AI-103 · CHEAT SHEET · CERTIFYDECK

Microsoft Certified: Azure AI Apps and Agents Developer Associate

Dominios, pesos oficiales y habilidades clave del temario de AI-103, para repasar antes del examen.

Nivel: Associate Duración: 120 min Nota de corte: 700/1000 Fuente: guía de estudio oficial de Microsoft Learn

1. Plan and manage an Azure AI solution

25–30%

Choose the appropriate Foundry services for generative AI and agents

- Choose an appropriate model for each task, including large language models, small language models, multimodal models, and Foundry Tools
- Choose the appropriate Foundry services for generative tasks, grounding, vector search, agent workflows, or multimodal processing
- Choose an appropriate method for retrieval and indexing
- Choose appropriate memory, tool, and knowledge integration services for agent solutions

Set up AI solutions in Foundry

- Design Azure infrastructure for AI apps and agent-based solutions
- Choose appropriate deployment options
- Configure model and agent deployments
- Integrate Foundry projects with continuous integration and continuous deployment pipelines

Manage, monitor, and secure AI systems

- Manage quotas, scaling, rate limits, and cost footprints for model and agent workloads
- Monitor model performance, drift, safety events, and grounding quality
- Monitor data ingestion quality, search index health, and relevance performance
- Configure security, including managed identity, private networking, keyless credentials, and role policies

Implement responsible AI across generative AI and agentic systems

- Configure safety filters, guardrails, risk detection, and content moderation
- Apply responsible AI instrumentation, including evaluators, safety evaluations, and explanation tooling
- Implement auditing through trace logging, provenance metadata, and approval workflows
- Govern agent behavior with oversight modes, constraints, and tool-access controls

2. Implement generative AI and agentic solutions

30–35%

Build generative applications by using Foundry

- Deploy and consume large language models, small models, code models, and multimodal models
- Implement retrieval-augmented generation in an application
- Design workflows, tool-augmented flows, and multistep reasoning pipelines
- Evaluate models and apps, including detecting fabrications, relevance, quality, and safety
- Integrate generative workflows into applications by using Foundry SDKs and connectors

Build agents by using Foundry

- Define agent roles, goals, conversation-tracking approach, and tool schemas
- Build agents that integrate retrieval, function calling, and conversation memory
- Integrate agent tools, including APIs, knowledge stores, search, content understanding, and custom functions
- Implement orchestrated multi-agent solutions
- Build autonomous or semiautonomous workflows with safeguards and approval flow controls

Optimize and operationalize generative AI systems

- Tune generation behavior, such as prompt engineering and adjusting model parameters
- Implement model reflection, chain-of-thought evaluations, and self-critique loops
- Set up observability by implementing tracing, token analytics, safety signals, and latency breakdowns
- Orchestrate multiple models, flows, or hybrid LLM and rules engines

3. Implement computer vision solutions

10–15%

Design and implement image- and video-generation solutions

- Implement a solution that generates images from text prompts and reference media
- Implement a solution that generates videos from text prompts and reference media
- Configure image-editing workflows, including inpainting, mask-based edits, and prompt-driven modifications
- Implement workflows to edit generated videos
- Select and apply appropriate generation and editing controls provided by the platform

Design and implement multimodal understanding workflows

- Build a solution that analyzes visual context by using multimodal models
- Configure apps to produce concise or detailed captions for single or multiple images
- Implement a solution that enables question answering grounded in visual evidence
- Configure generation of alt text and extended image descriptions aligned to accessibility guidelines
- Implement visual understanding by configuring Azure Content Understanding in Foundry Tools to extract visual characteristics

Implement responsible AI for multimodal content

- Implement filters to classify unsafe or disallowed visual content
- Detect and mitigate indirect prompt injection by using embedded text in images
- Enforce visual policy rules, including watermarks, prohibited symbols, brand requirements, and inappropriate-content detection

4. Implement text analysis solutions

10–15%

Apply language model text analysis

- Extract entities, topics, summaries, and structured JSON outputs by using generative prompting and Foundry Tools
- Configure detection of sentiment, tone, safety issues, and sensitive content
- Build solutions that translate text by using Azure Translator in Foundry Tools or language-model-powered translation flows
- Customize language model outputs for domain tasks, such as compliance summarization and domain extraction

Implement speech solutions

- Convert speech to text and text to speech for agentic interactions
- Integrate speech as an agent modality, including custom speech models
- Enable multimodal reasoning from audio inputs
- Translate speech into other languages by using language models and Foundry Tools

5. Implement information extraction solutions

10–15%

Build retrieval and grounding pipelines

- Ingest and index content, such as documents, images, audio, and video
- Configure semantic search, hybrid search, and vector search for grounding
- Implement enrichment by using custom or built-in skills for text, images, and layout
- Configure retrieval-augmented generation ingestion flow, including documents and optical character recognition
- Connect retrieval pipelines directly to workflows and agent tools

Extract content from documents

- Extract information by using multimodal pipelines that combine optical character recognition, layout analysis, and field extraction
- Produce clean, grounded representations for agents and retrieval-augmented generation by using Content Understanding
- Implement analyzers for structured or markdown outputs by using Content Understanding