

中大成宇：開發深度思考AI大模型 港產「T1」反應快成本低

AI港人才②

伴隨人工智能（AI）技術的加速演進，AI大模型已成為產業的核心驅動力，如何優化現有的大模型，構建更高效的模型以適配社會需求，是當代AI發展的重要課題。

香港中文大學計算機科學與工程學系副教授成宇，在接受《大公報》專訪時表示目前專注於高效AI模型架構研究，當中涵蓋多模態模型以及快速模型架構兩大方向，並取得多項成果。其團隊擔任技術顧問的騰訊深度思考模型「T1」，已在今年3月底推出。此模型開發成本與DeepSeek相比更低，響應時間更是壓縮至10多秒。成宇希望，在一至兩年後AI可在特定的工種替代人力，5至10年內實現應用場景內80%的工作由AI完成。

大公報記者 陳煒琛

成宇與其團隊目前主要聚焦於多模態模型架構以及快速模型架構的研究。所謂多模態模型，是能夠綜合處理圖片、語音與文本等多模態信息的模型。而快速模型架構，則是對問題推理、響應時間更加快速的模型。

從單模態進化到多模態

成宇指出，「當前主流的深思考模型，如DeepSeek、ChatGPT等，主要是採用單模態架構，在文本處理領域具有強大的邏輯推理能力，但其應用場景相對局限。」相比之下，多模態模型能綜合處理多種複雜信息，能有效應對現時社會的多樣需求。此外，他表示，由於當前的模型規模較大，「在實時交互場景中面臨推理時間較長、算力資源成本高昂的痛點」，因此研發兼顧精度與效率的快速模型架構是發展AI的重要方向。

成宇去年10月份加入中大任職，而在加入中大之前，成宇在微軟美國研究院與OpenAI團隊合作期間，對其Copilot模型的結構進行輕量化優化，為其後續的快速推理技術研究積累了關鍵經驗。

在不斷推進當前研究工作的同時，回顧過往，早在2020年，成宇及其團隊便有極具前瞻性的構想。成宇表示，在多數大模型僅關注文本信息時，他與團隊就提出構建多模態大模型的想法。他透露，在當時他們便搭建了將文本、圖像、語音等信息融合起來的第一代的多模態模型框架。成宇直言，「這一成果為後續GPT-4o等多模態模型架構提供了技術雛形。」

此外，AI在迅猛發展的同時，不可避免地會出現諸如AI幻覺（即虛構現實）、AI偏見（如性別歧視）、AI安全（如生成危險有害內容）等方面問題。針對以上問題，成宇表示，在2023年他與團隊構建了一套客觀的評測機制，借助數據去檢測大模型存在的問題，進而及時進行修改。成宇透露，目前該評測機制已被不少企業採納。

5至10年內可做八成工種

騰訊研發的「T1」模型是標誌性落地科研成果。成宇表示，「T1」屬於較快速的深思考模型，在今年3月底已在微信「騰訊元寶」免費上線。成宇透露，大概是去年年中開始與騰訊合作，年底時便搭建了一個反應較快的基座模型。而後在這個基座模型的基礎上，利用大量的數據不斷地進行迭代，最終花費了七至八個月的時間完成了「T1」的搭建。

成宇表示，一般情況下，深度思考或深度推理類模型在維持其深度思考能力（如邏輯回溯、多步驗證）時，其響應速度會變慢，則需要優化其模型架構。他透露，在「T1」模型研發的第一階段團隊已解決此問題，成功壓縮了推理時間。

成宇補充，「與DeepSeek相比，T1響應速度可能

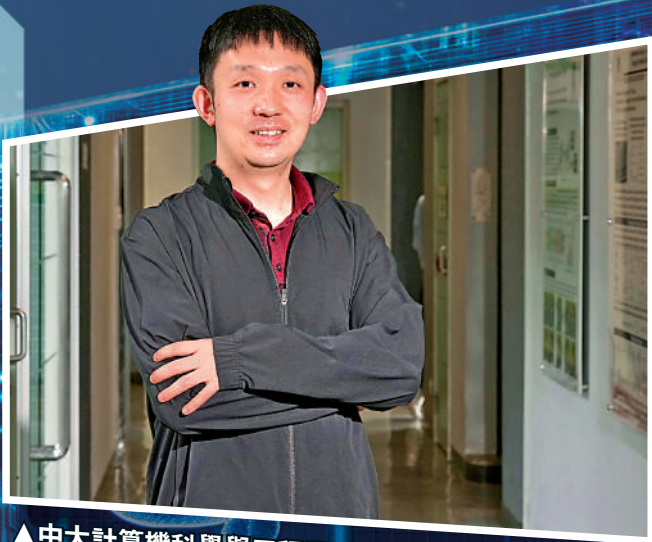
僅為其三分一到二分一，即10多秒便可達到同樣的處理效果」。

據悉，「T1」沿用了混元Turbo S的創新架構，並採用了Hybrid-Mamba-Transformer融合模式。而這也是工業界首次將混合的Mamba架構無損應用於超大型推理模型。這一架構有效降低了傳統Transformer結構的計算複雜度，減少了內存佔用，成宇表示，這顯著降低了訓練和推理成本。

儘管「T1」取得了不錯的成果，但在研發和應用推進過程中，亦面臨着不少挑戰。成宇坦言，AI領域的競爭激烈，新模型層出不窮，因此他們時刻面臨着外部的緊迫壓力。此外，AI模型漫長的迭代周期長同樣是一大難題。成宇進一步解釋，「當前的AI模型體量龐大，訓練一個模型往往需要耗時一至兩個月」。而在此期間，團隊須保持高度專注，持續去監控模型訓練過程中的各項表現指標，並根據實際情況動態調整架構，以確保最終成果的質量與性能。

目前，團隊也正探索如何將「T1」模型進行多模態升級，以增強其在不同場景下的應用能力。成宇舉例，讓「T1」做到在面對一道複雜的幾何題時，可以自主解析圖像、繪製輔助線和公式推導。成宇對AI未來的多模態應用充滿信心，他希望，一至兩年後可率先在特定的工種如代碼開發、平面式設計等領域實現AI替代人力的突破，「終極目標則為在5至10年內實現應用場景內80%工作由AI完成。」

►「T1」模型在今年三月底已在微信「騰訊元寶」免費上線。



▲中大計算機科學與工程學系副教授成宇與其團隊，目前主要聚焦於多模態模型架構以及快速模型架構的研究。
大公報記者麥潤田攝



「一定要以解決實際問題為出發點」

融入
各業

科研的意義在於解決實際問題、推動技術進步以及培養下一代科研人才。成宇強調，「科研不應局限於象牙塔，而應關注實際應用」，通過與企業與使用者的交流，學者可以更加了解AI技術的發展痛點進而提供更有有效的解決方案。成宇直言，可能當下的絕大部分研究並不能發揮大作用，但隨着時間的演變，無數科研的積累可以為未來的技術爆發奠定基礎，他形容「這是在正確的道路上不斷向前走」。

批判思維創新能力無法取代

而對於未來想要深耕於AI領域的學生，成宇提醒，一定要以解決實際問題為出發點。其次，成宇表示，AI不光是一個算法，若想AI模型具有效果，能解決問題，就會牽涉到非常大的數據和工作量，因此一定要加強動手能力和豐富實踐經驗。

除了研究者，成宇也關注全面AI素養的提升。他表示，AI將逐漸成為基礎工具融入各行業，公眾需掌握相關知識以便更高效地利用AI。在AI教育方面，他認為，對於一些知識性的東西，即純靠記憶、直覺去解決的問題，無需要求學生死記硬背，可以利用AI輔助工具去幫助學習。成宇表示，培養學生的批判思維和創新能力才是重點，因為這些是AI暫時無法取代的。



▲中大開設的人工智能相關的研究所，包含逾200名研究人員，畢業生又多一條出路。

港發展AI具優勢 惟產業轉型需時

迎接
挑戰

成宇2010年在清華大學本科畢業後，便前往美國西北大學求學，並在2015年取得博士學位。在求學以及工作的過程中，成宇表示他見證了兩次AI浪潮，首次是深度學習模型的崛起，其次是GPT等大模型的興起，他看到自然語言處理極速的快速演進，深感AI技術的強大，更加堅定了科研道路。

盼加大力度支持高校發展

談及來港發展的契機，成宇從個人與職業雙重維度作出解析。個人層面，他坦言想縮短與家人的距離。而從職業發展層面，則是因為香港提供的多元發展可能。他說，「無論是深耕學術研究、創業還是就業都能獲得較多機會。」同時他也表示，香港政府正極力擁抱AI，不斷優化創科政策、引進全球頂尖企業和科研人才。因此香港的創科環境是他回來的另一重要原因。此外，香港的高等教育資源亦是關鍵，尤其是在計算機和AI領域，中大等院校躋身前沿陣營。成宇透露，「以前在內地讀本科時便聽說過中大在AI領域的領先地位，所以

選擇來港。」

儘管香港優勢明顯，但成宇指出其AI發展面臨若干挑戰。其中核心難點在於產業定位的轉型。作為國際金融中心，香港若想在短期內轉型為AI中心或科技樞紐，成宇認為，「中間需要較長的過渡周期。」

其次，他表示，當前在港扎根落地的世界級企業仍需增多，這需要政策調整與配套設施完善形成協同支撐，而這同樣需要時間沉澱。成宇表示，「過渡過程有阻礙很正常，只要方向正確且發展態勢向好，穩步前行就可以突破階段性瓶頸。」

而作為一個科學學者兼教育工作者，成宇希望政府可以再加大對高校發展的支持力度。他指出，「當前高校定位可能比較適配傳統的、偏小的科研作坊氛圍。」他進一步解釋道，高校在算力基礎設施方面仍比較缺乏，很難去配合AI時代的「大基建」發展需求，因此希望政府在算力資源和各種保障上可以提供更多的投入，同時亦可加大对高校科研成果的宣傳，促進產學研合作。

中大AI相關研究所及AI香港課程

中大開設人工智能相關的研究所：包含逾200名研究人員。

- 中大天石機器人研究所
- 智能設計與製造研究所

- 醫療智能與擴展現實研究所
- 中大人工智能交叉學科研究所

中大開設人工智能相關課程：（11個）

- 人工智能：系統與科技學士學位課程
- 計算數據科學學士學位課程
- 計算機科學學士學位課程
- 計算機科學與工程學士學位課程
- 機械與自動化工程學士學位課程

- 計算機科學與工程學哲學碩士及哲學博士課程
- 計算機科學理學碩士課程
- 人工智能理學碩士課程
- 機械與自動化工程學哲學碩士及哲學博士課程
- 機械與自動化工程學理學碩士課程
- 機器人學理學碩士課程



▲人力勞動會否被機器人代替仍是未知數，但部分工種已是大趨勢。