

✓ LLM 會不會因為使用者的對話而「調整模型」？

編撰：屏東大學 周國華老師 (與 ChatGPT 共筆) 2025/11/26

一、模型本身（權重）——不會在對話中自動更新

像 GPT-5、GPT-4 等大型語言模型，在對話中不會自行更新權重。
也就是說：

- 不會因為你問了什麼
- 也不會因為我回了什麼
- 更不會因為你糾正了我

→ 模型參數 (weights) 都不會在對話中被重新訓練或改變。

這是為了確保：

- 所有人用到的是同一個穩定模型版本
- 模型不會被惡意對話「污染」
- 回答品質可預期且安全

！你問 10 次「 $2+2=5$ 」，模型並不會因此以後回答 5。

二、模型會「記住對話」嗎？——短期記憶會，但不影響模型本身

在對話中，模型可以利用「上下文」來維持連貫性。
這是 Context / Prompt Input 的作用，不是修改模型。

例如：

- 你告訴我你正在做 XBRL 教材，我後續可以接著這個話題回答
- 但這些資訊不會寫回模型的權重裡

當對話關閉，這些短期記憶就不再存在。

三、長期個人化記憶 (Memory) —— 要經過你明確允許

如果你有啟用像 ChatGPT 的「個人記憶 (Memory)」：

- 我會記住你允許記住的資訊 (例如你教 AIS、在 NPTU、做 XBRL)
- 這些記憶是外掛於模型之外的資訊庫
- 完全不影響模型本體

它只是協助：

- 回答更符合你的背景
- 提供你更個人化的建議
- 減少你每次都要重新描述自己

但這不等於模型被微調。

四、什麼情況下模型會被真正調整？—— 需正式訓練或微調

真正更新模型，需進行：

- 基礎訓練 (Pre-training)
- 微調 (Fine-tuning)
- RLAIIF / RLHF
- 系統性模型更新 (例如從 GPT-4 → GPT-5)

這都需要：

- 大量資料
- 訓練架構
- 專門硬體 (像 GPU / TPU)
- 模型公司的內部操作

→ 絕對不會在一般對話中發生。

結論

大型語言模型在與使用者互動時不會修改自己的模型參數。
模型的回答會根據上下文調整，但這只是利用前文資訊，不是模型訓練。
只有透過正式的微調或再訓練，模型的權重才會真正更新。
