

課程名稱：17800_如何打造最佳的 Azure OpenAI 應用體驗：

正確理解 Prompt Design、Fine-tuning、RAG 課程摘要

本摘要由 Gemini AI 協助生成，僅供參考

1. 適合對象與能力門檻

- **適合族群：**AI 應用開發者、軟體工程師、企業系統架構師以及想在企業內部導入 AI 應用的技術人員。課程內容專注於介紹如何運用 Azure OpenAI 服務建構最佳的應用體驗與選擇對應的技術。
- **能力門檻：**普通。建議具備基本的 API 呼叫概念、基本程式開發能力，以及對生成式 AI 與大型語言模型（LLM）的基本認知。

2. 課程核心大綱與架構圖

- **章節架構：**
 - **章節一：Prompt Design（提示詞設計）** — 學習提示詞設計原則、拆解複雜任務、邏輯結構化以及 Chain of Thought（思維鏈）概念。
 - **章節二：Tokens 運算與成本估算** — 理解 Token 計算方式、換算倍率（如 750 英文字約 1000 Tokens）與成本估算工具。
 - **章節三：Fine-tuning（模型微調）** — 了解 Fine-tuning 的原理、適用場景、資料準備格式（JSONL）與執行考量。
 - **章節四：RAG Pattern（檢索增強生成）** — 掌握結合 Azure Cognitive Search 與 Azure OpenAI 實現企業內部資料精準問答的架構。
 - **章節五：Embeddings（向量化與語意搜尋）** — 學習將文字轉為向量矩陣、Cos Similarity（餘弦相似度）檢索與語意搜尋的應用。

3. 必備前置知識清單

- **基礎程式開發與 API 呼叫概念：**需瞭解如何透過 HTTP / REST API（如 Chat Completions API 或 Completions API）進行資料傳輸與串接。

- **自然語言處理（NLP）基本認知：**瞭解大型語言模型（LLM）運作機制與基本對話互動模式。

4. 專有名詞與概念辭典

- **Prompt Design（提示詞設計）：**透過給予大型語言模型特定的提示指令，引導模型精準完成任務，免去大量微調成本並能節省 Token 消耗。
- **Tokens（詞元）：**語言模型用來理解與處理文字的最小單位（Subword）。英文約 750 字等於 1000 Tokens，是 API 計費與請求長度限制的核心單位。
- **Fine-tuning（微調）：**基於預訓練好的大模型，調整其文字產生的權重，使其回答符合特定格式或領域風格，非從頭重新訓練模型。
- **RAG（Retrieval-Augmented Generation，檢索增強生成）：**結合「資訊檢索」與「生成式模型」，讓 AI 根據企業私有真實資料回答問題，避免產生幻覺。
- **Embeddings（向量化）：**將字串轉換成高維度的數值向量（如 1024 維度的浮點數），用來代表文字的語意，以進行相似度比對與語意搜尋。

5. 課程學習職能對照

- **對應職能名稱：**AI 應用開發工程師 / 雲端解決方案架構師（Azure Cloud Solution Architect）
- **主要工作任務/能力指標：**
 - **分析與決策能力：**能評估業務場景需求，精準選擇使用 Prompt Design、Fine-tuning 或 RAG 模式來規劃 AI 應用架構。
 - **工具操作與執行能力：**能操作 Azure OpenAI 服務與 Azure Cognitive Search 建立檢索機制，並能準備 JSONL 訓練資料與進行向量（Embeddings）相似度比對檢索。