

## 從 GPT-3.5 到 GPT-5.0 的體驗與反思

周柏穎營養師 姜博文診所

### 前言

近年來，人工智慧 (Artificial Intelligence, AI) 技術以驚人的速度發展，特別是在語言模型 (Large Language Model, LLM) 領域。自 GPT-3.5 問世以來，AI 在語言理解與生成方面的能力已大幅提升，應用範圍涵蓋教育、研究、醫療與程式設計等多領域。以下將分享自 GPT-3.5 至 GPT-5.0 的實際使用心得，聚焦於程式撰寫與文獻整合之體驗，並探討此過程中 AI 帶來的改變與啟示。

### GPT-3.5：初識智能對話的驚訝

首次接觸 GPT-3.5 時，最明顯的感受是其自然流暢的語言生成能力。無論在日常對話、寫作輔助或程式碼撰寫上，GPT-3.5 都能快速給出條理清晰的回答，相較於人工搜尋與整理資訊，效率大幅提升。然而，GPT-3.5 亦存在明顯限制，尤其在專業領域與複雜推理上，時常出現邏輯不嚴謹或引用錯誤的情況，甚至產生無法查證的文獻 (即所謂「幻覺」)。在此階段，使用者仍需具備高度的判斷力與驗證能力。

### GPT-4.0：準確度與多樣性的飛躍

GPT-4.0 的推出標誌著語言模型進入新的階段。其在理解能力、邏輯推理與資訊整合上的表現皆有明顯提升。於程式撰寫方面，GPT-4.0 能更精確地生成可執行程式碼，減少語法錯誤。在文獻整合與資料比對時，也


能更準確地引用已知研究，幻覺情況明顯下降。以實際應用為例，GPT-4.0 在「食物圖像熱量判斷」任務中的準確度可達 75% 以上，能有效辨識單一食物種類與份量。雖在複合料理 (如炒飯、焗烤) 上仍有偏差，但整體進步相當顯著。程式撰寫的能力更是大幅度的強化，能更有效理解功能需求並生成相應程式碼，可行性明顯提高。就一個完全沒有撰寫程式碼的人來說，可以開發簡單的小工具，試錯時間也大幅度降低。

### GPT-5.0：更強大的上下文整合能力

GPT-5.0 的最大特徵在於「上下文整合」與「多來源資料融合」能力。它能延續先前對話脈絡，並在處理多層次指令時展現更強的綜合推理與個人化回覆能力。然而，實務上 GPT-5.0 對指令意圖的判斷有時不若前代精準，偶會出現答非所問或需以更長、更結構化的提示 (prompt) 才能獲得理想回答。但是，圖像辨識上面來說，已經進展到一個相當不錯的階段，相較於如同國小程度的 GPT-3.5 來說，GPT-5.0 的辨識感覺相當於大學生，至於對功能需求程式碼的撰寫，更是進步的非常的多，可以再更縮短驗證時間。

### AI 帶來的反思與展望

長期運用人工智慧於專業學習與臨床教育，讓我深刻體會到其潛力與挑戰。身為糖尿病衛教師，我特別感受



到 AI 在整合與更新醫學知識上的高效率與即時性。過去需花費大量時間交叉比對文獻、整理資料，而今 AI 可快速提供多元觀點，協助臨床人員在有限時間內掌握關鍵資訊，提升整體學習效能。特別在處理特殊個案或情境時，AI 更像是一位經驗豐富的輔助顧問，提供有價值的參考。儘管 AI 的資訊仍須專業驗證，其所帶來的學習模式轉變，確實為臨床教育開啟了新局。

AI 的使用並非為了取代專業，而是建立在既有知識與判斷力之上。這與早期文獻查詢時重視實驗設計與期刊可信度的原則相似。若缺乏基礎專業能力而完全依賴 AI 的判斷，將容易陷入「被工具引導而非運用工具」的風險。換言之，AI 是放大大類專業價值的輔助工具，而非取代思辨的捷徑。值得注意的是，在 GPT-4.0 世代之後，AI 的互動方式逐漸具備情感回饋的特質，使人更容易投入對話並感受到被理解。

### AI 與病友互動的臨床挑戰


AI 的普及也正改變病人就醫行為。越來越多病友主動向 AI 詢問健康建議，並攜帶結果至門診討論，形成「AI 天天相處」與「專業者每月十分鐘」的落差。這雖為臨床帶來挑戰，卻也是轉機。我們應將 AI 視為對話引子，肯定病友主動參與健康管理的態度，進一步引導其認識 AI 建議的侷限與風險。我在學習程式時的經驗便是明證：雖然 AI 能快速給出解答，但在缺乏底層理解時，仍會感到不安並渴望專業支持。這正是病

友對衛教師的需求所在——提供具備風險預判力的專業解析，成為資訊的最後防線，從而建立更深層的信任。

三個臨床挑戰方向：(1) 將 AI 視為「討論的引子」：與其否定病友問 AI，不如肯定他們主動管理健康的動機。我們可以說：「AI 提供的方向正確，但不夠全面，但針對您上週的血糖波動，我們需要微調一下...」。(2) 強化「專家驗證」的價值：衛教師不再是唯一的資訊源，而是「資訊的最後防線」。當病友發現 AI 的建議在您的專業分析下變得精準且個人化時，那 10 分鐘建立的信任將超越 AI 24 小時的陪伴。(3) 從「告知」轉向「教導判斷」：衛教的重點將轉向引導病友判斷 AI 回答的風險。例如，教導病友如何識別 AI 建議中可能導致低血糖的危險因子，這種專業的風險預判，是病友最需要的安全感。

### 結語：AI 時代下的臨床核心價值

AI 的出現並非削弱臨床專業價值，而是提醒我們轉化角色：從資訊提供者，走向判斷與情感支持的導引者。未來的臨床衛教將不再以「知識量」為唯一核心，而是強調「互動質感」。AI 能幫助我們節省搜尋資料的時間，卻無法取代對病友語氣的觀察、情緒的共感與臨床經驗的累積。在這個科技不斷進化的時代，唯有持續精進專業、擁抱人性關懷，才能讓 AI 成為提升醫療溫度的助力，而非冷漠的隔閡，真正實現科技與人性的融合。



## 參考文獻

1. OpenAI. GPT-4 Technical Report. arXiv preprint arXiv:2303.08774. 2023.
2. Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712. 2023.
3. Liu NF, Lin K, Hewitt J, et al. Lost in the middle: How language models use long contexts. *Trans Assoc Comput Linguist*. 2024;12:157 – 173. doi:10.1162/tacl\_a\_00638.
4. Du T, Huang H, Wang Y, Wang Y. Long-short alignment for effective long-context modeling in large language models. arXiv preprint arXiv:2506.11769. 2025.
5. Qu Y, Huang S, Li L, Nie P, Yao Y. Beyond intentions: A critical survey of misalignment in large language models. *Comput Mater Continua*. 2025;85(1):249 – 300. doi:10.32604/cmc.2025.067750.
6. Meyer A, Riese J, Streichert T. Comparison of the performance of GPT-3.5 and GPT-4 with that of medical students on a standardized medical licensing examination. *JMIR Med Educ*. 2024;10:e50965. doi:10.2196/50965.
7. Yeadon W, Peach A, Testrow CP. A comparison of human, GPT-3.5, and GPT-4 performance in a university-level coding course. *Sci Rep*. 2024;14:23285. doi:10.1038/s41598-024-73634-y.
8. OpenAI. Introducing GPT-5 [Internet]. OpenAI; 2025 Aug 7 [cited 2026 Jan 24]. Available from: <https://openai.com/index/introducing-gpt-5/>
9. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: Focus on clinicians. *J Med Internet Res*. 2020;22(6):e15154. doi:10.2196/15154.
10. Volpato R, DeBruine L, Stumpf S. Trusting emotional support from generative artificial intelligence: A conceptual review. *Computers in Human Behavior: Artificial Humans*. 2025;5:100195. doi:10.1016/j.chbah.2025.100195.