimum Foundry-specific values that the client code uses.

Which two values should you configure?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A The Azure portal home page URL
- ☐ B The Azure subscription display name
- ☐ C The prompt agent name
- ☐ D The TPM quota request number
- ☐ E The model catalog publisher name
- ☐ F The Foundry project endpoint

72 Multiple choice · Domain 2

Contoso is building App1 to analyze support messages by using Azure Language in Foundry Tools.

App1 must authenticate with the endpoint and key for a Microsoft Foundry resource.

You need to create the client object in Python for the Azure Language SDK.

Which client should you create?

- ☐ A An AIProjectClient that uses the project endpoint
- ☐ B A TextAnalyticsClient that uses AzureKeyCredential
- ☐ C A SpeechRecognizer that uses AudioConfig
- ☐ D An AzureOpenAI client that uses an audio deployment

73 Multiple choice · Domain 2

Fabrikam stores customer chat messages in multiple languages.

The app must route each message to a language-specific queue before any other processing occurs.

You need to identify the primary language and confidence score for each submitted document.

Which TextAnalyticsClient method should you call?

- ☐ A detect_language
- ☐ B begin_analyze_healthcare_entities
- ☐ C recognize_pii_entities
- ☐ D recognize_entities

74 Multiple choice, choose two · Domain 2

Northwind Traders is building App1 to process free-text customer feedback.

App1 must identify personal information and store a masked version of each message.

You need to implement the Azure Language SDK call and use the relevant response data.

Which two implementation details should you use?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Call recognize_entities with the documents to analyze.
- ☐ B Read redacted_text and the PII entity categories from the result.
- ☐ C Read primary_language.iso6391_name from the result.
- ☐ D Call recognize_pii_entities with the documents to analyze.
- ☐ E Call detect_language before storing the masked text.
- ☐ F Send the feedback to a SpeechRecognizer object.

75 Multiple choice · Domain 2

Tailwind Traders has an agent named Agent1 in the Foundry portal.

Agent1 must call deterministic text analysis tools for language detection and PII redaction without custom REST integration code.

You need to connect the appropriate tool to Agent1.

What should you add to Agent1?

- ☐ A Azure Speech in Foundry Tools as a Voice Live session
- ☐ B Azure Language in Foundry Tools as an MCP server
- ☐ C Foundry IQ as a knowledge base
- ☐ D An image generation model from the model catalog

76 Multiple choice, choose two · Domain 2

Adventure Works is building an automated compliance workflow for short text messages.

The workflow must return consistent fields for language code, confidence score, and redacted personal information.

You need to justify using Azure Language in Foundry Tools instead of a general-purpose model prompt.

Which two reasons support this choice?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A The Speech SDK provides the text redaction methods.
- ☐ B The SDK includes methods such as detect_language and recognize_pii_entities.
- ☐ C The OpenAI Responses API always returns deterministic structured fields.
- ☐ D Azure Language returns structured results for purpose-built analysis tasks.
- ☐ E Azure Language requires a deployed model name for every request.

77 Multiple choice · Domain 2

Wide World Importers deployed a speech-capable model in a Microsoft Foundry project.

A Python app must upload a recorded spoken prompt and receive a text transcript from the deployed model.

You need to choose the model type for the deployment.

Which model type should you use?

- ☐ A gpt-4o-transcribe
- ☐ B gpt-image-1
- ☐ C gpt-4o-mini-tts
- ☐ D text-embedding-3-small

78 Multiple choice, choose two · Domain 2

Litware is building App1 to transcribe short audio prompts by using a deployed speech-to-text model in Microsoft Foundry.

App1 must use the OpenAI Python SDK.

You need to implement the request that sends the audio file to the model.

Which two actions should you perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Call `client.responses.create` with only a text input string.
- ☐ B Create an `AzureOpenAI` client with the Foundry endpoint, key, and API version.
- ☐ C Set a Speech Synthesis Markup Language voice element.
- ☐ D Set `speech_synthesis_voice_name` on the speech configuration.
- ☐ E Create a `SpeechRecognizer` with an `AudioConfig` object.
- ☐ F Open the audio file in binary mode before sending the request.

79 Multiple choice · Domain 2

Alpine Ski House wants App1 to answer users with generated audio after it determines the response text.

App1 must use a deployed text-to-speech model in Microsoft Foundry by using the OpenAI Python SDK.

You need to call the appropriate client method.

Which method should you call?

- ☐ A `speech_recognizer.recognize_once_async`
- ☐ B `client.audio.speech.with_streaming_response.create`
- ☐ C `client.recognize_entities`
- ☐ D `client.audio.transcriptions.create`

80 Multiple choice · Domain 2

Graphic Design Institute is selecting a Microsoft Foundry model for a prototype that accepts spoken prompts and returns spoken answers.

The prototype must use a deployed model that supports low-latency speech-in and speech-out conversational interactions.

You need to identify the model capability to look for in the catalog.

Which capability should the model support?

- ☐ A Named entity recognition for written documents
- ☐ B Realtime audio processing for speech-in and speech-out interactions
- ☐ C Batch image generation from text prompts
- ☐ D Offline embedding generation for text similarity

81 Multiple choice · Domain 2

Woodgrove Bank is building App1 to transcribe recorded branch voicemail files.

App1 must use Azure Speech in Foundry Tools with the Speech SDK.

You need to invoke speech recognition for one audio file.

Which object and method should you use?

- ☐ A SpeechRecognizer and recognize_once_async
- ☐ B SpeechSynthesizer and speak_text_async
- ☐ C AzureOpenAI and audio.speech.create
- ☐ D TextAnalyticsClient and recognize_entities

82 Multiple choice, choose two · Domain 2

Fourth Coffee has a Python app that transcribes local WAV files by using Azure Speech in Foundry Tools. The app must connect to a Foundry resource and use the file as the audio source. You need to configure the SDK objects before creating the recognizer.

Which two objects should you configure?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A TextAnalyticsClient with an AzureKeyCredential
- ☐ B AIProjectClient with only the project endpoint
- ☐ C SpeechConfig with the Foundry resource key and endpoint
- ☐ D SpeechSynthesizer with a voice name
- ☐ E AudioConfig with the audio file name
- ☐ F AudioOutputConfig with the default speaker

83 Multiple choice · Domain 2

Lamna Healthcare Company is building App1 to read appointment reminders aloud. App1 must convert a plain text reminder to spoken audio by using Azure Speech in Foundry Tools. You need to call the correct Speech SDK method for plain text input.

Which method should you call?

- ☐ A speech_synthesizer.speak_ssml_async
- ☐ B speech_synthesizer.speak_text_async
- ☐ C client.detect_language
- ☐ D speech_recognizer.recognize_once_async

84 Multiple choice, choose three · Domain 2

Coho Winery is building App1 to create audio tours for visitors.

The narration must pause between sections, pronounce the product code SQL as sequel, and speak one sentence more slowly.

You need to choose SSML features for the text-to-speech request.

Which three SSML features should you use?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Adjust prosody for speaking rate.
- ☐ B Insert a break or silence element.
- ☐ C Train a custom speech recognition model.
- ☐ D Specify a phoneme for pronunciation.
- ☐ E Call detect_language on the narration text.
- ☐ F Call recognize_pii_entities on the tour script.
- ☐ G Run batch transcription on the audio file.

85 Multiple choice · Domain 2

Relecloud is building App1 for a kiosk that conducts natural spoken conversations with visitors.

The solution must provide real-time bidirectional audio and reduce the need to orchestrate separate speech recognition, model reasoning, and text-to-speech components.

You need to choose the Azure Speech in Foundry Tools feature to use.

What should you use?

- ☐ A SpeechSynthesizer
- ☐ B SpeechRecognizer
- ☐ C Azure Language MCP server
- ☐ D Voice Live API

86 Multiple choice · Domain 2

Contoso is building a support chat app that answers questions about customer-uploaded damage photos. You need to select a Foundry model capability for the app.

What should you deploy?

- ☐ A An image generation model deployment
 - ☐ B A vision-enabled chat model deployment
 - ☐ C A text embedding model deployment
 - ☐ D A speech synthesis model deployment
-

87 Multiple choice · Domain 2

Fabrikam is building a Python app that sends a local product photo to a vision-enabled chat model. The image is not available from a public URL. You need to include the image in the request body.

What should the app do?

- ☐ A Place the local file path in the text message content
 - ☐ B Convert the image to an embedding vector and send the vector
 - ☐ C Convert the image to a base64 data URL and place it in the image URL field
 - ☐ D Upload the image as the system message content
-

88 Multiple choice, choose two · Domain 2

Northwind Traders is calling the Chat Completion API with a vision-enabled deployment to analyze a shelf photo.

The response must include a complete answer to the user's question.

You need to build the request correctly.

Which two actions should you perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Include both a text item and an image_url item in the user message content array
- ☐ B Set n_seconds to control the length of the response
- ☐ C Use the image generation endpoint instead of chat completions
- ☐ D Use a text embedding model to convert the photo to tokens
- ☐ E Set max_tokens or max_completion_tokens to allow enough response tokens

89 Multiple choice · Domain 2

Tailwind Traders sends photos of printed labels to a vision-enabled chat model.

Small text in the image must be interpreted as accurately as possible, even if the request uses more tokens.

You need to configure how the model processes the image.

Which parameter should you set?

- ☐ A Set image generation quality to high
- ☐ B Set n to a value greater than 1
- ☐ C Set output_format to png
- ☐ D Set image_url.detail to high

90 Multiple choice · Domain 2

Adventure Works wants a Foundry app that generates promotional images from text prompts.

The app must use a current OpenAI image generation model available in Foundry.

You need to choose a model family to deploy.

Which model family should you use?

- ☐ A DALL-E 3
- ☐ B GPT-4.1 text chat
- ☐ C GPT-Image series
- ☐ D Whisper

91 Multiple choice, choose two · Domain 2

Tailspin Toys uses a GPT-image model deployment to generate artwork through the image generation API.

The app must request four images at 1024x1024 pixels.

You need to set the request parameters.

Which two parameters should you set?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A temperature
- ☐ B n
- ☐ C max_tokens
- ☐ D n_seconds
- ☐ E size
- ☐ F detail

92 Multiple choice · Domain 2

Woodgrove Bank calls a GPT-image model from a Python app.

The app must save the generated image to a file.

You need to handle a successful image generation response.

What should the app do?

- ☐ A Extract bounding box coordinates from the first choice
- ☐ B Poll a video generation job until it returns a download URL
- ☐ C Decode the b64_json image data and write the bytes to an image file
- ☐ D Read response.output_text and save it as a PNG file

93 Multiple choice, choose three · Domain 2

Graphic Design Institute has a GPT-image-1 deployment for creating simple product icons.

The generated icons must have transparent backgrounds and be saved as PNG files.

You need to configure the image generation request.

Which three actions should you perform?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A Use a Sora 2 deployment
- ☐ B Set output_format to png
- ☐ C Set background to transparent
- ☐ D Call the image generation API with a prompt for the icon
- ☐ E Set image_url.detail to high
- ☐ F Set n_seconds to 5

94 Multiple choice · Domain 2

Proseware is building a Foundry app that creates short marketing videos from text prompts.

The app must generate an MP4 video programmatically.

You need to implement video generation.

Which two actions should you perform?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Deploy a Sora 2 model with the Azure OpenAI resource
 - ☐ B Use a vision-enabled chat completion request with an image URL
 - ☐ C Decode b64_json from an image generation response
 - ☐ D Create a video generation job, poll its status, and retrieve the video content after success
 - ☐ E Deploy a GPT-Image model and call image edits
-

95 Multiple choice · Domain 2

Lamna Healthcare Company is prototyping a Python app that lets a clinician upload an image and ask a question about it.

You need to send the user input to a deployed vision-capable model by using the OpenAI Responses API.

What should the app send?

- ☐ A A speech synthesis request with voice and audio format
 - ☐ B A user input that contains only the image file name
 - ☐ C A user input that contains input_text and input_image content items
 - ☐ D A generation request with n, size, and quality
-

96 Multiple choice · Domain 2

Wide World Importers sends image URLs from a storage account to a vision-enabled chat model.

The storage account is restricted by a private endpoint and firewall.

You need the model to process the image successfully.

What should you do?

- ☐ A Set image_url.detail to auto
- ☐ B Keep the private URL and increase max_tokens
- ☐ C Provide a publicly accessible image URL or send the image as base64 data
- ☐ D Move the prompt text to the system message

97 Multiple choice, choose two · Domain 2

Litware is building a keyless Python app for image analysis with a deployed vision-enabled model.

The Microsoft Entra role assignment for the developer identity is already complete.

You need to configure the app to call the deployment.

Which two values should you configure in the app?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A The Sora video job ID before submitting a request
 - ☐ B The Azure OpenAI endpoint
 - ☐ C The model deployment name
 - ☐ D The generated image b64_json value
 - ☐ E The Azure OpenAI API key
-

98 Multiple choice · Domain 2

Coho Winery is building an app that sends vineyard photos and user questions to a vision-enabled chat model.

You need to design how the app handles the model result.

What should the app expect to receive?

- ☐ A A required set of object detection bounding boxes
 - ☐ B An MP4 video download created from the prompt
 - ☐ C A natural language response based on the text and image inputs
 - ☐ D A base64-encoded generated image
-

99 Multiple choice · Domain 2

You have a Microsoft Foundry project named Project1.

You are building App1 to process supplier invoices that use common invoice fields and line items.

You need to extract vendor, invoice date, total, and line item fields with the least custom configuration.

Which analyzer should you use?

- ☐ A prebuilt-imageSearch
- ☐ B prebuilt-audioSearch
- ☐ C prebuilt-documentSearch
- ☐ D prebuilt-invoice

100 Multiple choice · Domain 2

Tailwind Traders creates a custom document analyzer for purchase request forms.

Each form includes a repeating list of products, quantities, and prices.

You need the analyzer result to return each product as a structured child record.

Which field definition should you use?

- ☐ A A classify field with enum values for every product name
 - ☐ B A boolean field for each possible line item
 - ☐ C A string field whose description asks for comma-separated product values
 - ☐ D An array field with object items and properties for product, quantity, and price
-

101 Multiple choice · Domain 2

Woodgrove Bank uses a custom document analyzer to process loan forms.

If a field is low confidence, a reviewer must see the source location in the original document.

You need the analyzer to return confidence scores and grounding for field values.

Which setting should you enable?

- ☐ A tableFormat
- ☐ B enableFormula
- ☐ C omitContent
- ☐ D estimateFieldSourceAndConfidence

102 Multiple choice, choose two · Domain 2

Fabrikam creates a custom document analyzer based on prebuilt-document.

The analyzer configuration already enables confidence and source estimation for fields.

The analyzer must return the exact receipt total when it appears on a form.

The analyzer must classify the document as Receipt or Exception.

You need to define the field schema.

Which two field definitions should you add?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A A Summary field with type string and method generate
- ☐ B A tableFormat field set to markdown inside fieldSchema
- ☐ C A DocumentKind field with type string, method classify, and enum values Receipt and Exception
- ☐ D A ReceiptTotal field with type boolean and no description
- ☐ E A ReceiptTotal field with type number and method extract
- ☐ F A baseAnalyzerId field set to prebuilt-audio

103 Multiple choice · Domain 2

Northwind Traders stores product photos in a catalog.

You need a prebuilt Content Understanding analyzer that can generate descriptions and insights for the images so they can be indexed for search.

Which analyzer should you use?

- ☐ A prebuilt-layout
- ☐ B prebuilt-invoice
- ☐ C prebuilt-imageSearch
- ☐ D prebuilt-audioSearch

104 Multiple choice · Domain 2

Adventure Works receives smartphone photos of completed warranty forms.

Most required values are printed text next to field labels, such as Serial Number and Purchase Date.

You need to choose the most appropriate Content Understanding approach for extracting the form fields from the photos.

Which approach should you use?

- ☐ A Use a document analyzer with a field extraction schema for the labeled form values
- ☐ B Use prebuilt-videoSearch to split the form into scenes
- ☐ C Use prebuilt-imageSearch to generate only a one-paragraph image description
- ☐ D Use prebuilt-audioSearch to create a transcript from the image

105 Multiple choice, choose two · Domain 2

Litware creates a custom Content Understanding analyzer for chart images.

The result must include the chart title.

The result must classify the chart type as bar, line, or pie.

You need to define the field schema for the analyzer.

Which two fields should you include?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A A Caller field with type string
- ☐ B A KeyFrameTimes field with type array for video frames
- ☐ C A VendorName field with method extract for invoice headers
- ☐ D A ChartType field with type string, method classify, and enum values bar, line, and pie
- ☐ E A Title field with type string
- ☐ F A Transcript field with type string and method extract

106 Multiple choice · Domain 2

Contoso records customer support calls.

App1 must return a transcript with agent and customer roles, a call summary, sentiment, mentioned people, and categories.

You need to choose a prebuilt audio analyzer for post-call analysis.

Which analyzer should you use?

- ☐ A prebuilt-audioSearch
- ☐ B prebuilt-imageSearch
- ☐ C prebuilt-videoSearch
- ☐ D prebuilt-callCenter

107 Multiple choice · Domain 2

Relecloud builds an app that analyzes recorded support calls.

Calls are expected to be in en-US or fr-FR, and the app should avoid relying on automatic multilingual transcription when the expected locales are known.

You need to configure the Content Understanding analyzer for transcription quality.

Which configuration option should you set?

- ☐ A estimateFieldSourceAndConfidence
- ☐ B tableFormat
- ☐ C locales
- ☐ D segmentPerPage

108 Multiple choice, choose two · Domain 2

Alpine Ski House analyzes training videos by using prebuilt-videoSearch.

The app must let users find spoken instructions at precise times and inspect representative visuals from the same video.

You need to decide which returned outputs App1 should use.

Which two outputs should App1 use?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A Audio locale configuration options
 - ☐ B Image chart type enum values
 - ☐ C Document tableFormat settings
 - ☐ D Key frame times and key frame images
 - ☐ E Invoice line item fields
 - ☐ F Transcript phrases with timestamps
-

109 Multiple choice, choose three · Domain 2

Fourth Coffee creates a custom analyzer for product demo videos.

The analyzer must process video content, split each video into logical product demo segments, and classify the sentiment for each segment.

You need to configure the analyzer schema.

Which three configuration elements should you include?

Each correct answer presents part of the solution. Choose three.

NOTE: Each correct selection is worth one point.

- ☐ A tableFormat set to html
- ☐ B baseAnalyzerId set to prebuilt-document
- ☐ C prebuilt-invoice as the analyzer ID for the video files
- ☐ D enableSegment set to true with contentCategories that describe the desired segments
- ☐ E A Sentiment field with method classify and allowed sentiment values
- ☐ F baseAnalyzerId set to prebuilt-video
- ☐ G estimateFieldSourceAndConfidence set to true for document fields

110 Multiple choice · Domain 2

You are building a REST client named App1 for Azure Content Understanding in Foundry Tools. App1 sends an analyze request and receives a 202 Accepted response with an Operation-Location header. You need App1 to retrieve the completed analysis result.

What should App1 do next?

- ☐ A Submit the same file to a chat completions endpoint
 - ☐ B Read all extracted fields from the 202 response body without polling
 - ☐ C Send a PUT request to recreate the analyzer before every polling attempt
 - ☐ D Send GET requests to the Operation-Location URL until the status is Succeeded
-

111 Multiple choice · Domain 2

Proseware builds a Python client app that uses a prebuilt Content Understanding analyzer. You need to start an asynchronous analysis operation for a file URL by using the Content Understanding Python SDK.

Which client method should you call?

- ☐ A poller.result
 - ☐ B begin_create_analyzer
 - ☐ C get_analyzer
 - ☐ D begin_analyze
-

112 Multiple choice, choose two · Domain 2

Wide World Importers builds a lightweight REST app that uses prebuilt Content Understanding analyzers. The app already has the Foundry resource endpoint and subscription key. You need to submit a file for analysis by using the REST API.

Which two elements should the app include to identify the analyzer and the content to analyze?

Each correct answer presents part of the solution. Choose two.

NOTE: Each correct selection is worth one point.

- ☐ A The analyzer ID in the analyze request URL
- ☐ B A result status value of Succeeded in the request body
- ☐ C A tableFormat value for every media type
- ☐ D An inputs array in the JSON body that contains the file URL
- ☐ E An Operation-Location header in the initial request
- ☐ F A key frame URL before the video is analyzed

Part 2: Answers and explanations

A quick answer key first, then the reasoning for every question.

1	D	2	B	3	B	4	C	5	C	6	AE	7	C	8	BD
9	B	10	CD	11	A	12	AB	13	CDE	14	A	15	A	16	B
17	C	18	D	19	AE	20	D	21	C	22	A	23	AE	24	CDG
25	D	26	A	27	B	28	AC	29	AD	30	D	31	A	32	ABF
33	C	34	C	35	A	36	BD	37	B	38	A	39	A	40	BC
41	D	42	A	43	AB	44	B	45	ACF	46	A	47	BD	48	CD
49	D	50	A	51	AE	52	C	53	ABF	54	C	55	D	56	AC
57	B	58	A	59	C	60	AC	61	D	62	ACD	63	B	64	C
65	AC	66	D	67	C	68	D	69	AE	70	B	71	CF	72	B
73	A	74	BD	75	B	76	BD	77	A	78	BF	79	B	80	B
81	A	82	CE	83	B	84	ABD	85	D	86	B	87	C	88	AE
89	D	90	C	91	BE	92	C	93	BCD	94	AD	95	C	96	C
97	BC	98	C	99	D	100	D	101	D	102	CE	103	C	104	A
105	DE	106	D	107	C	108	DF	109	DEF	110	D	111	D	112	AD

1 Multiple choice · Identify AI concepts and capabilities

ANSWER: D

D Fairness

Fairness requires AI systems to avoid affecting similar groups differently and to minimize discriminatory outputs from biased data or design. The decisive detail is that patients with similar clinical factors receive different recommendations based on region. This is a fairness concern rather than an inclusiveness concern because the issue is unequal outcomes, not access to the system.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Inclusiveness focuses on avoiding exclusion from using or benefiting from the solution, not unequal recommendations for similar cases.
- **B.** Transparency focuses on explaining how the system works and its limitations, not unequal treatment across comparable groups.
- **C.** Accountability focuses on governance and responsibility for AI systems, not measuring differences in outcomes across groups.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [What is Responsible AI - Azure Machine Learning](#)

2 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

- B** Evaluate outcomes across relevant groups and adjust the training data to reduce bias.

Microsoft Learn describes fairness as minimizing bias in training data and testing AI systems for fair treatment. Screening job applicants is a high-impact scenario where similar qualifications should receive similar recommendations. Evaluating outcomes across relevant groups and adjusting data is the action most directly aligned with fairness.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A text-only interface supports inclusiveness by improving access, but it does not directly address biased hiring recommendations.
- **C.** Encryption and access control support privacy and security, but they do not directly test or reduce discriminatory outcomes.
- **D.** A user guide supports transparency, but it does not directly minimize bias in training data or model outcomes.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [What is Responsible AI - Azure Machine Learning](#)

3 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

- B** Require a minimum confidence score before the robot moves the object.

Reliability and safety require AI applications to account for probabilistic behavior and mitigate risks. The responsible AI unit gives the example of using confidence values before a vision-based robot interacts with physical objects. A confidence threshold helps the system fail safely when uncertainty is high.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Describing the training dataset supports transparency, but it does not directly prevent unsafe physical actions.
- **C.** Deleting temporary images supports privacy and security, but it does not directly control the robot when confidence is low.
- **D.** Translating the interface supports inclusiveness, but it does not make the physical action safer.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [What is Responsible AI - Azure Machine Learning](#)

4 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

- C** Content filters

Content filters are guardrails that detect and block harmful, illegal, or offensive content in prompts and responses. Mitigating the risk of harmful output before and after deployment is how an AI solution supports the reliability and safety principle.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Model benchmarks help compare models, but they do not filter harmful generated content at runtime.
- **B.** Model interpretability explains model predictions, but it is not the control that filters harmful generated content.
- **D.** Personal data filtering helps detect or block personal information in outputs, not harmful or offensive content in general.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#)

5 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

- C** Privacy and security

Privacy and security require protecting training, input, and operational data that can include personal information. The responsible AI unit gives a similar facial identification example that deletes temporary images and prevents unnecessary access. The decisive detail is data protection, not whether the system treats groups equally or explains its decisions.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Transparency concerns disclosure and explanations, not protecting and deleting personal data.
- **B.** Fairness concerns discriminatory outcomes or unequal treatment, not retention and access controls for personal images.
- **D.** Inclusiveness concerns avoiding exclusion of users, not controlling access to biometric data.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [What is Responsible AI - Azure Machine Learning](#)

6 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: A, E

- A** Restrict access to resources and operations by user or group.
- E** Enable the personal data filter for model outputs.

Privacy and security require protecting personal and business information in AI systems. Microsoft Learn describes access restrictions as a privacy and security control, and Foundry personal data filtering can flag or block outputs that contain known personal information. Together, these controls reduce the chance that sensitive customer data is exposed.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Text captions support inclusiveness for users who cannot use audio, but they do not reduce data exposure in responses.
- **C.** A limitations page supports transparency, but it does not directly prevent sensitive data from being exposed.
- **D.** Evaluating approval rates supports fairness, but it does not directly protect customer data in generated content.

Microsoft Learn: [What is Responsible AI - Azure Machine Learning](#) · [Personally Identifiable Information \(PII\) Filter](#)

7 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

- C** Display text captions for spoken responses.

Inclusiveness means AI benefits should be open to everyone and solutions should not exclude some users. The responsible AI unit gives the example of adding text captions to a speech-based AI agent for users with hearing impairments. Captions address the access barrier described in the scenario.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Testing predictions across groups supports fairness, but it does not address access for users who cannot rely on audio.
- **B.** Tracking deployment approvals supports accountability, but it does not reduce exclusion from the user experience.
- **D.** Encrypting recordings supports privacy and security, but it does not make the kiosk usable for users with hearing impairments.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#)

8 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: B, D

- B** Provide a text-based path in addition to voice interaction.
- D** Test the experience with users who have accessibility needs.

Inclusiveness focuses on ensuring AI solutions do not exclude some users. Providing an alternative to voice interaction and testing with users who have accessibility needs both reduce access barriers. This differs from fairness because the scenario is about who can use the assistant, not whether similar groups receive different outcomes.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Hiding the AI role conflicts with transparency and does not make the assistant more inclusive.
- **C.** Limiting support to the newest browser can exclude users and is the opposite of the stated goal.
- **E.** Collecting more passport data increases privacy risk and does not improve access for users with different needs.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [What is Responsible AI - Azure Machine Learning](#)

9 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

- B** Disclose the AI role and describe the assistant's capabilities and limitations.

Transparency means making users aware of how an AI system works and its potential limitations. The responsible AI unit gives a bank loan scenario where the organization discloses AI use and describes features of the training data without revealing confidential information. The scenario asks for communication that helps customers understand the system.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Evaluating recommendations across groups supports fairness, but it does not directly inform customers how the assistant works.
- **C.** A confidence threshold supports reliability and safety, but it does not disclose the AI role or limitations.
- **D.** Restricting records supports privacy and security, but it does not explain the system to customers.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [Transparency Note for Azure OpenAI in Microsoft Foundry Models](#)

10 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: C, D

- C** Provide documentation that explains App1 capabilities and limitations.
- D** Show citations to the policy sources used in grounded answers.

Transparency requires helping people understand how an AI system works, what its limitations are, and how to use its output appropriately. Microsoft Learn recommends providing system documentation and citing information sources when generated content is based on references. Those measures help employees judge and verify App1 responses.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Evaluating demographic groups supports fairness, but it does not provide explanations or citations to users.
- **B.** Encryption supports privacy and security, but it does not help employees understand or verify the AI output.
- **E.** Blocking harmful content supports safety, but it does not explain capabilities, limitations, or answer sources.

Microsoft Learn: [Overview of Responsible AI practices for Azure OpenAI in Foundry Models](#) · [Transparency Note for Azure OpenAI in Microsoft Foundry Models](#)

11 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

- A** Accountability

Accountability means the people and organizations that develop and distribute AI systems remain responsible for how those systems operate. Microsoft Learn describes governance frameworks and lifecycle metadata, including who published a model, why changes were made, and when models were deployed or used. The practice in the scenario creates accountability evidence.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Transparency focuses on helping users understand the system and its limitations, not tracking governance responsibility for deployments.
- **C.** Inclusiveness focuses on avoiding exclusion from the solution, not deployment lineage and governance.
- **D.** Fairness focuses on avoiding discriminatory outcomes, not recording who approved model changes.

Microsoft Learn: [Introduction to AI concepts: Responsible AI](#) · [What is Responsible AI - Azure Machine Learning](#)

12 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: A, B

- A** Capture lineage metadata for model publishing and deployment.
- B** Monitor deployed applications and alert on lifecycle events.

Accountability requires governance practices that show how AI systems are developed, changed, deployed, and operated. Microsoft Learn lists lineage metadata and monitoring or alerts as MLOps capabilities that improve accountability. These practices provide reviewable evidence for both lifecycle decisions and production operation.

WHY THE OTHER CHOICES ARE WRONG

- **C.** Reducing demographic outcome differences supports fairness, but it does not directly create accountability records.
- **D.** Captions support inclusiveness, but they do not provide lifecycle governance evidence.
- **E.** Blocking customer identifiers supports privacy and security, but it does not document model decisions or operations.

Microsoft Learn: [What is Responsible AI - Azure Machine Learning](#)

13 Multiple choice, choose three · Identify AI concepts and capabilities

ANSWERS: C, D, E

- C** Identify harms by using red team testing and stress testing.
- D** Apply guardrails and then repeat measurement to test effectiveness.
- E** Define metrics and test sets to measure harm frequency and severity.

Responsible AI practices for Azure OpenAI in Foundry Models use an iterative lifecycle of identifying, measuring, mitigating, and operating. Red teaming and stress testing help identify harms, metrics and test sets help measure them, and guardrails are a safety system mitigation. Repeating measurement after mitigation helps verify that the controls are effective for the scenario.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Disabling harmful content controls increases safety risk and is not a responsible mitigation approach.
- **B.** Waiting for production incidents ignores the need to identify and measure risks before deployment.
- **F.** Using customer prompts to train foundation models without permission conflicts with Microsoft data privacy commitments.

Microsoft Learn: [Overview of Responsible AI practices for Azure OpenAI in Foundry Models](#) · [Guardrails and controls overview in Microsoft Foundry](#) · [Responsible AI for Microsoft Foundry](#)

14 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

- A** Customer prompts and completions are not available to other customers or model providers.

Privacy and security include protecting customer data and preventing unauthorized disclosure. The data, privacy, and security documentation says prompts and completions are not available to other customers or model providers and are not used to train foundation models without permission or instruction. It also states that fine-tuned models are exclusive to the customer.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Microsoft Learn states that models are stateless for inferencing and prompts and completions are not stored in the model.
- **C.** Microsoft Learn states that fine-tuned models are available exclusively for the customer's use.
- **D.** Microsoft Learn states that prompts and completions are not available to other customers and are not used to improve products without permission or instruction.

Microsoft Learn: [Data, privacy, and security for Foundry Models sold by Azure in Microsoft Foundry](#)

15 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

- A** Words, subwords, punctuation, or other common character sequences

Large language models break training text into tokens and assign each token a unique identifier. Tokens can represent words, subwords, punctuation, and other commonly used character sequences.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Tokens can include complete words, but they can also include subwords, punctuation, and other sequences.
- **C.** A sentence is made from multiple tokens rather than being the only unit in the vocabulary.
- **D.** Token identifiers represent vocabulary items, not labels for entire documents.

Microsoft Learn: [Introduction to generative AI and agents: Large language models \(LLMs\)](#)

16 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

- B** Attention

A transformer uses attention layers to consider each token in the context of surrounding tokens. Attention assigns weights to reflect how much nearby tokens influence the current token representation.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Batch deployment is a deployment pattern, not a transformer technique for learning contextual relationships.
- **C.** Keyless authentication controls access to a service endpoint, not the model architecture.
- **D.** Content filtering evaluates prompts and responses for safety concerns, not token influence during transformer training.

Microsoft Learn: [Introduction to generative AI and agents: Large language models \(LLMs\)](#)

17 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

- C** The model predicts the next token, appends it, and repeats the process.

The decoder predicts the most probable next token based on the prompt and the sequence generated so far. The predicted token is added to the sequence, and the process repeats until the model finishes the completion.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Language generation is based on token prediction from learned patterns and context, not only exact document matching.
- **B.** A normal prompt does not permanently retrain the model parameters.
- **D.** API keys authenticate requests and are unrelated to token prediction.

Microsoft Learn: [Introduction to generative AI and agents: Large language models \(LLMs\)](#)

18 Multiple choice · Identify AI concepts and capabilities

ANSWER: D

- D** Retrieve relevant handbook passages and add them to the prompt as grounding context.

RAG searches selected data sources for information relevant to the user's question and adds the retrieved content to the prompt. The language model then generates a grounded response from the supplied context without changing its learned parameters.

WHY THE OTHER CHOICES ARE WRONG

- **A.** RAG uses the user prompt together with retrieved context rather than ignoring the user question.
- **B.** RAG does not require replacing the tokenizer or retraining the model.
- **C.** A system message can guide behavior, but large or changing knowledge is commonly retrieved at run time.

Microsoft Learn: [Introduction to retrieval-augmented generation concepts: Understand retrieval-augmented generation](#) · [Introduction to retrieval-augmented generation concepts: Retrieve information and generate a response](#)

19 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: A, E

- A** Divide source text into focused chunks.
- E** Encode the chunks as embedding vectors.

A RAG indexing pipeline commonly chunks source text into focused passages and encodes those chunks as embedding vectors. The chunks, embeddings, and metadata are stored in a search index or vector-enabled data store for retrieval.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Useful metadata such as title, source, category, and permissions should be added to support retrieval and access control.
- **C.** A deployment type controls model inference processing and does not create a vector index.
- **D.** Temperature affects response generation behavior, not the preparation of a vector index.

Microsoft Learn: [Introduction to retrieval-augmented generation concepts: Prepare data for retrieval](#)

20 Multiple choice · Identify AI concepts and capabilities

ANSWER: D

D A small language model

Microsoft Learn notes that different model sizes serve different needs, and a lightweight phone application may run best on a small language model. Small language models can be appropriate when the task is narrow and the application values efficient responses.

WHY THE OTHER CHOICES ARE WRONG

- **A.** An image generation model creates images from prompts and does not fit a text FAQ assistant.
- **B.** A speech to text model transcribes audio and does not generate FAQ answers from text questions.
- **C.** A video generation model creates video scenes and does not fit a narrow text question answering scenario.

Microsoft Learn: [Get started with generative AI and agents in Azure: Generative AI models](#) · [Microsoft Foundry Models overview](#)

21 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

C Multimodal chat

The model catalog includes multimodal models that support images and other inputs in chat scenarios. Azure OpenAI model descriptions identify models that can accept text and image input and produce text output.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Image generation creates new images from prompts rather than interpreting uploaded photos to answer questions.
- **B.** Batch processing is a deployment and processing pattern, not a model capability for image understanding.
- **D.** Embeddings convert inputs into vectors for similarity search rather than answering questions about images directly.

Microsoft Learn: [Get started with generative AI and agents in Azure: Generative AI models](#) · [Foundry Models sold by Azure](#)

22 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

A Reasoning

Reasoning models are designed for reasoning and problem-solving tasks with increased focus and capability. Microsoft Learn lists complex document comparison and advanced problem solving as suitable reasoning model scenarios.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Image generation creates visual content and does not target complex contract analysis.
- **C.** Speech synthesis produces spoken audio from text and does not perform contract reasoning.
- **D.** Embeddings support similarity search, but embeddings alone do not generate a multi-step comparison explanation.

Microsoft Learn: [Azure OpenAI reasoning models - GPT-6 Astra and GPT-5](#) · [Get started with generative AI and agents in Azure: Generative AI models](#)

23 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: A, E

- A** Model details such as capabilities and supported inference tasks
- E** Benchmark metrics or model leaderboards

The Foundry model catalog provides benchmark data, leaderboards, filters, and model card details to help compare models. Model details include descriptions, supported data types, capabilities, and inference tasks that help match a model to a workload.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Temperature configures response behavior for inference and does not describe a catalog comparison capability.
- **C.** API keys authenticate calls to deployed resources and do not compare model quality or capabilities.
- **D.** A firewall rule can help secure access, but it is not a model catalog selection signal.

Microsoft Learn: [Microsoft Foundry Models overview](#) · [Get started with generative AI and agents in Azure: Generative AI models](#)

24 Multiple choice, choose three · Identify AI concepts and capabilities

ANSWERS: C, D, G

- C** Audio
- D** Image generation
- G** Embeddings

Embeddings models convert text into vector representations for semantic retrieval. Image generation models create original images from natural language, and audio models support tasks such as speech to text, translation, and text to speech.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Developer deployment is for fine-tuned model evaluation and does not identify the needed model capabilities.
- **B.** Content filtering is a safety capability and does not create embeddings, images, or call transcripts.
- **E.** Managed compute is a deployment option, not a model capability for any of the three planned features.
- **F.** Global Standard is a deployment type rather than a model capability or inference task.

Microsoft Learn: [Foundry Models sold by Azure](#) · [Microsoft Foundry REST Reference - Azure OpenAI - Embeddings](#)

25 Multiple choice · Identify AI concepts and capabilities

ANSWER: D

- D** A system message

A system message is a high-priority set of instructions and context sent to a chat model to steer how it responds. It can define the role, tone, scope, boundaries, safety guidelines, and tool or data usage guidance.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A vector index chunk is retrieved grounding content and does not define the assistant's overall behavior.
- **B.** A deployment SKU controls deployment processing characteristics, not the assistant's role or boundaries.
- **C.** A model benchmark helps compare models and does not set runtime behavior for the assistant.

Microsoft Learn: [Safety system messages](#) · [Get started with generative AI and agents in Azure: Using a generative AI model](#)

26 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

A Global Standard

Microsoft Learn recommends starting with Global Standard for most workloads unless you have a specific reason to choose another deployment type. Global Standard is the first option for many newly released models and provides broad model availability.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Developer deployment is for evaluating fine-tuned models and is not the recommended starting point for most workloads.
- **C.** Data Zone Standard is appropriate when prompts and responses must be processed within a Microsoft-specified data zone.
- **D.** Regional Provisioned is appropriate when reserved throughput and Azure geography processing are required.

Microsoft Learn: [Understanding deployment types in Microsoft Foundry Models](#)

27 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

B Global Provisioned

Provisioned deployment types reserve model processing capacity for predictable throughput. Global Provisioned combines global routing with guaranteed capacity and lower, more consistent latency than Global Standard.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Developer deployment is intended for temporary fine-tuned model evaluation, not production throughput.
- **C.** Standard is suited to variable or bursty workloads and does not reserve model processing capacity.
- **D.** Global Batch is for large asynchronous jobs and trades real-time responsiveness for batch processing.

Microsoft Learn: [Understanding deployment types in Microsoft Foundry Models](#)

28 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: A, C

A Max output tokens

C Temperature

The Foundry playground exposes model parameters such as temperature, max output tokens, and system instructions. Temperature controls creativity versus determinism, while max output tokens caps response length.

WHY THE OTHER CHOICES ARE WRONG

- **B.** The project name organizes assets and does not control model response behavior.
- **D.** The deployment region affects where a model is available or processed, not the creativity or output length of a response.
- **E.** The model family affects available capabilities, but it is not the runtime parameter that directly controls creativity in this requirement.

Microsoft Learn: [Get started with generative AI and agents in Azure: Using a generative AI model · Microsoft Foundry REST Reference - Azure OpenAI - Responses](#)

29 Multiple choice · Identify AI concepts and capabilities

ANSWERS: A, D

- A** Data Zone Standard
- D** Data Zone Provisioned

Data Zone deployment types process prompts and responses only within the specified Microsoft data zone, such as US, EU, or APAC. Data Zone Standard fits variable interactive traffic, while Data Zone Provisioned adds reserved model processing capacity for predictable throughput.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Regional Provisioned keeps processing within an Azure geography, not within a Microsoft-specified data zone.
- **C.** Global Standard can process prompts and responses in any Azure region and does not meet the data zone requirement.
- **E.** Managed compute is a deployment option for supported open or custom models and does not identify the data zone deployment types for these apps.
- **F.** Developer deployment is for fine-tuned model evaluation and does not provide the requested production data zone options.

Microsoft Learn: [Understanding deployment types in Microsoft Foundry Models](#)

30 Multiple choice · Identify AI concepts and capabilities

ANSWER: D

- D** Generative AI

Generative AI enables applications to create new content, including natural language dialogs, images, video, code, and other formats. A product description written from user notes is original text generated from a prompt, so generative AI is the appropriate workload.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Object detection identifies the locations of objects in images, not new written content.
- **B.** Information extraction retrieves data from existing content, not original product descriptions.
- **C.** Speech recognition converts spoken language into text, not rough notes into original descriptions.

Microsoft Learn: [Introduction to AI concepts: Generative AI and agents](#)

31 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

- A** Agentic AI

AI agents are software applications built on generative AI that can reason over and generate natural language. Agents can also use knowledge and action tools, such as search, calendars, or email, to automate tasks in response to context.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Image classification predicts the main label for an image and does not perform business actions.
- **C.** Speech synthesis converts written text to spoken audio but does not reason over prompts or use tools.
- **D.** Key phrase extraction identifies main concepts in text but does not use tools to schedule callbacks or send email.

Microsoft Learn: [Introduction to AI concepts: Generative AI and agents](#)

32 Multiple choice · Identify AI concepts and capabilities

ANSWERS: A, B, F

- A** Generate a concise answer to a user's product question in natural language.
- B** Transcribe recorded customer calls into text.
- F** Identify products that appear in uploaded shelf photos.

Generating a natural language answer is a generative AI scenario. Transcribing calls is a speech recognition scenario, and identifying products in photos is a computer vision scenario. These are all common AI workloads described in Microsoft Learn.

WHY THE OTHER CHOICES ARE WRONG

- **C.** Role-based access control is an authorization requirement, not an AI workload.
- **D.** Database replication is an infrastructure availability requirement, not an AI workload.
- **E.** Virtual networking is an infrastructure configuration requirement, not an AI workload.

Microsoft Learn: [Introduction to AI concepts: Generative AI and agents](#) · [Introduction to AI concepts: Speech](#) · [Introduction to AI concepts: Computer vision](#)

33 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

- C** Information extraction

Information extraction solutions find information in unstructured sources such as scanned documents, forms, and images. Receipt processing typically combines OCR with a model that interprets extracted values and maps them to fields such as vendor, date, and total.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Sentiment analysis classifies emotional tone in text and does not map receipt values to fields.
- **B.** Speech synthesis converts text into audible speech and does not process receipt images.
- **D.** Image generation creates new visual content and does not extract existing receipt data.

Microsoft Learn: [Introduction to AI concepts: Information extraction](#)

34 Multiple choice · Identify AI concepts and capabilities

ANSWER: C

- C** Tokenization

Tokenization is the first step in analyzing a body of text, or corpus. It breaks text into tokens, which can be words, partial words, or combinations of words and punctuation that can be counted or analyzed.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Speech synthesis produces spoken audio from text and is not a text preparation technique.
- **B.** Image generation creates visual output from prompts and does not divide text into tokens.
- **D.** Semantic segmentation classifies pixels in an image and does not prepare text for term counts.

Microsoft Learn: [Introduction to natural language processing concepts: Tokenization](#)

35 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

- A** Named entity recognition

Named entity recognition identifies entities in text, such as people, locations, organizations, dates, and quantities. It is the appropriate technique when the required output is a list of typed mentions from an article.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Summarization reduces the volume of text while retaining the main points.
- **C.** Key phrase extraction identifies main topics or phrases, not typed mentions such as people and dates.
- **D.** Sentiment analysis identifies whether text is positive, negative, or neutral.

Microsoft Learn: [Introduction to AI concepts: Text and natural language](#) · [Get started with text analysis in Azure: Understand text analysis in Foundry](#)

36 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: B, D

- B** Key phrase extraction
- D** Summarization

Key phrase extraction identifies the main concepts in text, which can be used as searchable tags. Summarization reduces the amount of text while retaining the main points, which meets the requirement for shorter review versions.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Sentiment analysis classifies emotional tone and does not create tags or summaries.
- **C.** Speech recognition transcribes spoken audio and does not tag or shorten text reviews.
- **E.** Semantic segmentation identifies pixels that belong to objects in images.

Microsoft Learn: [Introduction to AI concepts: Text and natural language](#) · [Get started with text analysis in Azure: Understand text analysis in Foundry](#)

37 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

- B** Term Frequency - Inverse Document Frequency (TF-IDF)

TF-IDF calculates a score based on how often a term appears in one document compared with how often it appears across the full collection. A high TF-IDF score indicates that a term appears often in one document but rarely in others, which helps distinguish articles in the same corpus.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Part-of-speech tagging labels grammatical categories and does not rank distinctive terms in a corpus.
- **C.** Stop word removal excludes common words from analysis but does not score terms by document relevance.
- **D.** Simple frequency analysis counts terms in a document but does not account for how common they are across the corpus.

Microsoft Learn: [Introduction to natural language processing concepts: Statistical text analysis](#)

38 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

A Speech recognition

Speech recognition, also called speech-to-text, converts spoken language into written text. It is the correct capability for producing transcripts from uploaded audio recordings.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Image classification predicts labels for images and does not process speech recordings.
- **C.** Speech synthesis converts written text into spoken audio, not audio into text.
- **D.** Key phrase extraction finds main concepts in text after the speech has already been transcribed.

Microsoft Learn: [Introduction to AI concepts: Speech](#) · [Introduction to AI speech concepts: Speech recognition](#)

39 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

A Speech synthesis

Speech synthesis, also called text-to-speech, converts written text into spoken audio. It is used when an application must communicate text content audibly to users.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Speech recognition converts spoken audio into text, not text into audio.
- **C.** Entity detection finds typed mentions in text and does not create spoken output.
- **D.** Object detection locates objects in images and does not read alerts aloud.

Microsoft Learn: [Introduction to AI concepts: Speech](#) · [Introduction to AI speech concepts: Speech synthesis](#)

40 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: B, C

B Speech recognition

C Speech synthesis

A two-way voice assistant needs speech recognition to transcribe caller speech into text. It also needs speech synthesis to convert the assistant response from text into natural-sounding audio.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Optical character recognition extracts text from images and does not process a phone conversation.
- **D.** TF-IDF scoring ranks distinctive terms in documents and does not enable voice interaction.
- **E.** Object detection locates objects in images and does not support spoken requests or responses.

Microsoft Learn: [Introduction to AI speech concepts: Speech-enabled solutions](#)

41 Multiple choice · Identify AI concepts and capabilities

ANSWER: D

D Text normalization

Text normalization prepares written text for speech synthesis by expanding abbreviations, numbers, dates, times, and symbols into spoken forms. This prevents the system from trying to pronounce raw symbols or digits unnaturally.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Key phrase extraction identifies main concepts in text and does not expand abbreviations or symbols for speech.
- **B.** Object detection identifies objects in images and is unrelated to pronouncing text.
- **C.** Beam search decoding is used in speech recognition to choose a likely word sequence, not to expand written text for speech.

Microsoft Learn: [Introduction to AI speech concepts: Speech synthesis](#)

42 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

A Image classification

Image classification predicts a text label based on an image's contents. It is appropriate when the goal is to identify the main subject or category in a single product photo.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Speech recognition converts spoken audio to text and does not classify product photos.
- **C.** Semantic segmentation classifies individual pixels and is not needed to return one category label.
- **D.** Object detection finds individual objects and their locations, which is more than predicting the main product category.

Microsoft Learn: [Introduction to AI concepts: Computer vision](#) · [Introduction to computer vision concepts: Computer vision tasks and techniques](#)

43 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: A, B

A Object detection

B Semantic segmentation

Object detection finds individual objects in an image and returns their regions, commonly as bounding boxes. Semantic segmentation also provides location information by classifying individual pixels based on the object to which they belong.

WHY THE OTHER CHOICES ARE WRONG

- **C.** Image classification predicts an overall label and does not return object locations.
- **D.** Speech synthesis converts text into audio and does not analyze images.
- **E.** Sentiment analysis classifies emotional tone in text and does not analyze object locations in images.

Microsoft Learn: [Introduction to AI concepts: Computer vision](#) · [Introduction to computer vision concepts: Computer vision tasks and techniques](#)

44 Multiple choice · Identify AI concepts and capabilities

ANSWER: B

- B** Text-to-image generation

Image generation models can create images in response to natural language descriptions. Modern image generation commonly uses diffusion, starting with random pixel values and iteratively removing noise until the image matches the prompt.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Object detection locates existing objects in an image and does not create new images.
- **C.** Sentiment analysis classifies text as positive, negative, or neutral and does not create visual output.
- **D.** Optical character recognition converts visible text in images into text data and does not create images.

Microsoft Learn: [Introduction to computer vision concepts: Image generation](#) · [Get started with computer vision in Azure: Image generation models](#)

45 Multiple choice, choose three · Identify AI concepts and capabilities

ANSWERS: A, C, F

- A** Create a product concept image from a text prompt.
- C** Predict a label for the main subject in a photo.
- F** Find each bicycle in a warehouse image and return bounding boxes.

Image classification predicts a label for the main subject in a photo. Object detection locates objects in an image, and text-to-image generation creates new visual output from prompts. These are computer vision or image-generation capabilities.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Ranking terms by TF-IDF scores is a statistical text analysis technique.
- **D.** Detecting whether text is positive or negative is sentiment analysis, a text analysis technique.
- **E.** Converting a call recording to text is speech recognition, not computer vision or image generation.

Microsoft Learn: [Introduction to computer vision concepts: Computer vision tasks and techniques](#) · [Introduction to computer vision concepts: Image generation](#)

46 Multiple choice · Identify AI concepts and capabilities

ANSWER: A

- A** Optical character recognition

Optical character recognition, or OCR, converts visual text in scanned documents, photographs, screenshots, and similar files into editable, searchable text data. It is the foundational technique for extracting text from images of printed forms.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Sentiment analysis determines emotional tone in text and does not read text from images.
- **C.** Image generation creates new images and does not convert existing visual text into data.
- **D.** Speech synthesis turns text into spoken audio and does not extract text from form images.

Microsoft Learn: [Introduction to AI concepts: Information extraction](#) · [Introduction to AI-powered information extraction concepts: Optical character recognition \(OCR\)](#)

47 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: B, D

- B** Field extraction and mapping
- D** Optical character recognition

Information extraction commonly uses OCR to detect and extract text from image-based documents. Field extraction then maps individual text values to meaningful business fields, such as vendor name, invoice date, and total.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Prosody generation controls rhythm and intonation in synthesized speech, not invoice extraction.
- **C.** Image classification predicts an image label and does not map invoice values to business fields.
- **E.** Text-to-image generation creates new images from prompts and does not extract invoice data.

Microsoft Learn: [Introduction to AI-powered information extraction concepts: Overview of information extraction](#) · [Introduction to AI-powered information extraction concepts: Field extraction and mapping](#)

48 Multiple choice, choose two · Identify AI concepts and capabilities

ANSWERS: C, D

- C** Summarize a voicemail and extract the callback number left by the caller.
- D** Analyze a recorded meeting to list speakers, assigned actions, and attendance counts.

Information extraction can analyze audio files to provide transcriptions, summaries, and key insights such as caller details or callback numbers. It can also analyze video recordings to extract meeting information such as attendance, speakers, discussion summaries, and assigned actions.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Generating a new logo from a prompt is image generation, not extraction from audio or video.
- **B.** Classifying a review as positive or negative is sentiment analysis on text.
- **E.** Converting written text to spoken audio is speech synthesis, not information extraction.

Microsoft Learn: [Get started with AI-powered information extraction in Azure: Extract information from audio and video](#)

49 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** A system prompt

A system prompt sets the behavior, tone, role, tools, and guardrails for a generative AI model. A user prompt is the individual question or instruction that asks the model to produce a response. Put durable behavior rules in the system prompt so they apply across user turns.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A user prompt is the specific end-user request, not the persistent behavior and guardrails for the assistant.
- **B.** Temperature controls creativity compared to determinism, not the role, tone, and constraints of the assistant.
- **C.** Max output tokens caps response length, but it does not define the assistant role or policy boundaries.

Microsoft Learn: [Introduction to generative AI and agents: Prompts](#) · [Get started with generative AI and agents in Azure: Using a generative AI model](#)

50 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: A

- A** Summarize the attached policy changes for store managers in five bullet points.

Effective prompts are clear and specific, and they can include context, audience, examples, and desired structure. The correct prompt identifies the source content, audience, and output format. That guidance helps the model produce a more useful response.

WHY THE OTHER CHOICES ARE WRONG

- **B.** This prompt is vague and does not define the audience, scope, or output format.
- **C.** Temperature is a model setting, not a user prompt that gives task context and output requirements.
- **D.** This prompt does not specify what content to create or how the output should be structured.

Microsoft Learn: [Introduction to generative AI and agents: Prompts](#)

51 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, E

- A** Include relevant earlier turns or a summary of them in the next prompt.
- E** Include the new user question in the next prompt.

Generative AI chat applications often maintain conversation history and include it, or a summarized version of it, in later prompts. The next prompt still needs the current user question. Together, these inputs help the model answer in context.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Removing history prevents the model from using previous turns to understand follow-up questions.
- **C.** More output tokens can allow longer answers, but it does not provide the missing conversational context.
- **D.** A previous response is conversation context, not a replacement for the system prompt that defines behavior.

Microsoft Learn: [Introduction to generative AI and agents: Prompts](#)

52 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** Clearly label retrieved chunks as context and tell the model to treat them as data.

A RAG prompt should distinguish trusted application instructions from retrieved content. The retrieved chunks provide grounding context, but they should be treated as data rather than instructions. The prompt can also tell the model to answer from supplied context and acknowledge insufficient evidence.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Trusted application instructions are needed to define the assistant behavior and how retrieved context should be used.
- **B.** Retrieved documents should be treated as data, because source content could contain text that attempts to redirect the model.
- **D.** Putting retrieved content into the system instructions can blur the boundary between trusted instructions and untrusted data.

Microsoft Learn: [Introduction to retrieval-augmented generation concepts: Retrieve information and generate a response](#)

53 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, B, F

- A** Specify the desired format and maximum length.
- B** State the audience and purpose of the email.
- F** Provide an example of the preferred style.

Prompt quality improves when the prompt is clear, specific, and includes useful context. Asking for structure, such as a length or bullet format, helps shape the response. Examples can guide the model toward a desired style.

WHY THE OTHER CHOICES ARE WRONG

- **C.** TPM is a deployment capacity setting, not content that improves prompt quality.
- **D.** Removing context makes the request less specific and can reduce the relevance of the generated email.
- **E.** Leaving the answer length open does not guide the model toward the required output structure.

Microsoft Learn: [Introduction to generative AI and agents: Prompts](#)

54 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** The Foundry model catalog

The Foundry model catalog is the central place to discover and use generative AI models. Model entries can include descriptions, capabilities, benchmark results, supported inference tasks, and responsible AI documentation. Use it before choosing and deploying a model.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A knowledge source provides content for retrieval, but it is not the catalog for choosing a model.
- **B.** A Key Vault secret can protect credentials, but it does not help compare model capabilities or benchmarks.
- **D.** An App Service plan hosts web apps, but it is not where Foundry models are discovered and compared.

Microsoft Learn: [Get started with generative AI and agents in Azure: Generative AI models](#)

55 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** The deployment name

When you create a deployment, the deployment name is used in the model parameter during inference. The deployment name routes requests to that configured deployment. This allows code to call the correct deployed model.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The catalog publisher can describe the model source, but it does not route requests to a deployment.
- **B.** The resource group organizes resources, but it is not used as the model parameter for routing inference.
- **C.** The subscription contains Azure resources, but it does not identify the specific model deployment for inference.

Microsoft Learn: [Deploy Microsoft Foundry Models in the Foundry portal](#)

56 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, C

- A** Select Deploy and choose default or custom deployment settings.
- C** Open Discover, select Models, and choose a model.

The Foundry portal deployment flow starts by opening Discover, selecting Models, and choosing a model to review. You then select Deploy and choose default or custom settings before creating the deployment. After deployment, you can test the model in the playground.

WHY THE OTHER CHOICES ARE WRONG

- **B.** A hosted agent is a code-based agent path, not a prerequisite for deploying a model.
- **D.** A knowledge base can ground agents, but it is not required to deploy a model from the catalog.
- **E.** MCP tools are for agent integrations and are not required to create a model deployment.

Microsoft Learn: [Deploy Microsoft Foundry Models in the Foundry portal](#)

57 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** The Foundry Playgrounds

The Foundry Playgrounds let you try prompts, compare models, and capture working settings before writing code. Playground settings include temperature, max output tokens, and system instructions. The chat interface shows the model responses to submitted prompts.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Role assignments manage access, but they do not provide prompt testing or model response settings.
- **C.** Cost analysis helps review spending, but it is not the interface for testing prompts and model settings.
- **D.** A knowledge source configuration page manages retrieval sources, not playground chat settings for a deployed model.

Microsoft Learn: [Get started with generative AI and agents in Azure: Using a generative AI model](#)

58 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: A

- A** Max output tokens

Max output tokens caps response length and affects token consumption. The training content notes that larger prompts and higher max output token settings consume more TPM and can contribute to rate-limit errors. Lowering max tokens is one way to reduce token traffic.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Benchmark scores help compare models, but they do not change token consumption for a deployment.
- **C.** System instructions define behavior, but lowering max output tokens is the setting that caps generated response length.
- **D.** Temperature changes creativity compared to determinism, but it does not directly cap generated token volume.

Microsoft Learn: [Get started with generative AI and agents in Azure: Generative AI models](#) · [Get started with generative AI and agents in Azure: Using a generative AI model](#)

59 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

C `get_openai_client`

The Foundry SDK exposes a project client and an OpenAI-compatible client. In Python, `AIProjectClient` provides `get_openai_client` to obtain an authenticated client for Responses API calls. The lightweight chat client can then call `openai.responses.create`.

WHY THE OTHER CHOICES ARE WRONG

- **A.** `conversations.create` creates a conversation after an OpenAI-compatible client already exists.
- **B.** `deployments.list` is for enumerating deployments, not for creating a client to call the Responses API.
- **D.** `agents.create_version` creates or versions an agent, but it does not return the OpenAI-compatible client.

Microsoft Learn: [Get started with Microsoft Foundry SDKs and endpoints](#) · Quickstart: [Get started with Microsoft Foundry SDK](#)

60 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, C

- A** Use `DefaultAzureCredential` to authenticate with Microsoft Entra ID.
- C** Use the project endpoint in the form `https://<resource>.services.ai.azure.com/api/projects/<project>`.

The Foundry SDK uses a project endpoint such as `https://<resource-name>.services.ai.azure.com/api/projects/<project-name>`. Python samples create `AIProjectClient` with that endpoint and `DefaultAzureCredential`. The user signs in with Azure CLI so the credential can obtain a token.

WHY THE OTHER CHOICES ARE WRONG

- **B.** The endpoint host is based on the Foundry resource name or custom subdomain, not the resource group name.
- **D.** The browser portal URL is not the project endpoint used by the SDK for service calls.
- **E.** The Foundry SDK samples use Microsoft Entra ID through `DefaultAzureCredential`.

Microsoft Learn: [Get started with Microsoft Foundry SDKs and endpoints](#) · Quickstart: [Get started with Microsoft Foundry SDK](#)

61 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

D `model`

In a Responses API call, the `model` parameter names the model deployment that should handle the request, and the `input` parameter carries the user request. The service routes the request to that deployment, and `output_text` returns the generated answer.

WHY THE OTHER CHOICES ARE WRONG

- **A.** An API key is a credential value, not the parameter that selects the deployed model.
- **B.** The project is represented by the project endpoint and client, not by a `project` parameter in this call.
- **C.** The endpoint is used when creating the project client, not as the deployment selector in `responses.create`.

Microsoft Learn: [Quickstart: Get started with Microsoft Foundry SDK](#) · [Azure AI Projects client library for Python](#)

62 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, C, D

- A** Display the returned results.
- C** Collect user input.
- D** Call a remote service or API.

A lightweight client application collects user input, calls a remote service or API, and displays results. In Foundry, the model runs remotely and the client keeps a small code footprint. This pattern is suitable for a simple CLI or small app.

WHY THE OTHER CHOICES ARE WRONG

- **B.** The model runs remotely, so the client does not need to run GPU inference on the user device.
- **E.** A lightweight Foundry chat client calls a remote model deployment rather than training the foundation model locally.
- **F.** The client uses the Foundry project endpoint rather than replacing it.

Microsoft Learn: [Get started with generative AI and agents in Azure: Using a generative AI model](#)

63 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** A software development kit

Applications can call Foundry endpoints with REST requests, but most developers prefer software development kits (SDKs), which wrap the REST interfaces in language-specific code libraries. A lightweight chat client can use the Foundry SDK to call the model.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A deployment script can create resources, but it is not the client library used to call model APIs.
- **C.** Benchmark tables help compare models, but they do not build service requests for a programming language.
- **D.** A knowledge source provides content for retrieval, but it does not abstract REST API calls in application code.

Microsoft Learn: [Get started with AI in Azure: Using Microsoft Foundry endpoints](#) · [Get started with Microsoft Foundry SDKs and endpoints](#)

64 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** Instructions

A Foundry agent brings together a model, instructions, and tools. Instructions are the system prompt that defines the agent role, behavior, style, constraints, and output rules. Configure instructions when the requirement is about how the agent should behave.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A knowledge source provides grounding content, but it does not by itself define the agent role and constraints.
- **B.** A quota limit controls capacity, but it does not define how the agent should respond.
- **D.** A benchmark helps compare models, but it does not define the agent role, tone, or behavior.

Microsoft Learn: [Get started with generative AI and agents in Azure: Creating an agent](#) · [Quickstart: Create a prompt agent](#)

65 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, C

- A** Write instructions that tell the agent when to use the knowledge base and cite sources.
- C** Add the knowledge base connection to the agent as an MCP tool.

Foundry IQ knowledge bases expose an MCP tool for agent integration. A typical workflow adds the knowledge base connection to an agent as an MCP tool and defines instructions for when and how to use it. Instructions should require grounding and citations when the knowledge base supplies evidence.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Citations help users inspect supporting sources, so disabling them conflicts with the grounded response requirement.
- **D.** Separate model deployments for documents do not provide retrieval from the knowledge base.
- **E.** Higher temperature can make answers more creative, but it does not connect the agent to the knowledge base.

Microsoft Learn: [Get started with Microsoft Foundry IQ: Connect an agent to Foundry IQ](#)

66 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** Realistic questions with clear answers, ambiguous cases, missing-answer cases, and permission variations

Microsoft Learn recommends testing knowledge-enhanced agents with realistic questions, questions requiring multiple sources, ambiguous questions, no-answer questions, and users with different permissions. You should evaluate retrieval separately from the final response. This helps find content gaps, weak retrieval, permission issues, and instruction problems.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Creative writing prompts do not test whether factual answers are grounded in authorized retrieved content.
- **B.** Small-talk prompts do not test retrieval, citations, missing evidence, or user permissions.
- **C.** Model benchmark prompts compare model behavior, but they do not validate the connected knowledge base experience.

Microsoft Learn: [Get started with Microsoft Foundry IQ: Connect an agent to Foundry IQ](#) · [Introduction to retrieval-augmented generation concepts: Evaluate a RAG solution](#)

67 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** A prompt agent

Prompt agents are defined through configuration, including instructions, model selection, and tools. You can author them in the Foundry portal, test them in the playground, and call them from application code. Foundry runs the prompt agent without application runtime code or containers to maintain.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A voice-based prompt agent is for real-time spoken conversations, not the fastest text-only portal-first option.
- **B.** An ephemeral agent is assembled in application code and is not a persisted portal-first agent resource.
- **D.** A hosted agent is for custom code packaged as a container or source package, which adds runtime code to maintain.

Microsoft Learn: [What is Microsoft Foundry Agent Service?](#) · [Quickstart: Create a prompt agent](#)

68 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

D agent_name="Agent1"

Calling get_openai_client with agent_name returns an OpenAI-compatible client that is bound to that agent, so later responses.create calls go to the agent without naming it again. The Foundry SDK requires allow_preview=True on the project client to use the agent_name option.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The model argument selects a model deployment in a response call, not the agent to bind to the client.
- **B.** An API key is a credential, not the parameter for selecting an agent by name.
- **C.** A conversation identifies a chat thread, not the agent resource to bind when the client is created.

Microsoft Learn: [Quickstart: Create a prompt agent](#) · [azure.ai.projects.AIProjectClient class](#)

69 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, E

A Pass conversation.id in each responses.create call for the thread.

E Create a conversation by calling openai.conversations.create.

Creating a conversation gives the chat its own thread. Passing the conversation ID in each responses.create call keeps the first question and the follow-up in the same thread, so the agent can use the earlier turn as context.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Creating a new agent version changes the agent definition and is not needed for each chat turn.
- **C.** The model parameter is for selecting a model deployment in model calls, not for sending user input.
- **D.** Removing instructions would weaken the consistent agent behavior across turns.

Microsoft Learn: [Quickstart: Create a prompt agent](#)

70 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

B response.output_text

The responses.create call returns a response object, and its output_text property contains the text that the agent generated. The conversation ID only links the turns of a chat together, and the response ID identifies the response rather than containing its text.

WHY THE OTHER CHOICES ARE WRONG

- **A.** response.id identifies the response object, not the generated answer text.
- **C.** project.endpoint identifies the project service endpoint, not the generated answer.
- **D.** conversation.id identifies the chat thread, not the generated answer.

Microsoft Learn: [Quickstart: Create a prompt agent](#)

71 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWERS: C, F

- C** The prompt agent name
- F** The Foundry project endpoint

A client for a prompt agent needs the Foundry project endpoint, in the `https://<resource>.services.ai.azure.com/api/projects/<project>` format, and the name of the agent to call. The code uses the endpoint to create the project client and the agent name to create an agent-scoped OpenAI-compatible client.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The Azure portal home page URL is not the service endpoint used by `AIProjectClient`.
- **B.** The subscription display name is not used by the client to connect to a project or select an agent.
- **D.** A TPM quota request number is not part of the lightweight agent client configuration.
- **E.** The model catalog publisher name does not identify the agent that the client should call.

Microsoft Learn: [Quickstart: Create a prompt agent](#) · [Get started with Microsoft Foundry SDKs and endpoints](#)

72 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** A `TextAnalyticsClient` that uses `AzureKeyCredential`

Azure Language in Foundry Tools is accessed through a Microsoft Foundry resource endpoint and can authenticate by key or Microsoft Entra ID. The Python example creates a `TextAnalyticsClient` with an `AzureKeyCredential` for key-based authentication.

WHY THE OTHER CHOICES ARE WRONG

- **A.** An `AIProjectClient` accesses Foundry project APIs, not the Text Analytics client methods shown for Azure Language.
- **C.** A `SpeechRecognizer` is used with Azure Speech, not with Azure Language text analysis.
- **D.** An `AzureOpenAI` client is used for model calls, not for the Azure Language SDK methods.

Microsoft Learn: [Analyze text with Azure Language in Foundry Tools: Azure Language in Microsoft Foundry Tools](#) · [Get started with text analysis in Azure: Create a client application that analyzes text](#)

73 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: A

- A** `detect_language`

The language detection API evaluates text and returns language identifiers with a confidence score for each document. The Python sample calls `client.detect_language` with a `documents` collection and reads `primary_language` values from the response.

WHY THE OTHER CHOICES ARE WRONG

- **B.** `begin_analyze_healthcare_entities` is for health entity extraction, not general language routing.
- **C.** `recognize_pii_entities` identifies and redacts personal information, not the document language.
- **D.** `recognize_entities` identifies categorized entities, not the primary language of the document.

Microsoft Learn: [Analyze text with Azure Language in Foundry Tools: Detect language](#) · [What is Azure Language in Foundry Tools](#)

74 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: B, D

- B** Read redacted_text and the PII entity categories from the result.
- D** Call recognize_pii_entities with the documents to analyze.

The PII extraction unit shows that recognize_pii_entities identifies personal information such as names, email addresses, phone numbers, and social security numbers. The same response can include PII entities and redacted_text, which is the masked text to store.

WHY THE OTHER CHOICES ARE WRONG

- **A.** recognize_entities returns general named entities, but it is not the PII redaction method.
- **C.** primary_language.iso6391_name is returned by language detection, not by PII extraction.
- **E.** detect_language can identify a document language, but it does not identify or mask PII.
- **F.** SpeechRecognizer transcribes audio and does not analyze text for PII.

Microsoft Learn: [Analyze text with Azure Language in Foundry Tools: Extract personally identifiable information \(PII\)](#) · [Get started with text analysis in Azure: Understand text analysis in Foundry](#)

75 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** Azure Language in Foundry Tools as an MCP server

The Azure Language MCP server exposes Azure Language in Foundry Tools capabilities such as named entity recognition, sentiment analysis, language detection, and PII redaction. In the Foundry portal, you connect the server to an agent by selecting Azure Language in Foundry Tools and configuring the Foundry resource name.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Voice Live is for real-time voice conversations, not deterministic text analysis.
- **C.** Foundry IQ supports knowledge grounding, not Azure Language PII and language detection tools.
- **D.** An image generation model creates visual content and does not expose Azure Language analysis capabilities.

Microsoft Learn: [Get started with text analysis in Azure: Use Azure Language with an agent](#) · [What is Azure Language in Foundry Tools](#)

76 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: B, D

- B** The SDK includes methods such as `detect_language` and `recognize_pii_entities`.
- D** Azure Language returns structured results for purpose-built analysis tasks.

A general-purpose model prompt can word its answer differently on each run, while Azure Language returns structured results for purpose-built analysis tasks. Its SDK exposes methods such as `detect_language` and `recognize_pii_entities`, which return consistent fields such as a language code, a confidence score, and redacted text.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The Speech SDK is for speech-enabled apps, not for Azure Language text redaction.
- **C.** The module explains that model prompt outputs can vary, while Azure Language is better for consistent structured values.
- **E.** Azure Language SDK calls use a Foundry resource endpoint and credentials, not a deployed generative model name.

Microsoft Learn: [Get started with text analysis in Azure: Create a client application that analyzes text](#) · [What is Azure Language in Foundry Tools](#)

77 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: A

- A** `gpt-4o-transcribe`

The speech-capable generative AI module lists `gpt-4o-transcribe`, `gpt-4o-mini-transcribe`, and `gpt-4o-transcribe-diarize` as models that support speech-to-text operations. The model catalog documentation also lists `gpt-4o-transcribe` as a speech-to-text model for the audio API.

WHY THE OTHER CHOICES ARE WRONG

- **B.** `gpt-image-1` generates images and does not process speech audio.
- **C.** `gpt-4o-mini-tts` is used to generate audio from text, not to transcribe spoken prompts.
- **D.** `text-embedding-3-small` creates vector embeddings and does not transcribe audio.

Microsoft Learn: [Develop a speech-capable generative AI application: Transcribe speech](#) · [Foundry Models sold by Azure](#)

78 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWERS: B, F

- B** Create an AzureOpenAI client with the Foundry endpoint, key, and API version.
- F** Open the audio file in binary mode before sending the request.

To transcribe audio with a deployed speech-to-text model, the app creates an AzureOpenAI client with the Foundry endpoint, key, and API version. It then opens the audio file in binary mode and passes it to `client.audio.transcriptions.create` together with the model deployment name.

WHY THE OTHER CHOICES ARE WRONG

- **A.** `client.responses.create` with text input does not upload the spoken prompt audio file for transcription.
- **C.** A voice element in SSML controls synthesized speech output and does not send input audio to a model.
- **D.** `speech_synthesis_voice_name` configures text-to-speech output, not audio transcription.
- **E.** `SpeechRecognizer` is the Azure Speech SDK pattern, not the OpenAI SDK pattern for a deployed audio model.

Microsoft Learn: [Develop a speech-capable generative AI application: Transcribe speech](#) · [Azure OpenAI image and audio REST API reference \(2024-10-21\)](#)

79 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** `client.audio.speech.with_streaming_response.create`

The speech synthesis unit uses the AzureOpenAI client and calls `client.audio.speech.with_streaming_response.create` to generate speech and stream it to an output file. The request includes the deployed model, a voice, input text, and optional instructions for tone.

WHY THE OTHER CHOICES ARE WRONG

- **A.** `recognize_once_async` is an Azure Speech SDK speech recognition method, not a generative model speech output method.
- **C.** `recognize_entities` extracts text entities and does not create audio.
- **D.** `client.audio.transcriptions.create` converts audio to text, not text to speech.

Microsoft Learn: [Develop a speech-capable generative AI application: Synthesize speech](#) · [Foundry Models sold by Azure](#)

80 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** Realtime audio processing for speech-in and speech-out interactions

The model catalog documentation states that GPT-4o audio models support low-latency speech-in and speech-out conversational interactions or audio generation. For this prototype, you should select an audio model that supports realtime audio processing rather than a text-only or image-only model.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Named entity recognition analyzes text entities and is not a speech-in and speech-out model capability.
- **C.** Image generation creates visual output and does not handle spoken prompts or spoken answers.
- **D.** Embedding generation is for vectorizing text and does not provide a spoken conversation loop.

Microsoft Learn: [Foundry Models sold by Azure](#) · [Develop a speech-capable generative AI application: Choose a speech-capable model](#)

81 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: A

- A** SpeechRecognizer and recognize_once_async

The Speech to Text API unit shows that the SDK pattern creates a SpeechRecognizer from SpeechConfig and AudioConfig. The recognize_once_async method returns a SpeechRecognitionResult whose Text property contains the transcription when the Reason is RecognizedSpeech.

WHY THE OTHER CHOICES ARE WRONG

- **B.** SpeechSynthesizer and speak_text_async generate audio from text, not a transcript from audio.
- **C.** audio.speech.create is an OpenAI model speech output pattern, not the Azure Speech SDK recognition pattern.
- **D.** TextAnalyticsClient analyzes written text and does not transcribe audio files.

Microsoft Learn: [Create speech-enabled apps with Azure Speech in Microsoft Foundry Tools: Use the Speech to Text API](#) · [About the Speech SDK](#)

82 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: C, E

- C** SpeechConfig with the Foundry resource key and endpoint
- E** AudioConfig with the audio file name

The Azure Speech unit shows that SpeechConfig encapsulates the connection details for a Foundry resource, including endpoint or region and key. The Speech to Text API unit shows AudioConfig can define the input source, such as an audio file, before creating SpeechRecognizer.

WHY THE OTHER CHOICES ARE WRONG

- **A.** TextAnalyticsClient is for Azure Language text analysis, not Azure Speech recognition.
- **B.** AIProjectClient is not the Speech SDK object that defines the speech endpoint and audio input.
- **D.** SpeechSynthesizer and a voice name are used for text-to-speech, not speech-to-text input.
- **F.** AudioOutputConfig configures synthesized speech output, not an input file for transcription.

Microsoft Learn: [Create speech-enabled apps with Azure Speech in Microsoft Foundry Tools: Azure Speech in Foundry Tools](#) · [Create speech-enabled apps with Azure Speech in Microsoft Foundry Tools: Use the Speech to Text API](#)

83 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** speech_synthesizer.speak_text_async

The Text to Speech API unit shows that a SpeechSynthesizer object calls speak_text_async to convert text to spoken audio. The result reason is SynthesizingAudioCompleted when the service successfully generates the audio stream.

WHY THE OTHER CHOICES ARE WRONG

- **A.** speak_ssml_async is used when the input is SSML markup rather than plain text.
- **C.** detect_language identifies a text language and does not create spoken audio.
- **D.** recognize_once_async transcribes audio to text and does not synthesize speech.

Microsoft Learn: [Create speech-enabled apps with Azure Speech in Microsoft Foundry Tools: Use the Text to Speech API](#) · [What Is Azure Speech?](#)

84 Multiple choice, choose three · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, B, D

- A** Adjust prosody for speaking rate.
- B** Insert a break or silence element.
- D** Specify a phoneme for pronunciation.

SSML is an XML-based syntax for controlling text-to-speech output. Microsoft Learn states that SSML can insert pauses, specify phonemes for pronunciation, and adjust prosody such as pitch, volume, and speaking rate.

WHY THE OTHER CHOICES ARE WRONG

- **C.** A custom speech recognition model improves speech-to-text and is not an SSML feature for synthesis output.
- **E.** Language detection identifies the text language and does not control pauses, pronunciation, or rate.
- **F.** PII recognition redacts personal information and does not control speech output characteristics.
- **G.** Batch transcription converts recorded audio to text and does not configure synthesized narration.

Microsoft Learn: [Create speech-enabled apps with Azure Speech in Microsoft Foundry Tools: Use Speech Synthesis Markup Language](#) · [Speech Synthesis Markup Language \(SSML\) overview](#)

85 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** Voice Live API

The Voice Live API provides low-latency speech-to-speech interactions for voice agents through a single interface. The documentation explains that it integrates speech recognition, generative AI, and text-to-speech, which reduces the need to manually orchestrate those components.

WHY THE OTHER CHOICES ARE WRONG

- **A.** SpeechSynthesizer handles text-to-speech output but does not manage spoken input and model reasoning together.
- **B.** SpeechRecognizer handles speech-to-text recognition but does not provide the full real-time speech-to-speech agent interface.
- **C.** The Azure Language MCP server exposes text analysis tools and does not provide real-time voice conversations.

Microsoft Learn: [Voice Live API Overview](#) · [How to use the Voice Live API](#) · [Get started with speech in Azure: Creating a speech-capable agent](#)

86 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: B

- B** A vision-enabled chat model deployment

Vision-enabled chat models in Microsoft Foundry can accept image input with text and return natural language answers about the image. This supports scenarios such as image description, visual question answering, and document or screenshot analysis.

WHY THE OTHER CHOICES ARE WRONG

- **A.** An image generation model creates new images instead of answering questions about an uploaded image.
- **C.** A text embedding model produces vectors for semantic search and does not analyze image pixels in a prompt.
- **D.** A speech synthesis model creates audio output and does not interpret visual input.

Microsoft Learn: [Get started with computer vision in Azure: Multimodal models for image analysis](#) · [How to use vision-enabled chat models](#)

87 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** Convert the image to a base64 data URL and place it in the image URL field

The vision-enabled chat API accepts an image as a publicly accessible URL or as base64-encoded image data. For local files, the app can encode the file as a data URL and use that value in the image URL field.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The model cannot read a local file path from the client computer.
- **B.** Embedding vectors are used for semantic similarity and are not the documented way to send an image to a vision-enabled chat model.
- **D.** The system message is for instructions, not for attaching image bytes.

Microsoft Learn: [How to use vision-enabled chat models](#)

88 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, E

- A** Include both a text item and an image_url item in the user message content array
- E** Set max_tokens or max_completion_tokens to allow enough response tokens

A vision chat request can use a user message whose content array contains text and an image_url item. The documentation also notes that max_tokens or max_completion_tokens should be set so the model output is not cut off.

WHY THE OTHER CHOICES ARE WRONG

- **B.** The n_seconds parameter is used for video generation, not for chat completion length.
- **C.** The image generation endpoint creates images and is not used to analyze a submitted photo with a chat response.
- **D.** A text embedding model does not accept the photo as visual input for image analysis.

Microsoft Learn: [How to use vision-enabled chat models](#)

89 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

D Set image_url.detail to high

The image_url field can include a detail value of low, high, or auto. The high setting uses high resolution processing for more detailed interpretation, at the cost of more token usage.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The quality parameter applies to image generation, not to how a vision chat model reads an input image.
- **B.** The n parameter requests multiple generated images and does not affect vision chat detail.
- **C.** The output_format parameter controls generated image file format and does not improve image analysis.

Microsoft Learn: [How to use vision-enabled chat models](#)

90 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

C GPT-Image series

Microsoft Foundry image generation documentation lists current GPT-Image models for generating images from prompts. It also states that DALL-E 3 was retired and directs new image generation work to GPT-image series models.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The image generation documentation states that DALL-E 3 was retired and new deployments should use a GPT-image series model.
- **B.** A GPT-4.1 text chat deployment is for chat responses and is not an image generation model family.
- **D.** Whisper is an audio model and does not generate images from prompts.

Microsoft Learn: [How to Use Image Generation Models from OpenAI](#) · [Get started with computer vision in Azure: Image generation models](#)

91 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: B, E

B n

E size

In an image generation request, n sets how many images to return and size sets the output dimensions, such as 1024x1024. Parameters such as detail, max_tokens, and temperature apply to image analysis or text generation requests instead.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Temperature is a sampling parameter for text generation and does not specify the number or dimensions of generated images.
- **C.** The max_tokens parameter controls text response length for chat requests, not the count or size of generated images.
- **D.** The n_seconds parameter controls duration for video generation jobs.
- **F.** The detail parameter applies to image input analysis in vision chat requests.

Microsoft Learn: [How to Use Image Generation Models from OpenAI](#)

92 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** Decode the b64_json image data and write the bytes to an image file

The image generation response includes a data array with base64-encoded image data in b64_json. An app can decode that base64 value and write the resulting bytes to a PNG or JPEG file.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Bounding box coordinates are not the documented output of GPT-image generation calls.
- **B.** Polling is used for asynchronous video generation jobs, not for a synchronous image generation response.
- **D.** response.output_text is used for text output and is not the base64 image payload from the image generation API.

Microsoft Learn: [How to Use Image Generation Models from OpenAI](#)

93 Multiple choice, choose three · Implement AI solutions by using Microsoft Foundry

ANSWERS: B, C, D

- B** Set output_format to png
- C** Set background to transparent
- D** Call the image generation API with a prompt for the icon

The image generation API creates images from prompts by using a deployed GPT-image model. The documentation lists output_format and background parameters, with transparent backgrounds requiring PNG output for GPT-image-1.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Sora 2 is a video generation model and is not the correct deployment for still PNG icon generation.
- **E.** The image_url.detail setting controls how a vision chat model interprets an input image.
- **F.** The n_seconds parameter controls the duration of generated video, not still image transparency.

Microsoft Learn: [How to Use Image Generation Models from OpenAI](#)

94 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: A, D

- A** Deploy a Sora 2 model with the Azure OpenAI resource
- D** Create a video generation job, poll its status, and retrieve the video content after success

The Sora 2 video generation overview describes Sora 2 as the model for text to video, image to video, and video remix scenarios. Video generation is asynchronous, so an app creates a job, polls for completion, and then retrieves the generated video content.

WHY THE OTHER CHOICES ARE WRONG

- **B.** A vision-enabled chat completion request analyzes images and returns text, rather than generating video.
- **C.** The b64_json field is used for generated still images, not for Sora 2 video generation jobs.
- **E.** GPT-Image models generate or edit still images, not MP4 video clips.

Microsoft Learn: [Sora 2 video generation overview \(preview\)](#) · [Get started with computer vision in Azure: Video generation models](#)

95 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** A user input that contains input_text and input_image content items

With the Responses API, a user input can contain a content array with an input_text item for the question and an input_image item for the picture. The model processes both inputs and returns a natural language response that the app can display.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A speech synthesis request produces audio and does not analyze uploaded images.
- **B.** A file name alone does not provide the model with image data or a user question.
- **D.** The n, size, and quality parameters are used for image generation, not for asking questions about an uploaded image.

Microsoft Learn: [Get started with computer vision in Azure: Multimodal models for image analysis](#)

96 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** Provide a publicly accessible image URL or send the image as base64 data

The vision-enabled chat documentation states that image URLs must be publicly accessible and that private endpoint, VNet-restricted, and firewall-restricted URLs are not supported. An app can instead send base64 image data when a public URL is not appropriate.

WHY THE OTHER CHOICES ARE WRONG

- **A.** The detail setting changes image processing resolution and does not bypass storage network restrictions.
- **B.** Increasing max_tokens can affect response length but does not make a private image URL accessible to the service.
- **D.** Changing the prompt role does not allow the model to fetch a blocked image URL.

Microsoft Learn: [How to use vision-enabled chat models](#)

97 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: B, C

- B** The Azure OpenAI endpoint
- C** The model deployment name

For keyless authentication, the documentation lists the endpoint and deployment name as required resource information for the app. The token is obtained through Microsoft Entra credentials instead of configuring an API key in the app.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A video job ID is created after submitting a video generation request and is not used to call an image analysis deployment.
- **D.** A b64_json value is returned by image generation and is not configuration for a vision analysis call.
- **E.** An API key is not required when the app uses keyless Microsoft Entra authentication.

Microsoft Learn: [How to use vision-enabled chat models](#)

98 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

- C** A natural language response based on the text and image inputs

Vision-enabled chat models combine visual understanding with natural language processing. In an app scenario, the model can analyze the submitted image and return a text answer to the user's question.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Vision-enabled chat models can answer questions in natural language and do not require the app to consume object detection boxes.
- **B.** An MP4 download is returned after a successful video generation job, not after a vision chat request.
- **D.** Base64-encoded image data is returned by image generation models, not by vision chat analysis.

Microsoft Learn: [Get started with computer vision in Azure: Multimodal models for image analysis](#) · [How to use vision-enabled chat models](#)

99 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** prebuilt-invoice

Content Understanding includes domain-specific prebuilt analyzers for common document categories. The prebuilt-invoice analyzer is designed to extract structured invoice data, including values and nested line items, without requiring you to define a custom schema first.

WHY THE OTHER CHOICES ARE WRONG

- **A.** prebuilt-imageSearch generates image descriptions and insights, not invoice-specific document fields.
- **B.** prebuilt-audioSearch analyzes audio recordings, not invoice documents.
- **C.** prebuilt-documentSearch is optimized for document ingestion and retrieval scenarios, not domain-specific invoice fields.

Microsoft Learn: [Azure Content Understanding in Foundry Tools prebuilt analyzers](#) · [Get started with AI-powered information extraction in Azure: Extract information from documents](#)

100 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** An array field with object items and properties for product, quantity, and price

Content Understanding field schemas support nested fields through object and array types. For line items, an array whose items are objects lets the analyzer return each product with separate typed properties such as quantity and price.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A classify field selects from fixed categories and is not intended to return repeated product records.
- **B.** Boolean fields can indicate true or false values, but they cannot represent quantities, prices, and repeated item objects.
- **C.** A single string would not preserve the repeating item structure as typed child fields.

Microsoft Learn: [Azure Content Understanding in Foundry Tools - What is an analyzer? Configuration and reference](#) · [Get started with AI-powered information extraction in Azure: Extract information from documents](#)

101 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

D estimateFieldSourceAndConfidence

Document analyzers can opt in to confidence and source grounding for field extraction. Enabling estimateFieldSourceAndConfidence returns a confidence score and a source location so applications can route low-confidence values for human review.

WHY THE OTHER CHOICES ARE WRONG

- **A.** tableFormat controls whether extracted tables are returned as HTML or Markdown, not field confidence.
- **B.** enableFormula controls formula extraction in documents and does not return field source locations.
- **C.** omitContent removes the original content object from some responses and does not request confidence or source grounding.

Microsoft Learn: [Azure Content Understanding in Foundry Tools Document Overview](#) · [Azure Content Understanding in Foundry Tools - What is an analyzer? Configuration and reference](#)

102 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: C, E

C A DocumentKind field with type string, method classify, and enum values Receipt and Exception

E A ReceiptTotal field with type number and method extract

The extract method is used for literal values such as totals that should be taken from the document. The classify method is used when the field value must be selected from a fixed set of categories such as Receipt and Exception.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A generate field is useful for summaries or interpretations, but it does not classify the document or extract an exact total as the field appears.
- **B.** tableFormat is a document analyzer configuration option, not a field definition for extraction or classification.
- **D.** A boolean field cannot return the numeric receipt total.
- **F.** prebuilt-audio is a base analyzer for audio content, not document forms.

Microsoft Learn: [Azure Content Understanding in Foundry Tools - What is an analyzer? Configuration and reference](#) · [Create a custom analyzer with Azure Content Understanding in Foundry Tools](#)

103 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

C prebuilt-imageSearch

The prebuilt-imageSearch analyzer is designed for image content and generates descriptions and insights that support search and retrieval. An app submits an image URL to this analyzer and receives structured output that it can index.

WHY THE OTHER CHOICES ARE WRONG

- **A.** prebuilt-layout extracts document layout elements such as paragraphs and tables, not general image insights.
- **B.** prebuilt-invoice is a domain-specific document analyzer for invoices.
- **D.** prebuilt-audioSearch is for audio recordings and conversations, not still images.

Microsoft Learn: [Azure Content Understanding in Foundry Tools image overview](#) · Quickstart: [Azure Content Understanding in Foundry Tools](#)

104 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: A

A Use a document analyzer with a field extraction schema for the labeled form values

The image overview notes that image analyzers are not optimized for scenarios based primarily on extracted text. When the goal is to extract and analyze text from images, a document field extraction schema is the recommended approach.

WHY THE OTHER CHOICES ARE WRONG

- **B.** prebuilt-videoSearch processes video files and is not appropriate for still form photos.
- **C.** Image analyzers are not optimized when analysis is based primarily on extracted text from images.
- **D.** prebuilt-audioSearch processes audio files and cannot transcribe a still image.

Microsoft Learn: [Azure Content Understanding in Foundry Tools image overview](#) · [Get started with AI-powered information extraction in Azure: Extract information from documents](#)

105 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: D, E

D A ChartType field with type string, method classify, and enum values bar, line, and pie

E A Title field with type string

The custom image analyzer example defines a Title field and a ChartType field for chart and graph images. ChartType uses classify with an enum so the analyzer chooses one of the allowed chart categories.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Caller is a field for audio call scenarios and does not identify chart details.
- **B.** Key frames are returned for video analysis and are not needed for a still chart image.
- **C.** VendorName is relevant to invoice documents, not chart images.
- **F.** Transcripts are produced from audio or video speech, not from chart images.

Microsoft Learn: [Create a custom analyzer with Azure Content Understanding in Foundry Tools](#) · [Azure Content Understanding in Foundry Tools image overview](#)

106 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

D prebuilt-callCenter

The audio overview lists prebuilt-callCenter for post-call analysis. It returns conversation transcripts with speaker role detection, call summary, sentiment, mentioned articles, companies, people, and relevant categories.

WHY THE OTHER CHOICES ARE WRONG

- **A.** prebuilt-audioSearch generates conversation transcripts and a summary, but it is not the specialized post-call analyzer for role detection, sentiment, and categories.
- **B.** prebuilt-imageSearch analyzes still images, not call recordings.
- **C.** prebuilt-videoSearch analyzes videos and returns video-specific outputs such as key frames and segments.

Microsoft Learn: [Azure Content Understanding in Foundry Tools audio overview](#) · [Get started with AI-powered information extraction in Azure: Extract information from audio and video](#)

107 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: C

C locales

Audio analyzers use the locales configuration to control language processing during transcription. The documentation recommends configuring locales when you do not want to rely on multilingual transcription behavior.

WHY THE OTHER CHOICES ARE WRONG

- **A.** estimateFieldSourceAndConfidence applies to document field source and confidence output, not audio locale selection.
- **B.** tableFormat controls extracted document table output and does not configure audio transcription languages.
- **D.** segmentPerPage applies to document segmentation and does not affect audio transcription language handling.

Microsoft Learn: [Azure Content Understanding in Foundry Tools audio overview](#) · [Azure Content Understanding in Foundry Tools - What is an analyzer? Configuration and reference](#)

108 Multiple choice, choose two · Implement AI solutions by using Microsoft Foundry

ANSWERS: D, F

D Key frame times and key frame images

F Transcript phrases with timestamps

Video analysis can return transcripts with speaker phrases and timestamps so an app can locate spoken content. It can also return key frames and key frame times that provide visual context for video segments.

WHY THE OTHER CHOICES ARE WRONG

- **A.** Locale configuration controls transcription setup and is not a returned output for locating video content.
- **B.** Chart type enum values are part of a custom image schema and do not locate spoken instructions in videos.
- **C.** tableFormat is a document setting and is not a returned video output.
- **E.** Invoice line items are returned by invoice document analyzers, not by video analysis for visual context.

Microsoft Learn: [Azure Content Understanding in Foundry Tools video overview](#) · [Quickstart: Azure Content Understanding in Foundry Tools](#)

109 Multiple choice, choose three · Implement AI solutions by using Microsoft Foundry

ANSWERS: D, E, F

- D** enableSegment set to true with contentCategories that describe the desired segments
- E** A Sentiment field with method classify and allowed sentiment values
- F** baseAnalyzerId set to prebuilt-video

A custom video analyzer should inherit from the prebuilt-video base analyzer. Video segmentation is configured with enableSegment and contentCategories, and a classify field with enum values can return sentiment for each segment.

WHY THE OTHER CHOICES ARE WRONG

- **A.** tableFormat controls document table output and does not split or classify video segments.
- **B.** prebuilt-document is for document-based custom analyzers, not product demo videos.
- **C.** prebuilt-invoice is for invoice documents and cannot process product demo videos.
- **G.** estimateFieldSourceAndConfidence is a document field setting and does not configure video segmentation.

Microsoft Learn: [Azure Content Understanding in Foundry Tools video overview](#) · [Azure Content Understanding in Foundry Tools - What is an analyzer? Configuration and reference](#) · [Create a custom analyzer with Azure Content Understanding in Foundry Tools](#)

110 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** Send GET requests to the Operation-Location URL until the status is Succeeded

Content Understanding analysis is asynchronous for these REST operations. The response provides an Operation-Location URL, and the client polls that URL until the result status is Succeeded.

WHY THE OTHER CHOICES ARE WRONG

- **A.** A chat completions endpoint is not the Content Understanding analyze result endpoint.
- **B.** The 202 response starts an asynchronous operation and the completed fields are retrieved later from the operation result.
- **C.** Recreating the analyzer is not part of retrieving an existing analyze result.

Microsoft Learn: [Quickstart: Azure Content Understanding in Foundry Tools](#) · [Get started with AI-powered information extraction in Azure: Extract information from documents](#)

111 Multiple choice · Implement AI solutions by using Microsoft Foundry

ANSWER: D

- D** begin_analyze

In the Content Understanding Python SDK, ContentUnderstandingClient.begin_analyze starts an analysis with an analyzer ID and the input to analyze, such as a file URL. It returns a poller, and the app calls poller.result to read the output after the long-running operation completes.

WHY THE OTHER CHOICES ARE WRONG

- **A.** poller.result waits for a poller after an operation is started, but it does not submit content for analysis.
- **B.** begin_create_analyzer creates a custom analyzer and does not analyze a file with an existing analyzer.
- **C.** get_analyzer retrieves analyzer metadata and does not submit content for analysis.

Microsoft Learn: [Quickstart: Azure Content Understanding in Foundry Tools](#)

ANSWERS: A, D

- A** The analyzer ID in the analyze request URL
- D** An inputs array in the JSON body that contains the file URL

A REST client sends a POST request to the analyze endpoint of a specific analyzer, so the analyzer ID is part of the request URL. The JSON body includes an inputs array with the URL of the file that Content Understanding should analyze.

WHY THE OTHER CHOICES ARE WRONG

- **B.** Succeeded is a status returned by the result operation after processing completes, not a request value.
- **C.** tableFormat is a document analyzer configuration option and is not required to identify content for every media type.
- **E.** Operation-Location is returned by the service after the request is accepted, not supplied by the client to start analysis.
- **F.** Key frame URLs are returned after video analysis and are not known before submitting the file.

Microsoft Learn: [Quickstart: Azure Content Understanding in Foundry Tools](#)