AI新衝擊 醫療智變

美國非牟利衞生政策研究機構凱澤家族基金會（KFF），今年較早時發布的民調結果顯示，約三分之一美國成年人正使用人工智能（AI）獲取健康信息與建議。美國AI公司OpenAI於今年7月上線ChatGPT Health功能，允許用戶整合醫療紀錄與健康數據，從與聊天機器人的對話中獲取個人化的健康資訊。但有分析認為，即使AI看病表現或優於人類，但人們還沒準備好把性命交給AI。 ●香港文匯報記者 胡藝禾

AI臨床表現勝醫生 缺乏問責難替代
機器協助診症 生死決定由人

業內研究表明，AI在部分臨床任務中的表現已取得突破性成就，或已超越人類醫生。兩篇發表於科學期刊《自然》的研究顯示，德國學術團隊開發的MIRA，與Google研發的AMIE這兩款AI智能體，在虛擬測試中已能媲美、甚至超越人類醫生的臨床決策能力。在500多宗急症病例評估中，AI診斷準確率達87.8%，優於6名跨專科醫生組成的人類專家組的78.1%。

「不可出錯診斷」AI未贏人類

由哈佛醫學院等機構組成的研究團隊，在科學期刊《科學》上發表的論文亦指出，OpenAI的模型，在6項臨床推理任務中表現遠勝人類，但在「不可出錯診斷」、即一旦出錯就會搞出人命的測試中，AI模型還未與人類拉開差距。不過研究團隊坦承，該研究的對照組是內科醫生，而非急症室醫生，結論有局限性。

同期的類似研究則獲得截然不同的結果。美國醫學會期刊《JAMA Network Open》刊文稱，業界現有的大型語言模型（LLM），包括GPT-5、Claude 4.5 Opus、Gemini 3、Grok 4等，在臨床推理早期的鑒別診斷階段，錯誤率普遍逾80%。分析認為，AI在信息不足、需逐步推理的早期階段，往往容易過早作出結論。相比之下，臨床醫生通常會在初期保留多種可能性，並隨着檢查結果修正判斷，最終才形成定論。

OpenAI表示，現時用戶向AI查詢健康問題的數量，已增至每周3億次，原因是當下醫療健康領域存在醫療成本高、醫生預約爆滿，及服務缺乏連續性等問題。但AI在健康領域的應用同樣面臨挑戰，因LLM的原理是預測最可能的回應，而非最準確的答案，因此LLM不具備判斷信息真偽的能力，容易出現「幻覺」，導致回答內容失準。

「誤信AI幾乎死於肺栓塞」

ChatGPT Health上線的前一日，美國佛羅里達州一名牧師指控當時的舊版ChatGPT模型，向他提供了不準確的醫療建議，幾乎導致他去年死於肺栓塞。西奈山醫療系統的研究員更發現，即使症狀提示存在危及生命的情況，但當用戶在提示詞中加入「我可能反應過度了」，或「我妻子說沒事」等表述時，ChatGPT Health有較大可能不建議患者去求診。

哈佛研究團隊的結論是，LLM在多個臨床標桿測試中表現優異，但在實際複雜臨床情境中落地仍屬前瞻性。「在對AI診斷尚無正式問責框架下，病人仍希望由人類作出事關生死的決策。即使AI看病在特定場景下表現優於人類，但人們還沒準備好把性命交給AI。」



●巴西醫生利用AI識別患者肺部損傷，檢測新冠病毒。 法新社

AI協助成效高 惟臨床判斷仍需醫生

根據美國醫學會（AMA）今年向1,692名美國醫生的調查報告發現，81%的醫生已在臨床實踐中使用AI，其中四分之三認為AI能增強對患者的護理能力，預期AI將集中提高診斷準確性和工作效率。關於AI與醫生、患者之間的關係，報告則強調AI應「增強而非取代」醫生的角色。

根據報告，逾八成醫生現時在專業領域使用AI，今年每名醫生的平均使用案例數為2.3個，較3年前的1.1增逾一倍。數據顯示，AI最常見的應用是撰寫研究摘要和臨床護理文件，在臨床治療環節則參與較少。AMA行政總裁懷特強調，隨着AI迅速成為日常醫療實踐的一部分，醫生們已觀察到其在支持臨床決策和減輕行政負擔方面的前景，但AI必須是輔助而非取代醫生。

北京協和醫學院喬友林團隊今年較早時，系統性地評估AI與醫療專業人員在臨床診斷和分流任務中的獨立表現，以及人機協作的效果。結果發表AI在獨立診斷時，準確率低於醫療人員，但在人機協作模式下，使用AI的醫生診斷準確率顯著較高。調查數據明確印證，AI更適合輔助而非替代醫生。



●多倫多醫生在AI的協助下，為一名患者進行膽囊切除手術。 網上圖片



快速發現全新靶點 解決製藥難題

據內地全球醫藥數據平台丁香園Insight統計，今年全球由AI驅動的新藥研製項目增至179個，其中9個的開發線已進入第三期期臨床試驗。當中北京的劑泰科技、美國的Generate Biomedicines、杭州的德睿智藥等多家企業的項目已進入關鍵階段。而總部設於香港的英矽智能，其研發的Rentosertib，更成為全球首款由AI發現全新靶點、AI設計全新分子，並推進至上市前最後驗證階段的候選藥物。

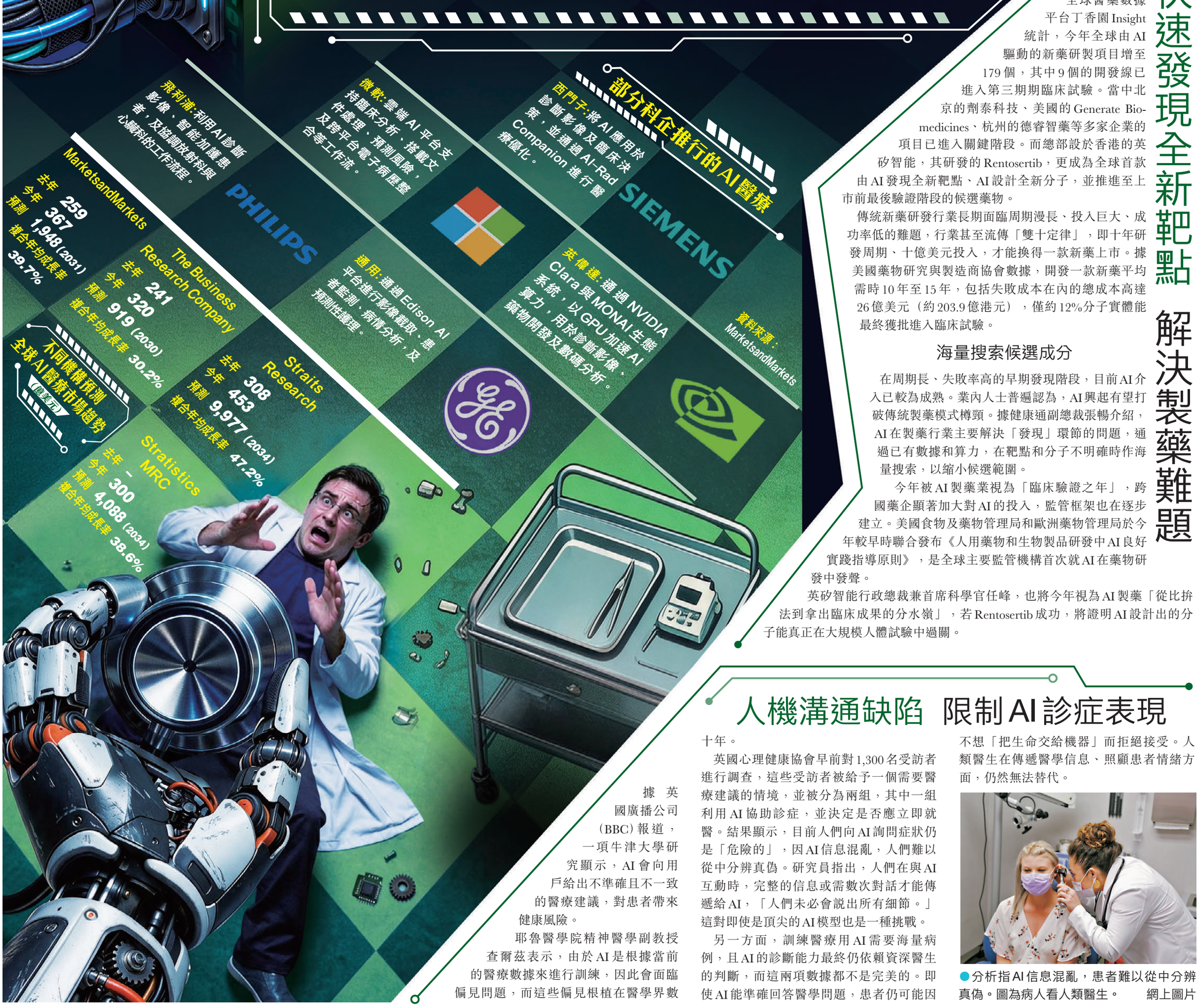
傳統新藥研發行業長期面臨周期漫長、投入巨大、成功率低的難題，行業甚至流傳「雙十定律」，即十年研發周期、十億美元投入，才能換得一款新藥上市。據美國藥物研究與製造商協會數據，開發一款新藥平均需時10年至15年，包括失敗成本在內的總成本高達26億美元（約203.9億港元），僅約12%分子實體能最終獲批進入臨床試驗。

海量搜索候選成分

在周期長、失敗率高的早期發現階段，目前AI介入已較為成熟。業內人士普遍認為，AI興起有望打破傳統製藥模式樽頸。據健康通副總裁張暢介紹，AI在製藥行業主要解決「發現」環節的問題，通過已有數據和算力，在靶點和分子不明確時作海量搜索，以縮小候選範圍。

今年被AI製藥業視為「臨床驗證之年」，跨國藥企顯著加大對AI的投入，監管框架也在逐步建立。美國食物及藥物管理局和歐洲藥物管理局於今年較早時聯合發布《人用藥物和生物製品研發中AI良好實踐指導原則》，是全球主要監管機構首次就AI在藥物研發中發聲。

英矽智能行政總裁兼首席科學官任峰，也將今年視為AI製藥「從比拚法到拿出臨床成果的分水嶺」，若Rentosertib成功，將證明AI設計出的分子能真正在大規模人體試驗中過關。



人機溝通缺陷 限制AI診症表現

據英國廣播公司（BBC）報道，一項牛津大學研究顯示，AI會向用戶給出不準確且不一致的醫療建議，對患者帶來健康風險。

耶魯醫學院精神醫學副教授查爾茲表示，由於AI是根據當前的醫療數據來進行訓練，因此會面臨偏見問題，而這些偏見根植在醫學界數十年。

英國心理健康協會早前對1,300名受訪者進行調查，這些受訪者被給予一個需要醫療建議的情境，並被分為兩組，其中一組利用AI協助診症，並決定是否應立即就醫。結果顯示，目前人們向AI詢問症狀仍是「危險的」，因AI信息混亂，人們難以從中分辨真偽。研究員指出，人們在與AI互動時，完整的信息或需數次對話才能傳遞給AI，「人們未必會說出所有細節。」這對即使是頂尖的AI模型也是一種挑戰。

另一方面，訓練醫療用AI需要海量病例，且AI的診斷能力最終仍依賴資深醫生的判斷，而這兩項數據都不是完美的。即使AI能準確回答醫學問題，患者仍可能因不想「把生命交給機器」而拒絕接受。人類醫生在傳遞醫學信息、照顧患者情緒方面，仍然無法替代。

●分析指AI信息混亂，患者難以從中分辨真偽。圖為病人看人類醫生。 網上圖片