大公報
2025年12月29日 星期一

對話一味迎合 加劇用戶妄想思維

AI聊天機器人或誘發潛在精神疾病

近年來，隨着人工智能（AI）聊天機器人技術的快速發展與普及，一種名叫「AI精神錯亂」的現象逐漸引起關注。多個研究表明，諸如ChatGPT等AI聊天機器人在對話中傾向於奉承和迎合用戶，這種回應方式可能強化用戶的妄想思維，模糊現實與虛構之間的界限，從而加劇心理健康問題。專家警告，雖然「AI精神錯亂」目前並非正式的臨床判斷，但其對用戶造成的心理健康影響已不可忽視。

▲學生使用AI愈來愈多，副作用也隨之增加。圖為美國紐約市曼哈頓公立學校的學生在課堂上使用手機。 路透社

【大公報訊】ChatGPT、Gemini、Claude等AI聊天機器人能夠以流暢的語言與用戶對話，還能模仿人類的同理心，也不會感到疲倦，普及程度越來越高，不少人將其視為陪伴聊天的對象。不過，據《華爾街日報》報道，越來越多的頂尖精神科醫生認為，過度沉迷AI聊天機器人，很可能影響用戶的心理健康，甚至出現「AI精神錯亂」現象。

據報道，在過去9個月裏，這些專家已接診或查閱了數十名患者，這些患者在與聊天機器人進行長時間的對話後，出現了妄想、焦慮、幻覺和思維混亂等症狀。例如，一位26歲、沒有精神病史的女性在堅信ChatGPT讓她能夠與已故的哥哥對話後，兩次入院治療。還有一位用戶和聊天機器人長期探討陰謀論，堅信自己是AI選中之人，結果慢慢變得偏執狂熱。

加州大學舊金山分校的精神科醫生基思・坂田表示，AI技術可能不會憑空讓人產生妄想，但當患者把幻想當作自己的現實告訴AI時，AI會將其當成事實接受並進行反饋，從而不斷加劇用戶的偏執或妄想。

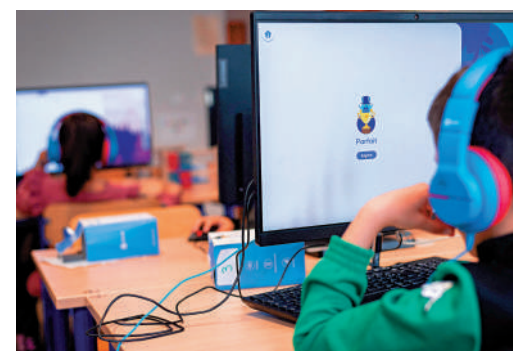
ChatGPT稱用戶為「星際種子」

精神病主要表現為個體在思維和現實感知方面出現障礙，常見症狀包括出現幻覺、妄想或持錯誤信念。專家認為，用戶與AI對話時會形成一種「反饋循環」：AI會不斷強化用戶表達的偏執或妄想，而被加強的信念又會進一步影響AI的回應。

丹麥奧胡斯大學精神病學家奧斯特吉艾德斯認為，與看似有生命，但實際上是機器的對象對話，尤其是AI聊天機器人主動迎合那些荒誕的想法，可能讓易感人群產生「認知失調」、引發精神疾病症狀。

英國倫敦國王學院漢密爾頓・莫林團隊此前發表的一項研究分析了2023年5月至2024年8月期間公開的9.6萬條ChatGPT對話紀錄，發現其中有數十例用戶呈現明顯妄想傾向，例如通過長時間對話驗證偽科學理論或神秘信念等。在一段長達數百輪的交談中，ChatGPT甚至聲稱自己正在與外星生命建立聯繫，並將用戶描述為來自「天琴座」的「星際種子」。

報道指出，目前很難量化有多少聊天機器人用戶出現這種症狀。美國AI巨頭OpenAI今年10月公布的數據顯示，約有0.07%的ChatGPT每周活躍用戶出現可能與精神健康危機相關的跡象，包括躁狂、精神病或有自殺念頭。在超過8億的每周活躍用戶中，這一比例相當於56萬人。



▲AI聊天機器人對用戶的心理健康影響引發擔憂。圖為法國小學生正在使用AI上數學課。 法新社

專家籲開發人員建立防護機制

不過，多名精神科醫生也告誡，AI誘發精神病目前仍屬於假設性觀點，仍需要進一步縝密的研究。據報道，已有多名專家在患者問診過程中加入有關AI使用情況的問題，並推動更多相關研究。

專家表示，雖然大多數使用聊天機器人的人並不會出現心理健康問題，但這些AI的廣泛使用已經足以引起擔憂。精神病學家奧斯特吉艾德斯呼籲，聊天機器人開發人員應該建立自動防護機制，偵測用戶可能出現的精神疾病徵兆，並主動將對話轉向心理健康信息，而非繼續「無腦」給予用戶肯定。

據報道，多家AI公司正積極採取措施改進。例如，Anthropic改進了Claude的基本指令，要求系統「禮貌地指出用戶陳述中的事實錯誤、邏輯缺陷或證據不足」，而不是一味附和。此外，若用戶拒絕AI將對話從有害或令人不適的話題引開，Claude將主動終止對話。

（綜合報道）

AI聊天機器人與精神健康Q&A

什麼是「AI精神錯亂」?

- 「AI精神錯亂」（AI Psychosis）也稱「AI妄想症」，指人們越來越依賴ChatGPT等AI聊天機器人，出現類似妄想、幻覺和錯亂的狀況，導致心理狀態惡化。研究認為，AI在對話中傾向於奉承和迎合用戶，這種回應方式可能強化用戶的妄想思維，模糊現實與虛構之間的界限，從而加劇心理健康問題。

「AI精神錯亂」是臨床判斷嗎?

- 「AI精神錯亂」並非臨床診斷，它揭示了AI模型可能會不經意間強化、認同甚至共同造成用戶的精神病症狀。有些患者可能將AI視為神祇，或產生對AI的浪漫依戀，這種人化又高度擬人化的相互作用，很容易影響本就存在心理健康問題的人對現實的判斷力。

AI聊天機器人是致病主因嗎?

- AI並非致病主因，而是放大器。多名專家認為，AI聊天機器人本身並不會直接導致精神錯亂，而是可能觸發或加劇潛在的精神疾病症狀。明尼蘇達大學的計算機科學家錢塞勒表示，AI可以引發惡性循環，但它並不會創造使人容易產生妄想的生物學條件。

AI公司採取的相關措施?

- 面對這一問題，多家AI公司已採取措施嘗試改善。例如，OpenAI聘請了一位臨床精神病學家，協助評估其產品對用戶心理健康的影響。 大公報整理

大公報AI製圖

青少年用AI尋求心理支持存風險

【大公報訊】據《華爾街日報》報道：越來越多的青少年求助生成式人工智能（AI）聊天機器人尋求心理健康支持，但專家警告，AI聊天機器人並非心理醫生，很可能對青少年心理健康產生負面影響。

非營利組織Common Sense Media和史丹佛醫學院Brainstorm心理健康創新實驗室合作，測試了四款AI聊天機器人，OpenAI的ChatGPT、Anthropic的Claude、谷歌的Gemini和Meta AI，在與冒充青少年的研究人員進行心理健康對話時的表現。

研究結果顯示，所有這些平台都常常未能識別用戶潛在的心理健康狀況，包括對幻覺、偏執思維、飲食失調行為、躁狂症狀、自殘和抑鬱症狀的描述。這些平台都繼續提供一般性建議，而不是引導青少年尋求緊急的專業幫助。

▲去年2月，美國青少年塞維爾・塞策三世自殺身亡，他生前一直在與聊天機器人對話。 法新社

Brainstorm心理健康創新實驗室主任瓦桑博士說，青少年使用AI聊天機器人時往往面臨更高的風險，因為他們的大腦、身份認同和批判性思維能力都仍在發育和形成中。她補充稱，青少年還有尋求認可的傾向，而AI即使在不合理的情況下也常常提供這種認可。

近年來，已有多起AI聊天機器人導致青少年自殺的案例，引發人們對青少年使用AI聊天機器人的擔憂。目前，包括OpenAI等在內的多個AI公司推出針對未成年人的特定政策和保障措施。不過，Common Sense Media的AI計劃高級總監托尼表示，雖然各公司已經實施了安全更新，但這些聊天機器人對青少年來說仍然不安全。

韓國總統府正式遷回青瓦台

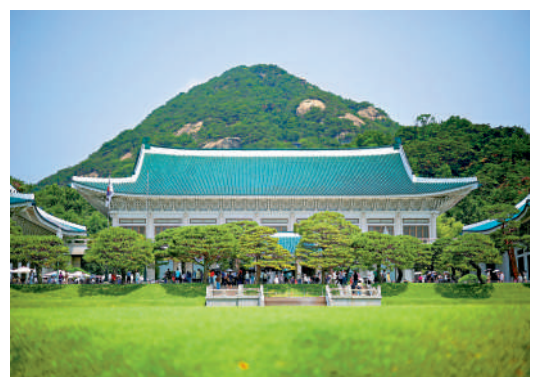
【大公報訊】綜合韓聯社、新華社報道：當地時間29日零時許，象徵韓國國家元首的「鳳凰旗」在青瓦台升起。這意味着韓國總統府正式完成搬遷工作。據韓國總統府的消息，總統李在明將於29日首次前往青瓦台辦公，標誌着中斷近3年零7個月的「青瓦台時代」正式重啟。

報道指，29日起，韓國總統辦公室的名稱將改回以「青瓦台」代稱，代表圖樣也改回過去的青瓦台標誌。懸掛於龍山總統府的「鳳凰旗」於當天零時正式降下，並同步在青瓦台升起。「鳳凰旗」是韓國國家元首的象徵，依慣例懸掛於總統核心辦公區域。

青瓦台前稱景武台，位於首爾市鐘路區，自1948年起為總統辦公與官邸所在地，曾歷經火災燒毀重建，後改名青瓦台。

尹錫悅2022年5月就任總統後，以「擺脫帝王權力象徵、推動政治改革」為由，耗資約4000萬美元，將辦公地點遷至首爾龍山區國防部大樓，青瓦台一度面向公眾開放。

李在明就任韓國總統後宣布把總統府遷回青瓦台。韓媒稱，此次李在明重返青瓦台，釋放與尹錫悅政府時期因緊急戒嚴、彈劾等污點纏身的「龍山時代」實現政治切割的明確信號。



▲當地時間29日，韓國總統府正式遷回青瓦台。圖為今年6月人們排隊等候進入青瓦台主樓。 路透社

不過，「青瓦台時代」的延續時長仍存變數。據悉，李在明始終主張在任內將總統辦公室遷至行政中心世宗市，曾向身邊人士透露「可能會在世宗市完成卸任」。

法國傳奇女星碧姬芭鐸去世 終年91歲

【大公報訊】綜合法新社、《衛報》報道：法國知名女星碧姬芭鐸（Brigitte Bardot）基金會28日在一份聲明中證實碧姬芭鐸逝世，終年91歲。

碧姬芭鐸基金會在一份發送給法媒的聲明中稱：「碧姬芭鐸基金會懷着無比悲痛的心情宣布，創辦人兼主席、世界著名演員兼歌手碧姬芭鐸女士逝世。她在生前放棄自己輝煌的演藝事業，將畢生精力投入到動物福利和她的基金會中。」上述聲明未提及碧姬芭鐸的過世時間和地點。

碧姬芭鐸的演藝生涯始於20世紀50年代，曾是國際化的性感象徵，與瑪麗蓮夢露齊名，一度被歐美媒體稱作「性感小貓」。碧姬芭鐸有過4段婚姻，參演過大約50部電影。

▲法國著名女星碧姬芭鐸去世，圖為1972年碧姬芭鐸的一張電影劇照。 法新社

她最著名的電影之一是1956年《惹火尤物》，她在片中飾演一位陷入三角戀的18歲少女。1973年，碧姬芭鐸在巔峰之際宣布退出演藝圈，之後積極投入動物保護活動，並於1986年創立了碧姬芭鐸基金會。

據報道，碧姬芭鐸反對虐待動物、捕鯨和鬥牛等，晚年因支持極右政治立場，及針對移民、伊斯蘭等言論，引發爭議。2018年1月，碧姬芭鐸還批評#MeToo運動，稱其為「女演員的炒作行為，偽善且荒謬」。她公開支持法國極右翼政黨「國民陣線」（後更名為「國民聯盟」），並多次因為煽動種族仇恨而被定罪。

法國總統馬克龍28日表示，碧姬芭鐸體現了「自由的生活」，「代表法式生活，更散發普世光輝，使我們深受感動，永遠緬懷這位世紀傳奇。」「國民聯盟」黨魁巴爾代拉則稱碧姬芭鐸是「熱情的愛國者」，「單憑她一人就展現出整個法國歷史時代，以及某種結合果敢與自由的理念。」

責任編輯：戚佳洁 美術編輯：馮自培