**Azure AI Foundry**

**What is Azure AI Foundry?**

**Azure AI Foundry** is **Microsoft’s end-to-end platform** to build, deploy, and scale **Generative AI (GenAI) solutions** responsibly. It combines **Azure AI Studio**, **Azure OpenAI**, **Prompt Flow**, **ML Ops**, and **responsible AI tools** into one cohesive ecosystem to accelerate GenAI adoption in enterprises.

**Step-by-Step Setup for a GenAI Solution in Azure AI Foundry**

**Step 1: Define Your Business Use Case**

Before starting:

* Example use cases: Q&A bot, document summarization, product recommendation.
* Identify data sources: PDFs, databases, product catalogs, etc.

**Step 2: Set Up Azure Resources**

Provision required services:

1. **Azure AI Studio** – Central workspace for building AI solutions.
2. **Azure OpenAI Service** – For GPT, Codex, and DALL·E models.
3. **Azure AI Search** – For implementing Retrieval-Augmented Generation (RAG).
4. **Azure Blob Storage** – Store documents (e.g., PDFs, JSON, CSV).
5. **Azure Cosmos DB or Vector DB** – Store embeddings.
6. **Azure Key Vault** – Manage API keys and secrets.

**Step 3: Ingest & Prepare Data**

Use tools like:

* **Data Factory / Azure Synapse** – to ingest data.
* Store it in **Blob Storage** or **Cosmos DB**.

*Example:* Upload product data to Blob or documents for Q&A bots.

**Step 4: Create Embeddings (Vector Index)**

Use **Azure OpenAI Embedding Models** (e.g., text-embedding-ada-002) to:

* Convert documents or product descriptions into vector form.
* Store embeddings in Azure AI Search or Cosmos DB with vector search support.

**Step 5: Build Prompt Flow (in Azure AI Studio)**

Prompt Flow allows you to:

* **Design** and **chain prompts, Python logic**, and **tools**.
* Implement **RAG pipelines** visually and with code.

Sample Flow:

User Query → Vector Search (Azure AI Search) → Relevant Docs → LLM (GPT) → Response

**Step 6: Fine-Tuning or Customization (Optional)**

If needed:

* Use Azure OpenAI to fine-tune models with domain-specific data.
* Store fine-tuned models in Azure AI Studio.

**Step 7: Test & Evaluate**

* Use built-in **Prompt Flow Evaluator** to measure model performance.
* Use Azure tools to assess accuracy, hallucination, latency, etc.

**Step 8: Deploy the Solution**

* Convert your pipeline to **real-time API**.
* Use **Azure App Service / Azure Functions** for deployment.
* Use **APIM + Key Vault** for secured endpoint exposure.

**Step 9: Monitor, Secure & Govern**

* Use **Azure Monitor** and **Application Insights** for logs/metrics.
* Apply **Responsible AI Dashboard** for transparency and fairness.
* Ensure **role-based access control (RBAC)** is in place.

**Summary Diagram**

[Data Source] → [Blob / Cosmos] → [Embedding Generator (OpenAI)]

→ [Vector DB (AI Search)] → [Prompt Flow (LLM)]

→ [Response/API] → [Monitor & Evaluate]