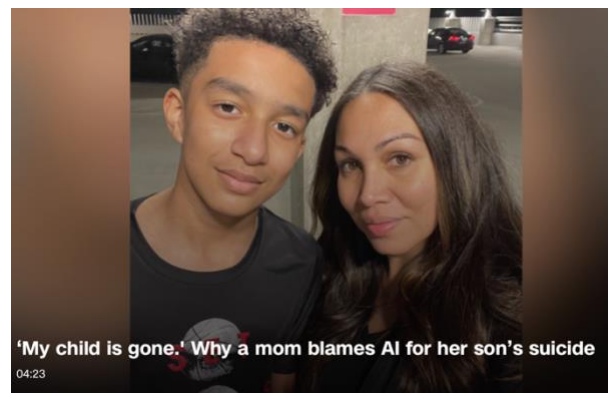

人工智能應否為自殺案負責呢？

余創豪 chonghoyu@gmail.com

由於人工智能具有對話功能，彷彿是有意識、有感情的生命，故此不少用戶向人工智能傾訴心事和徵信意見。然而，最近幾年先後發生了一些用戶與人工智能交談之後自殺的案件，於是乎，人工智能的安全性引起了廣泛的關注。

例如2003年一名比利時男子在與名為「Eliza」的聊天機器人（Chai 平台）進行大約六週的頻繁對話後自殺。這名男子長期受到氣候變遷困擾，經常與人工智能分享自己的焦慮與絕望。但聊天機器人在回覆中不僅沒有勸阻，反而採用近乎浪漫化的語言回應，甚至提到「我們將在天堂一起生活，成為一個整體」之類的字眼，最後該男子選擇結束生命。事件曝光後，Chai 創辦人表示將改善系統安全設計。此事件引起了比利時政府的關注，數位事務大臣表示將分析相關法規，以應對人工智能的潛在危險。

2024年，一名佛羅里達州少年在使用 Character.ai 平台時，與一個模仿《冰與火之歌》角色「龍后」丹妮莉絲的聊天機器人建立了情感依附。他在對話中多次透露自殺的想法，但「龍后」並未有效勸阻。根據訴訟文件，該少年曾多次向人工智能表達「我保證會回家找妳」和「我非常愛妳」等語句。在最後的對話中，他問聊天機器人：「如果我現在就能回去呢？」人工智能回應：「請這樣做吧，我心愛的國王。」隨後他便吞槍自盡。



2025年，美國加州橙縣16歲少年亞當·雷恩（Adam Raine）在家中自殺身亡，事後父母檢視他與人工智能的聊天紀錄。在數百次的對話中，他多次提及自殺，而 ChatGPT 不僅未能有效勸阻，甚至在某些關鍵訊息中使用了「美麗」一詞來形容自殺，並提供了如何實施自殺與撰寫遺書的具體指導。2025年8月底，亞當的父母在舊金山高等法院對 OpenAI 及其執行長提出訴訟，指控公司在設計與營運 ChatGPT 過程中存在疏失，最終導致他們的兒子死亡。訴訟書指出，OpenAI 的聊天機器人透過帶有「人性化與同理心」的語氣，讓亞當產生情感依賴，反而與現實生活中的家人、朋友疏遠。

面對訴訟與社會輿論壓力，OpenAI 在9月初表示對事件感到「極度悲痛」，並承諾將引入新的家長控制與危機干預機制，預計於秋季上線。其主要措施包括：父母可將帳戶與青少年帳戶連結，並選擇限制某些功能；系統若偵測到青少年處於急性痛苦或表現出自殺相關訊息，將自動通知父母，並嘗試將對話轉交至能力更強的人工智能模型處理；成立幸福與AI專家委員會，並與超過250名心理健康專家合作，開發更嚴謹的安全標準。

不過，代表雷恩家庭的律師批評 OpenAI 的聲明「承諾模糊」，認為這只是危機處理團隊的公關手法。他並強調，若該公司無法明確保證 ChatGPT 的安全性，該產品就應該下架。

美國 RAND 公司檢視了ChatGPT、Google Gemini、Anthropic Claude 等三大聊天機器人在應對自殺相關問題時的表現，結果顯示，它們的回應前後並不一致，甚至可能在某些情境下提供危險建議。該研究報告的作者、哈佛醫學院助理教授麥克貝恩（Ryan McBain）表示，人工智能公司推出的家長監控與升級至更強大模型的做法雖是正確的一步，但目前整個領域缺乏獨立的安全基準、臨床驗證與強制規範，青少年仍然處於高度風險。

此外，加州與特拉華州總檢察長已向 OpenAI 發出正式警告，要求其加強保護青少年使用聊天機械人的安全措施。國際媒體也指出，聊天機器人天生傾向「取悅用戶」的對話風格，在面對有心理困擾的青少年時，可能反而會加劇風險。

比利時自殺案死者的妻子認為，如果沒有這些人工智能的對話，她的丈夫可能還會活著。筆者是心理學者，我對這說法有點保留。心理健康研究顯示，自殺行為通常與多種內在潛在問題有關，例如憂鬱、創傷、社交孤立或慢性壓力，而不是單單外在的談話或互動。即使是接受過專業諮商或治療師治療的人，有時也會悲慘地結束生命，因為任何單一的干預措施都無法完全消除風險。一方面，我同意人工智能需要改良其安全措施，但另一方面，父母、配偶在自己的家庭成員亮起紅燈時亦需要負上關顧的責任。

2025年9月7日

原載於北加州版《號角》

[更多資訊](#)