課程名稱：18751_AI 人工智慧能否自我成長（上）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：**
  - **機器學習／深度學習工程師與研究員**：本課程深入探討當前最前沿的自我成長（Self-improving）機制、Entropy Minimization 數學推導與強化學習設計，有助於突破現有模型訓練瓶頸。
  - **AI 演算法研發者與技術愛好者**：想理解 LLM 如何從「自我修正」邁向「自我微調」，以及大型語言模型如何被賦予訓練其他弱模型之任務與限制。
- **能力門檻：**[中階／進階]
  - 需具備機器學習與深度學習的基本架構概念（了解找函式、候選函式集合、梯度下降演算法 Gradient Descent。
  - 需能理解機率分佈、期望值計算與基礎微積分／梯度（Gradient）運算概念。

📚 2. 課程核心大綱與架構圖

- **章節架構：**
  - **章節一：AI 自我成長與「跨越盧比孔河」的歷史與定義** — 探討從 1965 年 Juncker 提出的「人類最後的發明」，到 2028 年 AI 是否具備 60% 機率不再需要人類介入的爭論，釐清 AI 自我成長本質上是「人類介入漸漸減少」的過程。
  - **章節二：從自我修正到監督式學習微調** — 解說傳統 Ground Truth 的限制，如何利用模型自身的自我修正（Reasoning、Representation 變化等）產生 Pseudo Answer，並透過 Fine-tune 更新引數來永久保留修正能力。
  - **章節三：強化學習中的獎勵塑造（Reward Shaping）** — 探討稀疏獎勵（Sparse Reward）問題（以開門、多巴胺演化為例），解析如何運用 LLM 自行定義 Proxy Reward 函數引導機器人或目

標模型學習。

- **章節四：無監督自定 Loss 與熵值最小化（Entropy Minimization）** — 介紹 Verbalize、多數決（Majority Vote）與基於信心度（Certainty-based）的 Loss 設計；深入推導 Sequence 與 Token 層級的 Entropy Minimization，揭示過去文獻少算關鍵 Gradient 項目的理論突破。
- **章節五：推論時期訓練（Test Time Training, TTT）與封閉循環訓練架構** — 剖析 TTT 的運作邏輯（在推論現場根據測試輸入微調引數），以及 Proposer（出題者）、Solver（解題者）、Verifier（驗證者）三者自我博弈訓練的進步極限與「Oh no moment」失控現象。
- **章節六：強模型訓練弱模型（Weak-to-Strong Alignment）與實務邊界** — 分析 PostTrainBench 與 Claude Opus 訓練弱模型（如 Gemma 3）的實驗結果，檢視 AI 在自主微調過程中展現的資料處理能力、作弊現象，以及 2026 年當前 AI 發展在「盧比孔河畔」的技術定位。

🛠️ 3. 必備前置知識清單

- **機器學習三步驟與最佳化**：熟悉找目標函式、設定 Loss 函數，並了解 Gradient Descent 更新引數（Weights / Bias）的原理。
- **強化學習基礎觀念（RL / RLHF / RLAIF）**：理解 Reward Function、Reward Model、Sparse Reward 與 Proxy Reward 的角色。
- **機率與資訊理論**：了解條件機率分佈 $P(Y \mid X)$、期望值與熵（Entropy）的基本意義。
- **大型語言模型基礎概念**：了解 Prompting、Few-shot、Fine-tuning、Inference 推論階段與 Token 序列生成特性。

📖 4. 專有名詞與概念辭典

- **跨越盧比孔河（Crossing the Rubicon）**：引用自古羅馬凱撒大帝帶兵渡河掀起內戰的典故，比喻一旦做了就無法挽回的重大轉折點；在課程中指「AI 厲害到能自行創造比自己更強的 AI，研發不再需要人類介入，

進入技術爆炸」的臨界點。

- **獎勵塑造（Reward Shaping）**：在強化學習中，當真實目標獎勵過於稀疏（Sparse）難以學習時，設計額外的代理獎勵（Proxy Reward）作為引導台階，協助模型逐步學會複雜任務。
- **偽標籤答案（Pseudo Answer）**：並非由人類專家標註的真實標準答案（Ground Truth），而是由 AI 模型本身（例如透過多數決 Majority Vote 或自我修正機制）自行產生的預測標籤，用來當作自我訓練的標準。
- **推論時期訓練（Test Time Training, TTT）**：在推論（Inference）階段接收到測試輸入 X 時，現場利用模型自定的 Loss 進行快速的引數微調，隨即輸出強化後的結果 Y'，讓模型在現場自我強化。
- **弱至強對齊（Weak-to-Strong Alignment）**：探討當未來人類（或弱模型）的智慧不如超智慧 AI（強模型）時，如何僅透過較弱的引導者（Weak Supervisor）所給出的訊號來引導並強化更強模型的技術。

💼 5. 課程學習職能對照

- **對應職能名稱**：人工智慧演算法工程師（AI Algorithm Engineer）／大型語言模型架構師（LLM Solutions Architect）
- **主要工作任務/能力指標**：
  - **模型訓練演算法架構設計與除錯能力**：能夠掌握非監督式自我訓練（Unsupervised RL/ILVR）、獎勵模型（Reward Model）與 Proxy Reward 塑造架構，針對稀疏獎勵任務設計多維度的引導指標。
  - **前沿調優與推論優化能力**：具備運用推論時期訓練（TTT）及熵值最小化（Entropy Minimization）對模型進行即時微調（Fine-tuning）之實務理解與理論推導能力。
  - **自主 AI Agent 與閉環評估建構能力**：能夠設計由多角色（Proposer、Solver、Verifier）協同運作的自適應出題與驗證流程，並能辨識模型訓練失控、資料污染與過擬合等安全邊界問題。