

課程名稱：17797_Generative AI Foundations on AWS

Part 2 Picking the right foundation model 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：**初學者與 AI / 機器學習工程師。本課程提供基層基石模型（Foundation Models）選型與評估的架構觀念，並透過實作手把手教學，非常適合欲在企業或專案中導入基石模型的開發者。
- **能力門檻：**入門至普通。預期具備基本的 Python 程式設計基礎，以及對機器學習基礎概念的基本理解。

2. 課程核心大綱與架構圖

- **章節架構：**
 - **章節一：基石模型選型四大維度與權衡** — 剖析模態（語言與視覺）、生成需求、模型參數規模與準確度/成本/推論時間（Inference Runtime）之間的權衡，以及開源與專有模型的優缺點。
 - **章節二：基準測試與外部 Evaluation** — 介紹如何善用外部評測榜單（如 Hugging Face Open LLM Leaderboard、Stanford HELM）並建立屬於自家業務情境的 Unit Tests 與 Prompt 資料庫。
 - **章節三：實務手把手操作（Hands-on Walkthrough）** — 透過 AWS SageMaker Studio、Boto3 SDK 調用 Falcon 40B 模型，搭配 CNN/Daily Mail 資料集計算 ROUGE 指標進行文本摘要評估。

3. 必備前置知識清單

- **Python 與 SDK：**需熟悉基本 Python 語法，並了解如何使用 boto3 與雲端服務交互。
- **基礎機器學習概念：**了解基本模型評估指標（如 Precision, Recall, F1-score）與模型推論（Inference）的基本流程。

- 雲端平台操作：對 AWS 基礎服務及 SageMaker / Jupyter Notebook 環境有基本概念。

4. 專有名詞與概念辭典

- **基石模型 (Foundation Models)**：經過巨量資料預訓練的大型模型（如 GPT、BERT、Stable Diffusion），可作為多種下游任務（如分類、摘要、生成）的通用基礎。
- **模態 (Modality)**：指資料的形式或種類，例如文本 (Language) 與影像 (Vision)。不同模態會決定模型的輸入與輸出型態。
- **ROUGE 指標 (ROUGE Metric)**：一種常用於評估文字摘要品質的自動化指標，透過比較模型產生的文字與標準答案 (Ground Truth) 之間的字詞重疊度進行打分。

5. 課程學習職能對照

- **對應職能名稱**：AI 應用工程師 / 資料科學家 / 機器學習工程師
- **主要工作任務/能力指標**：
 - [分析與決策能力] 能根據業務情境的靈活度、成本、硬體限制與精度需求，選擇最佳的基石模型（開源 vs. 專有、規模大小）。
 - [工具操作與執行能力] 能在 AWS SageMaker 環境部署模型 endpoint，並整合 Hugging Face Evaluate 程式庫執行模型的品質評測（如 ROUGE 運算）。