

課程名稱：17827_Transforming AI 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 研發人員、軟體工程師、AI 產業分析師與科技創業者。原因在於內容深度探討 Transformer 模型的誕生背景、架構設計邏輯（如 RNN 與 Attention 的比較）、資料效率、合成資料以及 AI 未來的發展趨勢。
- **能力門檻：**進階。建議需具備深度學習（Deep Learning）、神經網絡架構（如 RNN, LSTM, Transformer）、GPU 算力與大語言模型（LLM）訓練的基本知識，才能深刻理解對談中的技術細節與幽默。

2. 課程核心大綱與架構圖

- **章節架構：**
 - **章節一：計算典範轉移與 AI 工廠的興起** — 探討從傳統電腦架構（如 1964 年 IBM System 360）到加速計算（Accelerated Computing）與生成式 AI（Generative AI）作為新型態工廠的演進。
 - **章節二：Attention Is All You Need 的誕生與命名由來** — Transformer 團隊回顧當時為解決機器翻譯（Machine Translation）及 RNN/LSTM 訓練痛點而設計自注意力機制（Self-Attention）的過程與命名趣事。
 - **章節三：模型的架構演進與資料瓶頸（Data Bottleneck）** — 討論目前 Transformer 模型過度依賴大數據的現況、合成資料（Synthetic Data）、推理階段計算（Reasoning-time compute）與高品質資料的重要性。
 - **章節四：Transformer 創始成員的創業與未來展望** — 八位作者分享各自在生物科技（Inceptive）、去中心化網路（Near Protocol）、企業級 LLM（Cohere）、Character.AI 及 Sakana AI 等領域的創業實踐與對下一代 AI（如 SSM、循環架構）的看法。

3. 必備前置知識清單

- **深度學習基礎架構（RNN / LSTM / Transformer）**：需理解傳統循環神經網路的訓練瓶頸與 Transformer 注意力機制的優勢。
- **反向傳播與梯度下降（Gradient Descent）**：理解模型如何透過數據進行參數優化。
- **GPU 與加速計算（Accelerated Computing）**：了解硬體算力（如 NVIDIA GPUs / DGX）如何支撐大模型訓練與超級電腦的構建。

4. 專有名詞與概念辭典

- **Transformer（轉換器模型）**：一種基於自注意力機制的深度學習架構，能夠並行處理數據並捕捉長距離依賴關係，取代了傳統 RNN 的順序處理，成為當今大語言模型（LLM）的基石。
- **Generative AI Factory（生成式 AI 工廠）**：黃仁勳提出的概念，將資料（Data）與能源（Energy）輸入給 GPUs / Dynamo 機器，進而產出具有價值的 AI 數據與軟體應用。
- **RAG（Retrieval-Augmented Generation / 檢索增強生成）**：在對談中提到作為外部知識檢索與模型結合的機制，補足模型內部參數記憶的不足。
- **SSM（State Space Model / 狀態空間模型）**：被討論作為可能補充或挑戰 Transformer 的新架構方向，試圖解決長文本處理與記憶體效率問題。

5. 課程學習職能對照（對標經濟部產業職能基準/iPAS）

- **對應職能名稱**：AI 演算法工程師 / 資料科學專家 / AI 架構師
- **主要工作任務/能力指標**：
 - **架構評估與優化能力**：能分析不同深度學習模型（RNN vs. Transformer vs. SSM）的優缺點與算力消耗。
 - **前沿技術洞察力**：理解大模型訓練中的資料質量、推理計算（Reasoning）與合成資料的應用趨勢。