2.1 生成式 AI 的演進與核心技術

2.1.1生成式人工智慧的歷史演進

生成式人工智慧（Generative AI）的發展標誌著技術範式的重大轉型，由傳統的模式識別與數據分析，轉向具備創造文本、圖像及音訊內容之能力。研究指出，此演進歷程呈現清晰的軌跡：從早期基於規則（rule-based）與統計機率模型的生成系統，逐步過渡至深度生成模型，最終演進為當代的大型基礎模型（He et al., 2025; Sengar et al., 2024）。在 2014 年至 2017 年間，變分自編碼器（VAEs）與生成對抗網絡（GANs）的出現，透過編碼器—解碼器架構或對抗學習機制，實現了高度寫實的內容生成（Bengesi et al., 2023; Kılınç & Keçecioglu, 2024）。隨後，自回歸模型、流模型（flow-based models）及擴散模型（diffusion models）進一步提升了生成的保真度與多樣性（Trigka & Dritsas, 2025）。自 2017 年 Transformer 架構問世以來，生成式技術已從特定領域模型轉向更具通用性的基礎模型與大語言模型（Hagos et al., 2024; Zhou, 2025）。

2.1.2.Transformer 核心機制：自注意力系統

當代生成式 AI 的突破核心在於 Transformer 架構，其以自注意力（Self-attention）機制徹底取代了傳統的循環架構（Vaswani et al., 2017; Hagos et al., 2024）。自注意力機制允許模型並行計算序列中各個標記（tokens）之間的關聯強度，藉此精準捕捉長距離的依賴關係與上下文語境資訊（Sengar et al., 2024; Zhou, 2025）。為解決並行處理中缺失的序列資訊，位置編碼（Positional encoding）技術，如旋轉位置嵌入（RoPE），被引入以有效整合位置資訊（Su et al., 2021; Bengesi et al., 2023）。藉由多層自注意力模組與前饋網路（feed-forward networks）的堆疊，模型展現了極強的表達能力，能應對高度複雜且稀疏的序列數據建模（Wang & Weinan, 2024; Ferraris et al., 2025）。顯見，這種架構上的創新為後續大規模預訓練提供了理論與技術基礎（Vaswani et al., 2017）。

2.1.3.大語言模型（LLM）的崛起、應用與轉型路徑

大語言模型（LLM）作為生成式 Transformer 的延伸，透過在海量語料庫上的大規模預訓練，展現出卓越的語義理解與內容產出能力（Matarazzo & Torlone, 2025; Zhou, 2025）。研究顯示，隨參數規模、數據量及計算資源的增加，模型效能遵循「規模定律」（Scaling laws）而提升，進而具備跨任務的通用解決能力（He et al., 2025; Shen et al., 2024）。當前 LLMs 已從單純的文本生成，擴展至醫療、智慧製造、商務管理及 6G 無線系統等多元應用領域（Kusiak, 2024; Linkon et al., 2024）。雖然仍面臨偏見、幻覺及隱私等技術挑戰，但透過精調（fine-tuning）與人類回饋強化學習（RLHF）等對齊技術，模型已能更安全且精準地執行下游任務（Bengesi et al., 2023; Myers et al., 2023）。顯見，這些技術能力的提升，為後續中小企業（SMEs）的數位轉型提供了關鍵的技術路徑。

參考文獻（References）

Bengesi, S., El-Sayed, H., Sarker, M., Houkpati, Y., Irungu, J., & Oladunni, T. (2023). Advancements in generative AI: A comprehensive review of GANs, GPT, autoencoders, diffusion model, and transformers. *IEEE Access, 12*, 69812-69837. https://doi.org/10.1109/access.2024.3397775

Çelik, A., & Eltawil, A. (2024). At the dawn of generative AI era: A tutorial-cum-survey on new frontiers in 6G wireless intelligence. *IEEE Open Journal of the Communications Society, 5*, 2433-2489.

https://doi.org/10.1109/ojcoms.2024.3362271
Ferraris, A., Audrito, D., Di Caro, L., & Poncibò, C. (2025). The architecture of language: Understanding the mechanics behind LLMs. *Cambridge Forum on AI: Law and Governance*. https://doi.org/10.1017/cfl.2024.16
Hagos, D., Battle, R., & Rawat, D. (2024). Recent advances in generative AI and large language models: Current status, challenges, and perspectives. *IEEE Transactions on Artificial Intelligence, 5*, 5873-5893. https://doi.org/10.1109/tai.2024.3444742
He, R., Cao, J., & Tan, T. (2025). Generative artificial intelligence: a historical perspective. *National Science Review, 12*. https://doi.org/10.1093/nsr/nwaf050
Kılınç, H., & Keçecioglu, Ö. (2024). Generative artificial intelligence: A historical and future perspective. *Academic Platform Journal of Engineering and Smart Systems*. https://doi.org/10.21541/apjess.1398155
Kusiak, A. (2024). Generative artificial intelligence in smart manufacturing. *Journal of Intelligent Manufacturing, 36*, 1-3. https://doi.org/10.1007/s10845-024-02480-6
Linkon, A., Sarker, M., Nabi, N., Rana, M., Ghosh, S., Rahman, M., Esa, H., & Chowdhury, F. (2024). Advancements and applications of generative artificial intelligence and large language models on business management: A comprehensive review. *Journal of Computer Science and Technology Studies*. https://doi.org/10.32996/jcsts.2024.6.1.26
Matarazzo, A., & Torlone, R. (2025). A survey on large language models with some insights on their capabilities and limitations. *ArXiv*. https://doi.org/10.48550/arxiv.2501.04040
Myers, D., Mohawesh, R., Chellaboina, V., Sathvik, A., Venkatesh, P., Ho, Y., Henshaw, H., Al-Hawawreh, M., Berdik, D., & Jararweh, Y. (2023). Foundation and large language models: Fundamentals, challenges, opportunities, and social impacts. *Cluster Computing, 27*, 1-26. https://doi.org/10.1007/s10586-023-04203-7
S. B., Chirchi, V., Kadry, S., Agoramoorthy, M., P, G., K, S., & A. S. (2024). The road ahead: Emerging trends, unresolved issues, and concluding remarks in generative AI—A comprehensive review. *International Journal of Intelligent Systems*. https://doi.org/10.1155/2024/4013195
Sengar, S., Hasan, A., Kumar, S., & Carroll, F. (2024). Generative artificial intelligence: A systematic review and applications. *Multimedia Tools and Applications, 84*, 23661-23700. https://doi.org/10.1007/s11042-024-20016-1
Shen, X., Li, D., Leng, R., Qin, Z., Sun, W., & Zhong, Y. (2024). Scaling laws for linear complexity language models. *ArXiv*. https://doi.org/10.48550/arxiv.2406.16690
Su, J., Lu, Y., Pan, S., Wen, B., & Liu, Y. (2021). RoFormer: Enhanced Transformer with rotary position embedding. *ArXiv*. https://doi.org/10.1016/j.neucom.2023.127063
Trigka, M., & Dritsas, E. (2025). The evolution of generative AI: Trends and applications. *IEEE Access, 13*, 98504-98529. https://doi.org/10.1109/access.2025.3574660
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998-6008.
Wang, M., & Weinan, E. (2024). Understanding the expressive power and mechanisms of transformer for sequence modeling. *ArXiv*. https://doi.org/10.48550/arxiv.2402.00522
Zhou, Z. (2025). Large language models and their evolution. *Applied and*

Computational Engineering. https://doi.org/10.54254/2755-2721/2025.25586