

## 專訪國立臺灣大學資訊工程學系教授陳信希

# 小型大語言模型興起 生成式 AI 應用範圍廣

採訪／施鑫澤 文／林裕洋

市面出現有愈來愈多參數量僅有 1B、2.7B、7B、13B 等小型大語言模型，搭配特定資料進行訓練之後，同樣能在特定領域有亮眼表現。

被譽為 AI 2.0 的生成式 AI，在各領域展現出強大能力之後，在 Gartner 公布 2024 年十大戰略性技術趨勢中，已將此列為不容忽視的重要科技之一。2023 年在生產環境中部署支援 GenAI 應用程式比例還不到 5%。Gartner 預測到 2026 年，超過 80% 企業將在生產環境中部署支援 GenAI 的應用程式。而在自然語言處理相關技術快速進化，對個人化和改善用戶使用體驗的需求不斷成長，以及硬體和機器學習技術的不斷進步，Lucintel 預測 2024 年至 2030 年之間，全球大語言模型市場複合年成長率將達到 10.1%。

儘管生成式 AI 結合大語言模型後的威力非常強大，但是由於背後需要仰賴高 AI 算力協助，且此類 AI 模型難以落地，很難用於提升員工效率之用。在 AI 普及化的浪潮下，除愈來愈多業者投入超大型的大語言模型之外，市面也出現參數量僅有 1B、2.7B、7B、13B 等小型大語言模型，主打能在 AI PC 設備中運作，搭配特定資料進行訓練之後，能稱職扮演協助員工提升工作效率的後盾。

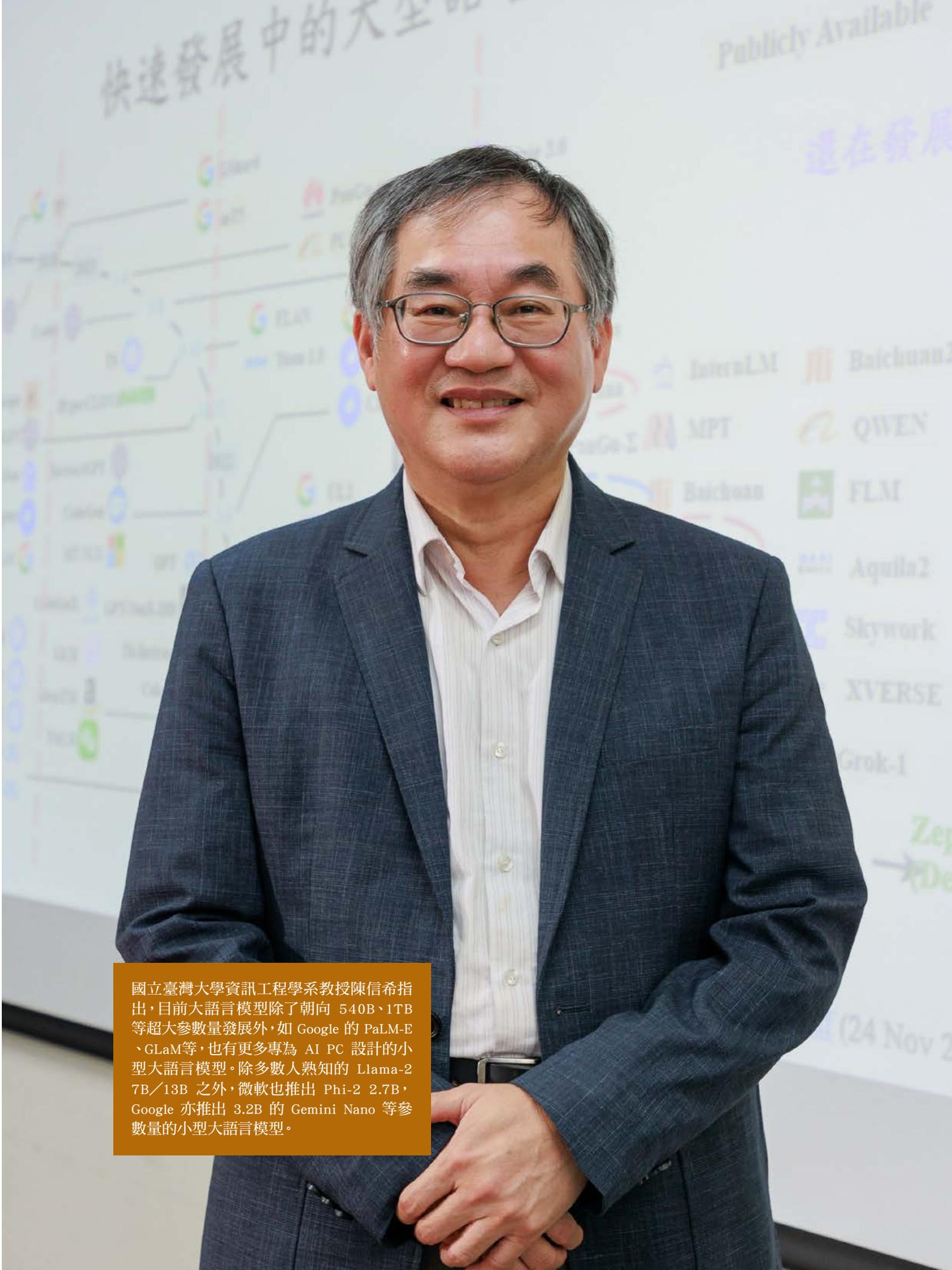
國立臺灣大學資訊工程學系教授陳信希指出，事實上，全球科學家投入大語言模型相關研究已經多年，近幾年在 AI 算力大幅飆升下，目前已在不同領域展現出優異成績，也加速各業投入大語言模

型的開發工作。當前大語言模型除了朝向 540B、1TB 等超大參數量發展外，如 Google 的 PaLM-E、GLaM 等，也有更多專為 AI PC 設計的小型大語言模型。除多數人熟知的 Llama-2 7B/13B 之外，微軟也推出 Phi-2 2.7B。Google 亦推出 3.2B 的 Gemini Nano 2 等參數量的小型大語言模型。

雖然小型大語言模型的參數量較少，不過在 BBH 資料集常識推理、語言理解、數學問題及程式碼撰寫等評比項目中，整體表現與 70B、175B 等大語言模型相去不遠，只需要搭配合適的資料進行微調，即可在不同領域之中扮演員工的最佳助手。

### ChatGPT 席捲全球 臺灣大學採取開放態度

隨著產業加快推動生成式 AI 專案，目前已在醫療、金融、製造、教育、媒體等領域展現出效益。以軟體開發應用場景為例，GitHub 攜手 Open AI 推出的 GitHub Copilot 是一種經由人工智慧訓練，專為輔助程式設計師開發的生產力工具，具備可自動完成程式碼、適應多種語言、支持程式碼重構、提供即時編輯建議，且能用 Visual Studio、VSCode 等環境之中。該工具會透過分析程式碼庫和使用者輸入的內容，自動生成自然語言、程式碼



國立臺灣大學資訊工程學系教授陳信希指出，目前大語言模型除了朝向 540B、1TB 等超大參數量發展外，如 Google 的 PaLM-E、GLaM 等，也有更多專為 AI PC 設計的小型大語言模型。除多數人熟知的 Llama-2 7B/13B 之外，微軟也推出 Phi-2 2.7B，Google 亦推出 3.2B 的 Gemini Nano 等參數量的小型大語言模型。

建議，幫助開發者提高生產力。

對程式設計師而言，GitHub Copilot 可帶來三大好處，首先是**減輕開發工作量**，減少開發人員的打字時間，從而釋放更多時間專注於設計和創新。其次能**生成更乾淨、更可讀的程式碼**，有助於提高軟體專案的產品品質和可靠性。第三點則是通過自然語言的交互詰問，創造新編程體驗、使編程更加直觀和有趣，有助於**新手工程師加速熟悉開發流程**。

陳信希表示，2023 年初臺灣大學就關注 ChatGPT 問世之後，對教育帶來的影響。過去一年多來，臺大建立專頁說明生成式 AI 工具之教學因應措施，進行生成式 AI 工具使用經驗調查，舉辦善用 ChatGPT 的教與學工作坊，設計不同的應用情境，協助老師與學生善用此工具提升教學與學習效益。



陳信希指出，在生成式 AI 浪潮下，對教育與就業部分自然也帶來衝擊，首先在市場對 AI 人才需求量極大的狀況下，自然也會帶動學校增加 AI 課程，讓學生能學習新技能以應對變化。然而值得關注之處，部分重複性和簡單性的工作可能被 AI 自動化取代，當然屆時也會帶來新的職業和工作。

## 加速數位內容生成 改變媒體發展生態

現今生成式 AI 已被廣泛應用於專業領域，如軟體開發、數位內容、智慧醫療等，其中在數位媒體內容生成部分，可分成**搜集資料、整理資料、核實事實、版面校對等四部分**，如可藉由 AI 技術協助整理逐字稿、將內容轉譯成多種語言、針對特定受眾撰寫文稿。

現今在 AI 技術快速發展，媒體產業正持續加快報導產出速度，節省人力成本，顛覆內容產業。如紐約時報、BBC、路透社等國際媒體，正使用 AI 自動生成財經股票、氣象、體育賽事等報導。而 OpenAI 推出的 Sora 影片生成服務，用戶僅需要輸入簡短文字描述，即可生成出逼真影片，大幅降低數位內容生成的時間，不過背後也有很大的使用疑慮。例如，這項技術雖然能加快創意內容的生成，但也衍生易於製作和散布假影片的問題。

「除文字方面的新聞內容生成之外，近來我們也看到媒體產業運用 AI 技術發展虛擬電視新聞主播，如臺灣民視新聞的 AI 主播-敏熙播報國際新聞、氣象新聞等。」陳信希解釋：「洛杉磯 Channel 1 也在 2024 年推出由人工智慧生成的虛擬主播，取代真人報導新聞。」

## 診別和醫師推薦服務 台大醫院優化掛號流程

人工智慧在醫療保健上有多面向的應用，例如在生物學資訊擷取、醫療諮詢提供、報告摘要和撰寫等方面，帶來極大幫助。不過，背後也可能製造醫療錯誤訊息，如誤解醫學術語、做出與醫療指南不一致的建議，可能上傳患者健康訊息的隱私問題。

在考量降低醫護人員工作負擔，為民眾提供更好醫療服務的前提下，目前在醫療環境中常見應用情境有分類任務和生成任務等兩大類。在**分類任務**部分，如可用於檢傷分級，給予急診病人資訊，判斷分級。在三日再回診部分，則預測住院病人出院後三日內再急診的可能性。至於疾病編碼部分，從病歷中判斷疾病分類（ICD）編碼。在**生成任務部**

分，如透過機器翻譯功能協助，將一段英文病例翻譯成中文病歷。而由圖像生成文字部分，則可從醫學影像生成報告，又或者給予問題可生成答案，如衛教問答。至於與病患學習相關的部分，則只需輸入症狀，推薦就診診別和醫師。

陳信希指出，為提供病患更好的服務體驗、減少民眾不知道掛哪科的焦慮和困惑，同時降低第一線人員的工作負擔，2023年3月14日臺大醫院正式推出**診別和醫師推薦服務**。就診病患只需於臺大醫院APP輸入自身的症狀問題描述，系統將問題串接健保分科做為選項，讓ChatGPT API回應適當就診健保科別及代碼。如此一來，民眾先透過ChatGPT回答選擇合適科別，而不一定要撥打總機電話詢問，提升掛號科別正確性並且減少時間與資源浪費，有助於加速就醫流程。

值得一提，儘管臺大醫院推出診別和醫師推薦服務之後備受好評，不過兼顧病患的個人隱私，以及鑑別診斷的專業性／正確性、資源分配之公平性和效益等，臺大醫院也會持續進行系統優化工作。

## 疾病分類輔助系統上線 大幅降低人力負擔

在AI科技持續進步，即便是參數量僅有7B、13B等基礎模型，只需要經過適當微調也能展現出相當強大的威力，在醫療產業展現出絕佳效果。如臺大醫院便選擇**預訓練的 Vicuna v1.5 為基礎模型**，運用臺大醫院的任務和資料集等進行多任務學習，如ICD-10疾分編碼預測、病理報告SNOMED編碼、自動化報告彙整系統等，提供疾分人員、醫檢師、醫師、護理師等使用。

根據臺大醫院統計資料顯示，以大語言模型為核心的疾病分類輔助系統上線之後，短短一秒即可產生住院病人的**ICD-10-CM／PCS疾病分類編碼**，包含診斷碼及處置碼等資訊，主診斷達到**80.77%**、整體一致率**86.28%**，準確率也較舊版的深度學習模型改善不少。

陳信希指出，考量到病歷資料屬於敏感性的個資，在保護病患隱私的前提下，我們採用**建置於地端的大語言模型**，可兼顧科技創新並符合法規要

求。臺大醫院不光擁有先進深度學習及生成式AI技術、大量優質訓練資料，結合優秀疾病分類團隊建立系統化專家知識回饋流程，可望擴大在醫院內部的應用範疇。

## AI 兩面刃 政府與產業應做好準備

生成式AI浪潮席捲全球，影響力幾乎涵蓋所有產業，根據Cognizant Impact公布研究指出，未來十年對美國經濟影響達到驚人的一兆美元，且恐有高達九成工作受到影響。若進一步分析對不同領域帶來的衝擊，在經濟、產業領域，確定某些行業可能受到衝擊，但也有不少會因引進以AI為核心自動化智慧技術，達到提高效率、降低成本的目的，也讓整體產業呈現蓬勃發展，為公司發展帶來新機會。

在生活影響層面部分，消費者可藉由引進智慧家居方案，享受到安全提升、生活更便利等優點。在醫療部分也能藉由AI技術提升醫療水準，如診斷、治療、以及藥物研發。在前述好的改變之外，自然也會有眾多挑戰需要克服，如在創造美好生活之餘，也得兼顧倫理和隱私等環節。另外，部分地區和社會群體無法充分享受科技帶來的好處，政府單位必須思考解決數位落差的問題。而在AI大行其道下，背後仰賴資料中心提供的強大算力，也會讓碳排放問題變得日益嚴重。

陳信希指出，在生成式AI浪潮下，對教育與就業部分自然也帶來衝擊，首先是**個人化、靈活、遠距學習方式等將會改變傳統教育模式**，且在市場對AI人才需求量極大的狀況下，自然也會帶動學校增加AI課程，讓學生能學習新技能以應對變化。然而值得關注之處，部分重複性和簡單性的工作可能被AI自動化取代，當然屆時也會帶來新的職業和工作。

在AI時代來臨之際，陳信希建議政府、教育單位、企業等，必須在政策、教育、就業方面等做好準備，才能以應對不斷變化的世代，進而在全球市場中保有強大競爭力。