

演講內容

本場演講先回顧資料處理與人工智慧的發展脈絡：從早期以「程式」明確指示電腦行為，到以演算法從大量資料中學習模式，再發展到今日能主動生成文字與影像的生成式 AI。

在模型面向，教授介紹深度生成模型的分類，包括以 VAE、Boltzmann Machine 等為代表的模型，以及以 GAN、GSN 等為代表之模型；接著說明 Transformer 與 GPT 系列模型的演進，從 GPT、GPT-2 到具多能力的 GPT-4，並談論對未來 GPT-5 期望提升推理能力與降低錯誤

教授在演講中特別點出，目前生成式 AI 最大的問題之一，是經常產生看似合理、實際上卻錯誤或無依據的內容「幻覺」(hallucination)。大型語言模型輸出在語言上流暢、語法正確，但其事實內容與真實世界不符，或缺乏可查證證據的情形，模型架構以「預測下一個 token」為目標，追求的是統計上最可能的字串，而非事實正確；在推論階段，為了給出連貫回覆，模型傾向「合理猜測」，而不是選擇沉默或表達不確定。

心得

透過本次演講，我對生成式 AI 的整體輪廓有較系統性的理解：從 GAN、VAE 等生成模型，到 GPT 系列與異質運算平台的關聯，不再只停留在「會聊天、會寫程式」的印象，而其背後原因牽涉演算法、硬體與系統整合的多層面技術。

結合理論研究對 hallucination 的分析，我開始反省自己使用生成式 AI 的方式。未來在課業或專題中，我會將這 CHATGPT 定位為「協助與整理」的工具，而非直接複製貼上的答案來源。

五、參考文獻

A. Alansari, H. Luqman, A Comprehensive Survey of Hallucination in Large Language Models: Causes, Detection, and Mitigation, arXiv:2510.06265, 2025.

S. M. T. I. Tonmoy et al., A Comprehensive Survey of Hallucination Mitigation Techniques in Large Language Models, arXiv:2401.01313, 2024.