

運用 ChatGPT 撰寫作業的風險

一、什麼是 ChatGPT

ChatGPT (2023) 的全名是 Chat Generative Pre-trained Transformer，這是由美國的 OpenAI 實驗室在 2022 年 11 月所推出的人工智慧聊天機器人程式，它可藉由文字問答的方式，回答人類的問題或執行指定的內容產製任務，例如根據人類輸入的題目和要點，產生一篇完整的文章或劇本，甚至編寫簡單的程式，由於 ChatGPT 所產生的內容已極度近似真人撰寫的文稿，因此也引發了 ChatGPT 能否用於取代人類創作者或代筆學校作業、學術論文的疑慮。

二、ChatGPT 的原理

根據臺大電機系副教授李宏毅的觀察(李宏毅, 2022), 如圖一所示, ChatGPT 極有可能是經由文字接龍的方式學習產生句子, 再經由人類的訓練篩選出最符合人類喜好的回應文字。

圖一 ChatGPT 的學習四階段



本圖片由臺灣學術倫理教育資源中心繪製，內容來源：李宏毅（2022）

首先從任意的句首或單一的字詞開始，ChatGPT 從網路上的大量文本，學習到以「文字接龍」的方式，將句首或字詞拉長到整個句子。此過程無關該字詞或該句子的真實意涵，最終 ChatGPT 會學習到哪些句子的「接龍」方式，在什麼樣的場合下，會有較高的機率是獲得人類肯定的。在其他的相關研究中已經指出，人類並不需要澈底列舉所有可能的問題或答案，僅需要大約數萬則題目，就可以得到可接受的訓練效果 (Jocelyn, 2022a)。

在這個過程中，ChatGPT 基本上完全不理解這些問題和答案中的意涵及相關知識，它只是單純地根據訓練過程組織出較有可能被人類肯定的文字。

三、ChatGPT 的侷限

如前所述，ChatGPT 所做的事情，基本可以視為把網路上眾人給的答案抄下來，並且以人類最偏好的文字形式重新編寫成一篇文章，當中其實沒有任何知識基礎的決策判斷成分，因此 ChatGPT 無法自行檢驗資訊的真偽與正確性，並且無從判定確切的資訊來源為何。

一個問題如果極其顯而易見、或極其荒謬到沒有討論價值，就會鮮少有人公開討論，那麼 ChatGPT 也無法正確地回答，例如：李宏毅教授曾詢問 ChatGPT 「numbers」 有幾個字母，ChatGPT 無法計算就回答了 8 個（李宏毅，2022）；學者 Gary Marcus 提及，有使用者請 ChatGPT 撰寫一篇關於「吉拿棒用於人類腎臟移除手術」的文章，而 ChatGPT 卻貌似專業的寫出一篇「吉拿棒尺寸小，用於手術中能更為精準且更容易控制，也降低併發症的機率，整體讓病患的情況改善」的荒誕敘述（Jocelyn，2022b）。由此可見，ChatGPT 的侷限在於它終究是語言的人工智慧（AI）模型，本身並沒有任何知識判斷或生產的能力。

儘管 ChatGPT 學習了大量的文句且經過人工的喜好選擇，能在遣詞用句與文法方面表現優秀，但其產出內容的邏輯是否合理、論述是否符合事實、資訊來源的可信度等方面卻都無法提供保證。

四、運用 ChatGPT 撰寫作業的風險

由於 ChatGPT 是沒有知識判斷能力的語言模型，因此若運用來撰寫學校作業、研究報告、學位論文等作品，將可能產出具有邏輯漏洞的文稿，甚至有違反學術倫理的疑慮。例如可能的問題漏洞有：

1. **時空與人物錯亂**：由於 AI 語言模型並不理解它所產製的內容，它只是單純地模仿網路上收集來的文字，重新編織成一篇文章，所以文中可能會出現不屬於同一時代、或不應該出現於同一個地區的人事物，或者明顯的人名誤植。
2. **邏輯錯誤或自相矛盾**：由 AI 語言模型所產生的文章，可能會取材自論點完全相反的文字素材，並且揉合在同一篇文章，因此整篇文章可能會出現論點在文內前後自相矛盾的邏輯錯誤；此外，關於數字的計算，ChatGPT 也經常犯下顯著的錯誤，例如前面提到的「numbers 有幾個字母」問題，或是文章內出現加減乘除等計算式，ChatGPT 通常無法接上正確的計算結果。

3. **無法提供資料來源或引用出處：**如前所述的，AI 語言模型所做的是根據過去訓練而得的「經驗」在組裝文字，它無法感知哪些話是出自哪一個知識領域，也無法得知哪些文字資料是引用自哪個出處來源，因此便無法標註這些資訊的出處。

如果一篇由 ChatGPT 所撰寫出的文稿，不僅無法提供資訊來源，還可能出現名詞定義胡亂填空、數據計算錯誤、結論前後矛盾甚至出現內容虛構等情形，那麼這樣的寫作品質就很值得憂慮了。如果也涉及違反學術倫理的抄襲、造假、變造等不當行為，將影響研究者的學術誠信及聲譽。

五、結語

學術研究非常看重原創性，並且重視自我思考與創造的過程，然而 ChatGPT 的產製內容是由既有的資料庫所重新組織而成的結果，缺乏知識判斷且毫無經過理解，這些都使得 ChatGPT 回答的文稿不具原創性，甚至在資料來源的標示上有所侷限，無法列出參考來源，更遑論正確性不足的問題，這些都使得 ChatGPT 回答的文稿若作為研究報告、碩博士論文、學校作業等，都將擁有許多風險且容易衍生爭議。

ChatGPT 就像其他的生成式 AI (generative AI) 一樣，擁有廣泛的應用範圍，能代替人類做許多瑣事，然而，就像 Deepfake 技術被用於產生假影片、假新聞等，ChatGPT 也可能被有心人士濫用於造假、剽竊、或是產製不當內容，惡意應用來產生大量的假消息。

身處這個 AI 時代的我們，除了期待相關機關立法限制這些生成 AI 技術的惡意濫用以外，本身也應該具備查證與拒絕不當內容的能力。除此之外，所有的數位內容無論是否由 AI 所產生，都應該標註出處和作者，就像「生產履歷」一般，數位內容應當都要能夠追溯到作者；也就是說，不能追溯到作者的作品，就不應該被使用，否則就可能誤用或誤信了由 AI 惡意生成的內容。這是在這個處處是 AI 產物的時代，人人應有的基本素養。

參考資料

- ChatGPT(2023,2月20日)。載於維基百科。<https://zh.wikipedia.org/zh-tw/ChatGPT>
- 李宏毅(2022,12月7日)。**ChatGPT(可能)是怎麼煉成的—GPT社會化的過程**〔影片〕。YouTube。<https://youtu.be/e0aKI2GGZNg>
- Jocelyn(2022a,12月8日)。**ChatGPT究竟如何煉成?台大教授李宏毅提可能的訓練步驟**。INSIDE報導。<https://www.inside.com.tw/article/30032-chatgpt-possible-4-steps-training>
- Jocelyn(2022b,12月4日)。**ChatGPT為何有能力同時給錯誤和正確答案?學者 Gary Marcus 提自身觀點解惑**。INSIDE報導。<https://www.inside.com.tw/article/29960-gary-marcus-openai-chatGPT>

作者：國立陽明交通大學 教育研究所 楊舒凱專案管理師
文字編輯：國立陽明交通大學 人文與社會科學研究中心 詹韻蓉計畫助理
專家審查：國立陽明交通大學 教育研究所 周倩教授
國立陽明交通大學 人文與社會科學研究中心 潘璿安助理研究員

注意事項：本文內容僅作者個人觀點，不代表教育部立場，相關學術倫理規範請以主管機關之公告資訊為主。

※更新日期：2023年3月17日

