

Training for AI Engineer Interns – Notes (Sample)

Includes multiple sections to test semantic retrieval and citations.

Deployment checklist

- Create DeepSeek-R1 (chat) deployment and text-embedding-3-small (embeddings) deployment in Azure AI Foundry.
- Set AZURE_AI_ENDPOINT to <https://services.ai.azure.com/models> (must end with /models).
- Run healthcheck: embeddings ping, chat ok, MongoDB \$vectorSearch.

RAG answer policy

- Answer using ONLY the retrieved context.
- If context is insufficient, say: The answer was not found in the documents.
- Include citations: (source#chunk_id).

Troubleshooting

- 401 Unauthorized: endpoint/key mismatch.
- 404 Not Found: wrong base endpoint (missing /models) or wrong deployment name.
- Mongo error: vector index missing or numDimensions mismatch (e.g., 1536).

Mini FAQ (Chinese-English mix) – Sample

Q: █ /models █ endpoint█

A: [REDACTED] /chat/completions ■ /embeddings■

Q: 

A: [REDACTED] source [REDACTED] chunks [REDACTED] !delete manual.pdf [REDACTED]