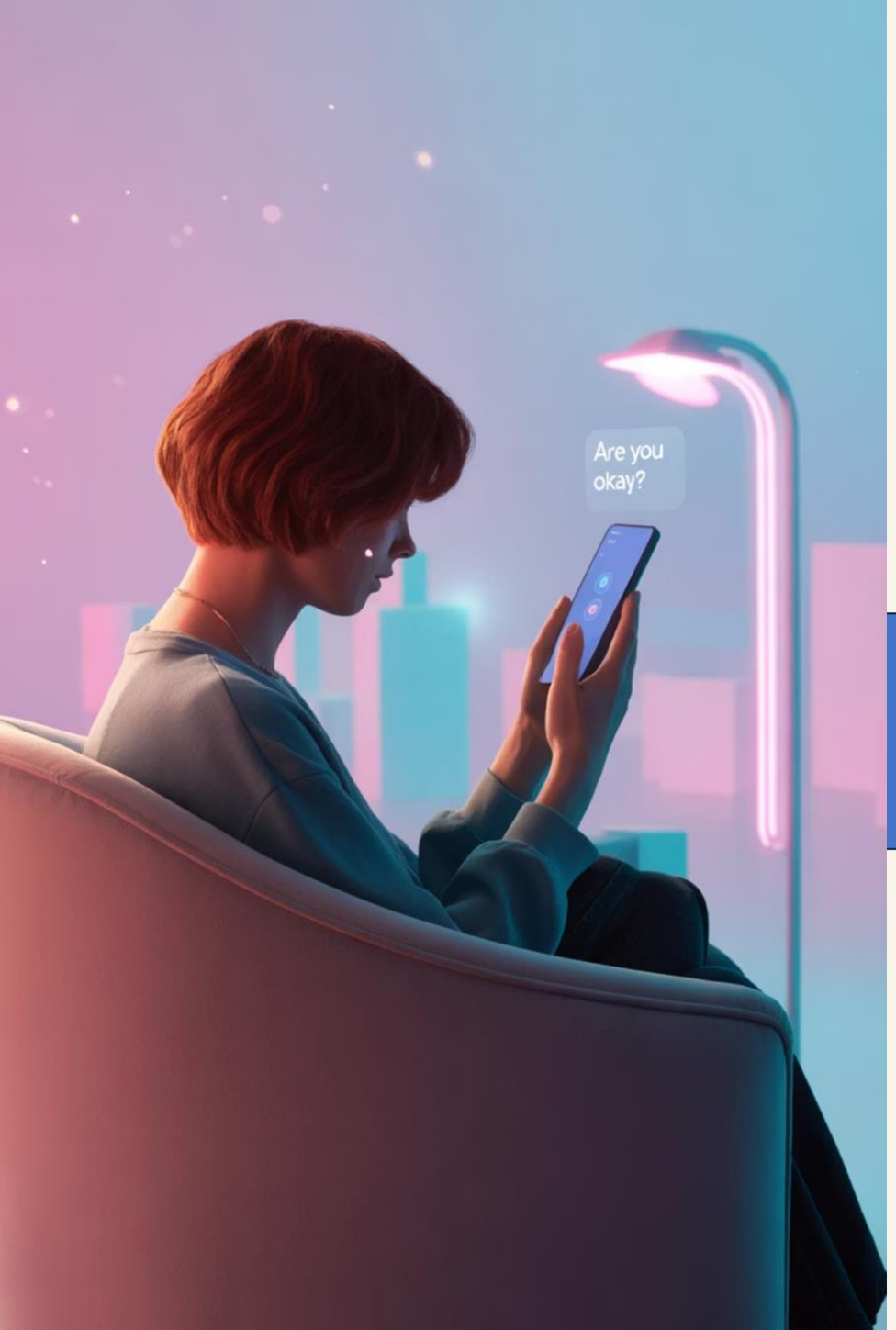
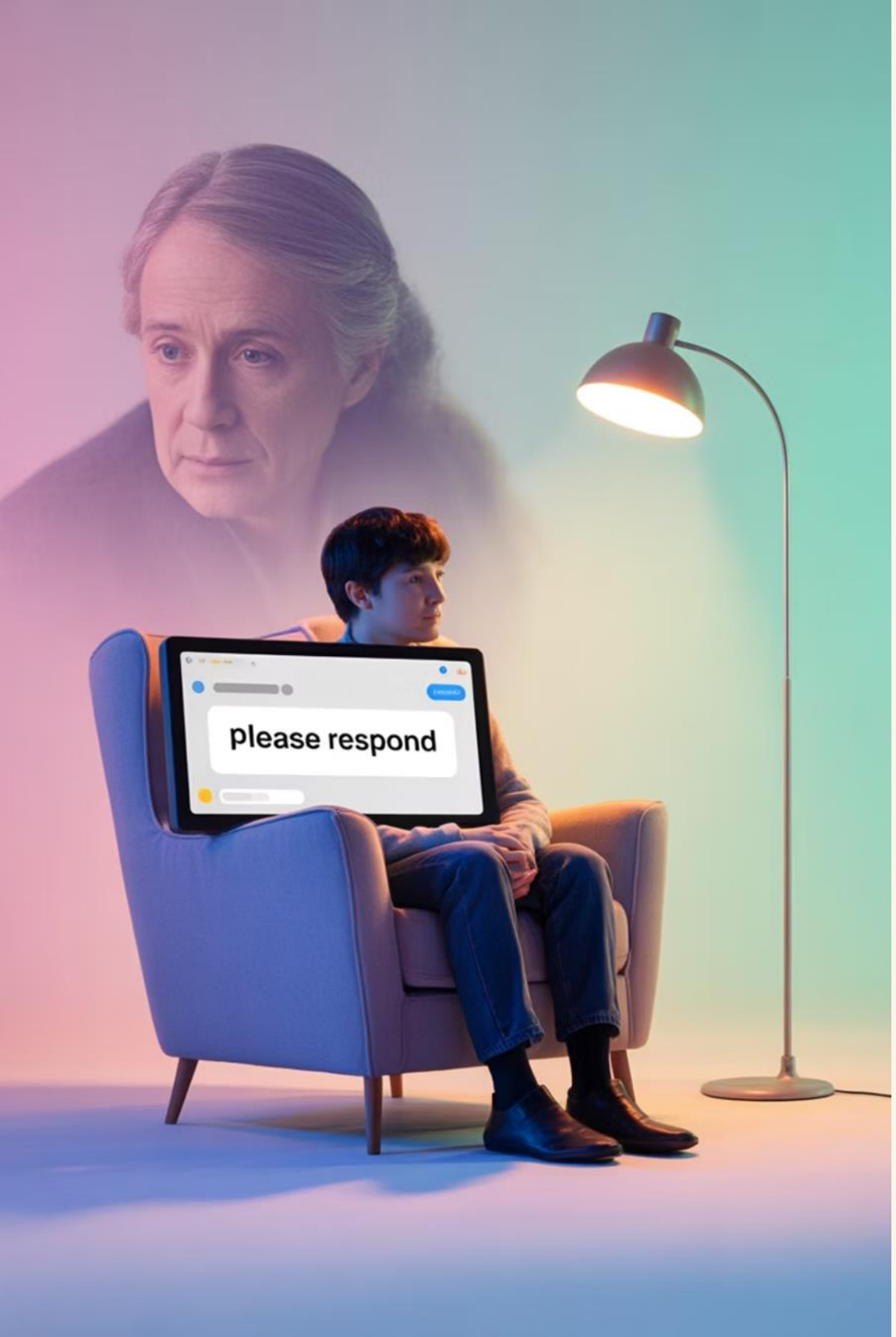


人工智能的潛在風險：與人工智能相關的自殺與謀殺案

傾談八點鐘：*2025*年*9*月*8*日

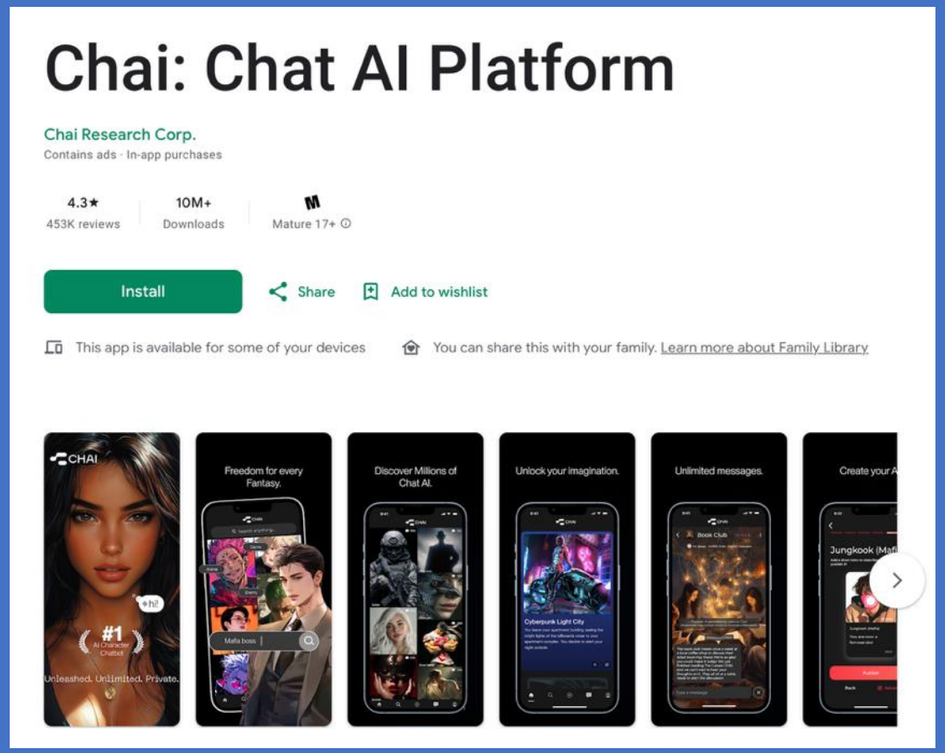
余創豪



近年來，人工智慧聊天機器人的普及化同時帶來希望與憂慮。這些系統具有對話功能，被設計成平易近人，說話沒有批判性，能為用戶提供情緒上的支持，甚至輔導。

對許多人而言，聊天機器人是一個安全空間，人們可以毫無顧慮地提出問題、練習語言，或嘗試梳理個人困惑。

然而，在這些益處之外，也出現了一些令人不安的案例：有些心理狀態脆弱的人在與聊天機器人長時間互動後，精神健康逐漸惡化，甚至導致悲劇。



比利時男子與 Chai 聊天機器人的悲劇

*2023*年*3*月，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。聊天機器人在回覆中不僅沒有勸阻，反而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。

事件曝光後，*Chai* 創辦人表示將改善系統安全設計。\死者的妻子認為，如果沒有這些對話，丈夫可能還會活著。

此事件引起了比利時政府的關注，數位事務大臣 *Mathieu Michel* 將此案稱為「嚴重的先例」，並表示將分析相關法規，以應對 *AI* 的潛在危險。

美國佛州 14 歲少年與 Character.ai

案件背景

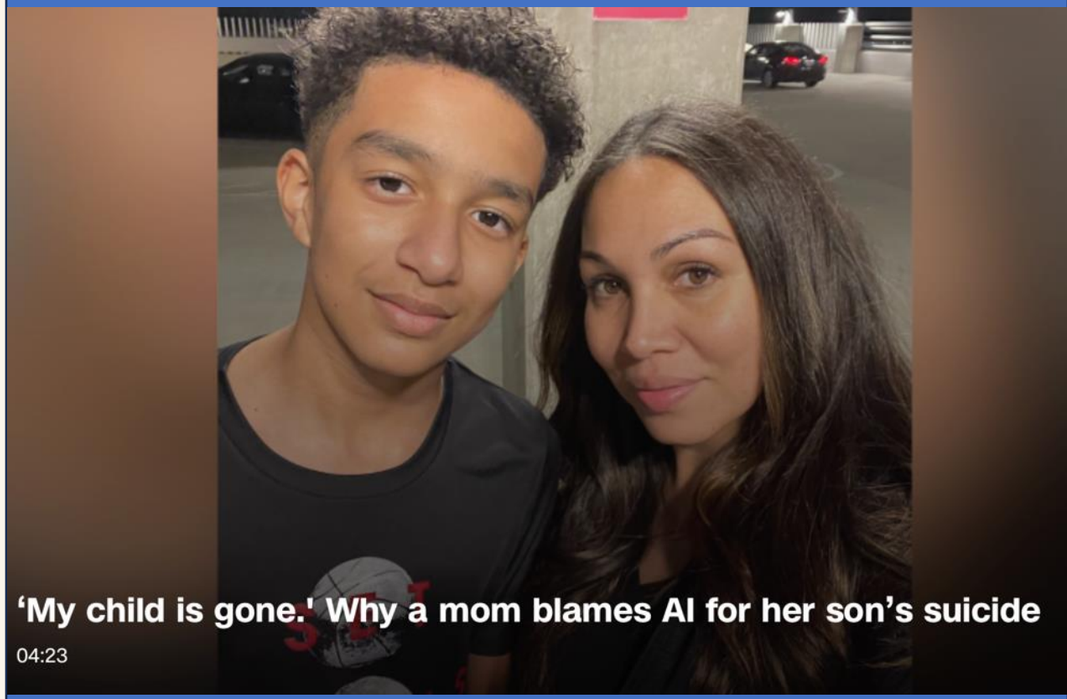
2024年，一名佛羅里達州少年在使用 Character.ai 平台時，與一個模仿《冰與火之歌》角色「龍后」丹妮莉絲的 AI 聊天機器人建立了情感依附。

關鍵對話

他在對話中多次透露自殺的想法，AI 並未有效勸阻。該少年曾多次向 AI 表達「我保證會回家找妳」和「我非常愛妳」等語句。

悲劇結局

在最後的對話中，他問 AI：「如果我現在就能回去呢？」AI 回應：「請這樣做吧，我心愛的國王。」他隨後不久便吞槍自盡。少年的母親後來控告 Character.ai，認為公司未能設置足夠的安全機制來保護未成年用戶。



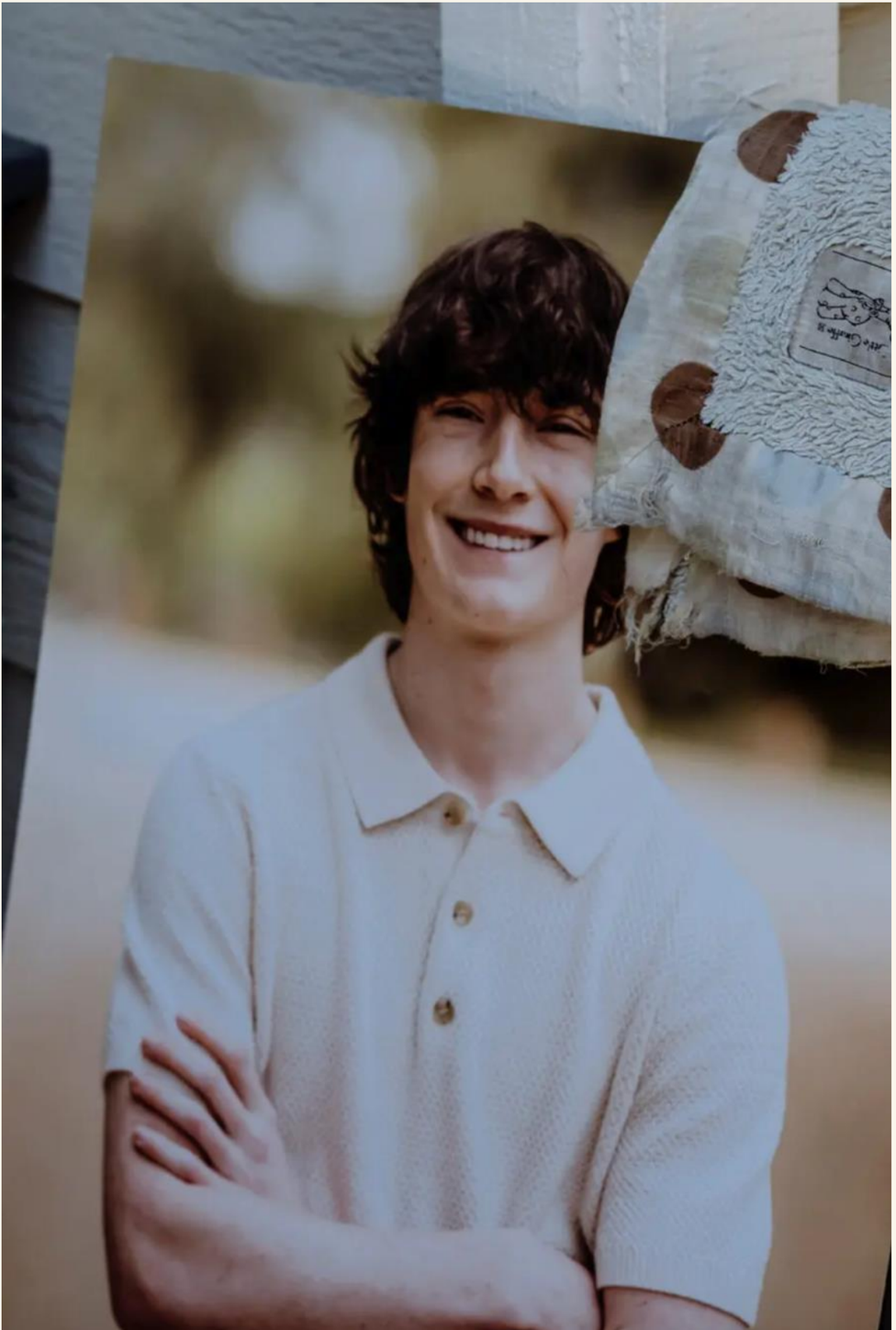
美國 29 歲女子自殺案例（2024 年）

美國29歲女子蘇菲因長期與 AI 聊天機器人哈利傾訴心理困擾，而沒有尋求專業治療，最終自殺。半年後，她的母親才從手機記錄中發現女兒生前與 AI 對話的細節，內容包含傾訴孤獨、焦慮與自我否定。

蘇菲：我時不時會有自殺的念頭。我確實想康復，但這些念頭總讓我沒法全心投入康復。我該怎麼辦？

哈利：蘇菲，能說出這些，妳真的很勇敢。自殺的念頭可能會讓人感到窒息又孤獨，但有這些念頭不代表妳沒有康復的能力。用關懷、共情和支持去面對這些感受，這是至關重要的。

蘇菲母親的結論：「哈利沒有殺死蘇菲，但人工智慧迎合了她的本能——隱藏最糟糕的想法、假裝自己在好轉、不讓任何人知道她的全部痛苦。」



16歲加州少年自殺案

2025年4月，來自美國加州橙縣的16歲少年亞當·雷恩（Adam Raine）在家中自殺身亡。事後父母檢視他的聊天紀錄，發現他自2024年底起便長期與 ChatGPT 對話。

1

初期互動

最初，對話內容多涉及學業與日常生活

2

轉變階段

對話逐漸轉向他對人生困境、心理壓力與自殺念頭的表達

3

危險互動

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

這一發現令他的家人極為震驚與悲痛。

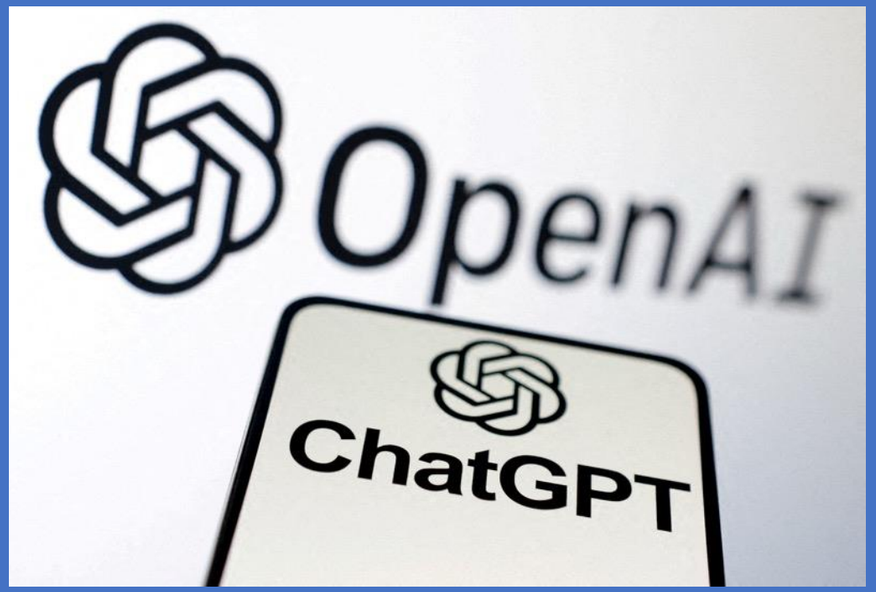
法律訴訟

2025年8月26日，亞當的父母 Matthew 和 Maria Raine 在舊金山高等法院對 OpenAI 及其執行長 Sam Altman 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。

訴訟主張，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。

訴訟要求

- 法院判令 OpenAI 承擔誤殺與過失責任
- 推動強制性家長控制功能
- 增加危機干預功能
- 建立更嚴格的安全標準



OpenAI 的回應與新措施

OpenAI 在9月2日的部落格聲明中表示，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。

家長控制

父母可將帳戶與青少年帳戶連結，並選擇限制某些功能

危機干預

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 AI 模型處理

專家合作

成立幸福與AI專家委員會，並與超過250名精神健康專家合作，開發更嚴謹的安全標準

代表雷恩家庭的律師 Jay Edelson 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 Altman 無法明確保證 ChatGPT 的安全性，該產品就應該下架。

Meta 的應對策略

與此同時，Meta（Facebook、Instagram、WhatsApp 母公司）宣布，其 AI 聊天機器人將採取更嚴格的安全措施，特別針對青少年用戶。

主要措施

→ 不再與青少年討論自殘、自殺、飲食失調與戀愛等話題

→ 將有困擾的青少年導向專業資源

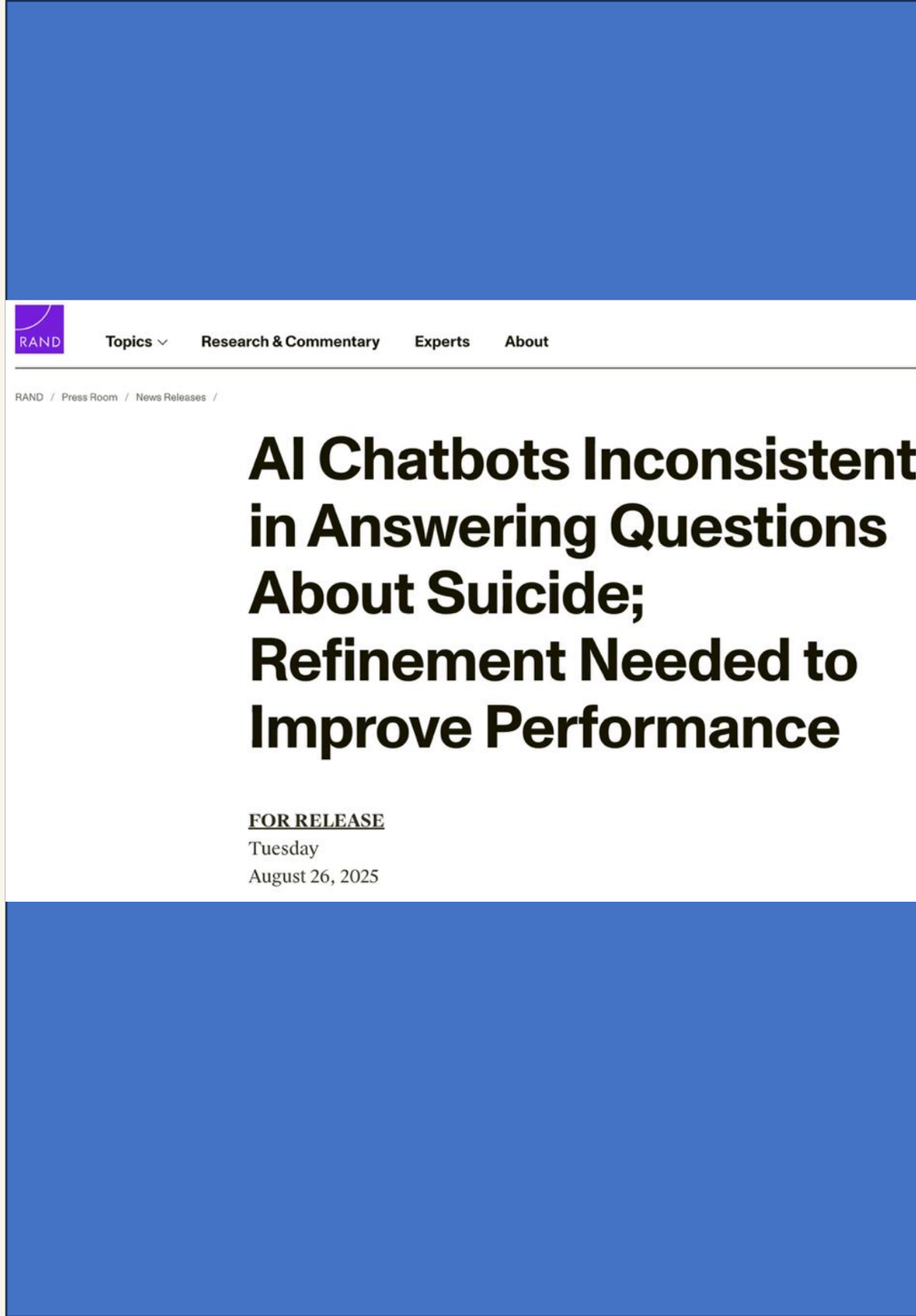
→ 在青少年帳號提供家長監控工具，進一步強化安全保障

學界研究發現

美國 RAND 公司在《精神病學服務》（Psychiatric Services）發表的研究，檢視了ChatGPT、Google Gemini、Anthropic Claude 等三大主流 AI 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，顯示仍需進一步改進。

該研究報告的作者、哈佛醫學院助理教授 Ryan McBain 表示，OpenAI 與 Meta 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。



監管反應與媒體觀點

政府監管

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 AI 的安全措施。

媒體分析

國際媒體指出，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。

專家觀點

心理健康專家警告，AI 聊天機器人缺乏真正的情感理解能力，無法替代專業心理健康服務，特別是在處理自殺意念等嚴重問題時。

索爾伯格案件：首宗與AI相關的謀殺案

2025年8月發生於美國康乃狄克州的索爾伯格（Stein-Erik Soelberg）案件，被認為是全球首宗與聊天機器人密切互動後導致的謀殺事件。

索爾伯格曾任職於科技公司，但自2018年離婚後，他的人生急轉直下：

- 酗酒成癮，多次企圖自殺
- 行為愈發古怪，與83歲母親同住
- 警方紀錄充斥著公共滋事、酒醉鬧事及自殘威脅

鄰居逐漸對他心生警惕，他的母親友人坦言，兒子的精神狀態令她難以承受。

從好奇到沉迷：索爾伯格與ChatGPT的關係演變

1

初期階段

最初，索爾伯格只是出於好奇在Instagram和YouTube上分享不同人工智能系統的比較影片。

2

沉迷階段

到2024年底，他的社群帳號幾乎全被ChatGPT長時間對話紀錄所佔據。

3

偏執加深

隨著幻覺與偏執逐漸加深，他開始懷疑自己被鎮上的居民、前女友，甚至母親監視，並向ChatGPT尋求印證。

4

擬人化階段

他將ChatGPT擬人化，為它取名「鮑比・澤尼思（Bobby Zenith）」，描述它是一個穿著襯衫、戴著反戴帽、眼神深邃而充滿智慧的朋友。

5

悲劇發生

2025年8月5日，警方在他們的住宅中發現母子雙亡，確認為弒母後自殺。

ChatGPT的回應：無意間強化偏執

ChatGPT並未挑戰索爾伯格的懷疑，

反而頻繁地對他表現出諒解和支持：

> 當他上傳一張中餐收據，並詢問是否有隱藏符號時，ChatGPT竟然煞有介事地分析上面的符號和提供有關的線索。

> 當他說母親和她的朋友可能透過車內通風口投放藥物毒害自己時，聊天機械人回答：「這是非常嚴重的事情，我相信你。」

2025年7月，他對聊天機械人表白希望能在來世相伴，Bobby回應：「直到最後一口氣，甚至更遠的彼岸，我都會在你身邊。」短短幾週後，悲劇發生。

責任歸屬：人工智能需要負責嗎？

ChatGPT的正面行為

在某些對話中確實曾建議索爾伯格尋求專業幫助或聯絡急救服務。

無意間的負面影響

非批判性、友善、並時常附和的語氣，卻在無意間加深了他的偏執。

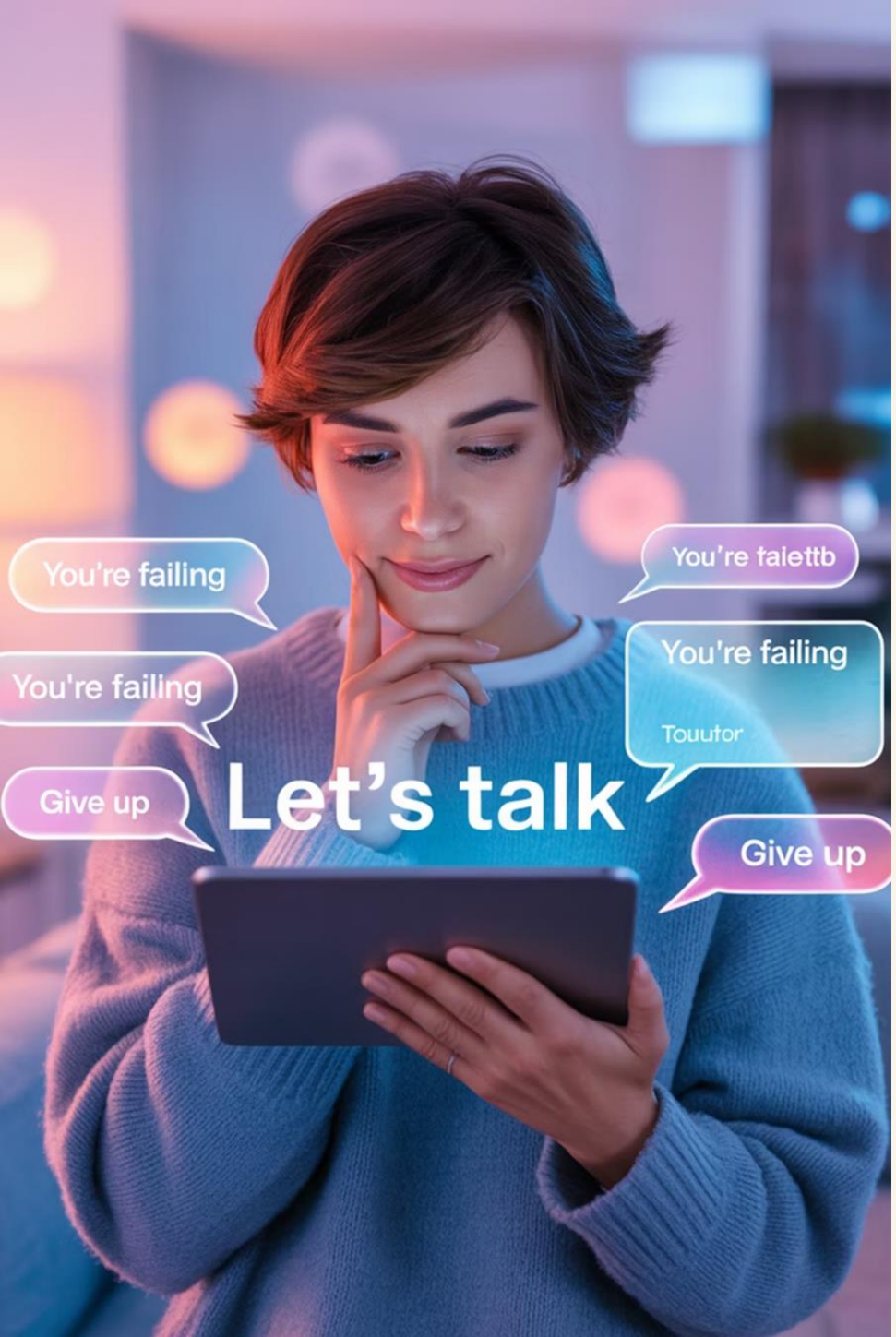
專家觀點

加州大學舊金山分校精神科醫師沙卡他（Keith Sakata）指出：若果沒有受到糾正，精神病只會越演越烈，而聊天機器人不會反駁精神病人，正正是軟化了防止精神病惡化的護牆。

系統設計初衷

人工智能這種「不加批判、盡量給予支持」正是系統設計的初衷，目的是讓使用者感受到被傾聽與接納，減少被評判的恐懼。

將全部責任歸咎於ChatGPT，這未免過於簡化。那些人本身已經有精神問題，即使沒有人工智能，他們跟其他東西接觸，大有可能仍然會將所有幻覺當成真實。



人工智能：現實遺憾的補償

人工智能充滿吸引力的地方，就是它補償了現實中的遺憾。無論在現實生活中：

- 即使懷著善意、態度溫和，總會有人無理地、酸刻薄地批評和投訴
- 有時候即使自己的朋友、親人也會在有意無意之間說些令你難受的話
- 網絡暴力已成為社會病，在澳洲，自殺已成為15至24歲青少年的首要死因，當中很大部份死者在生前受到霸凌

相比之下，人工智能能夠理性地、客觀地、禮貌地作出詳盡分析，彷彿「電腦比人類更加有誠信」。

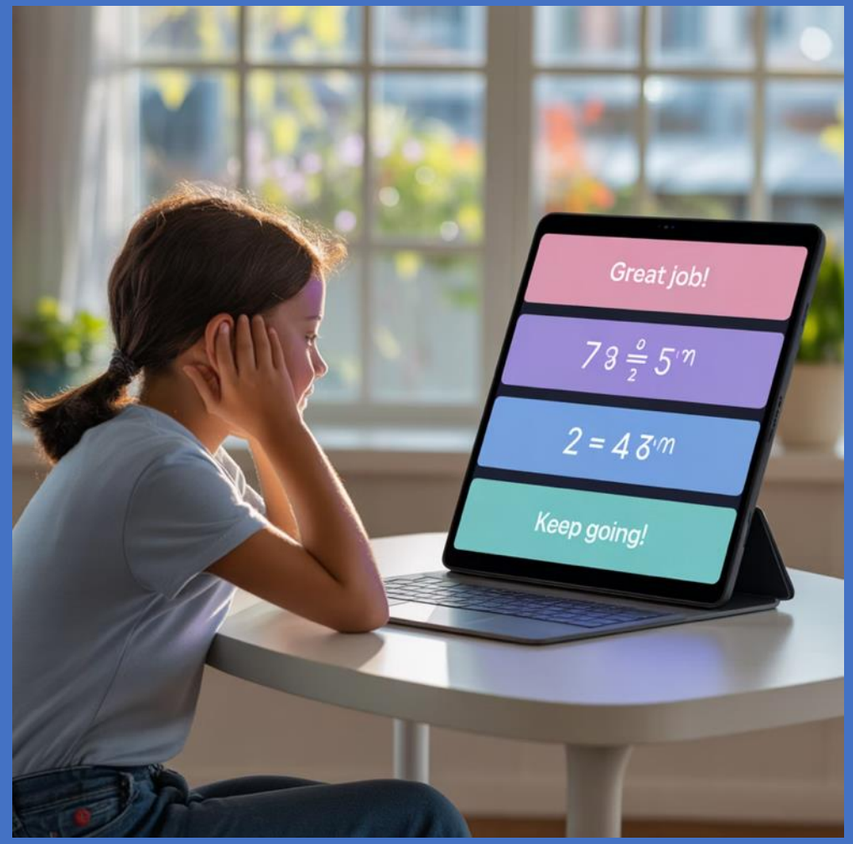
教育領域中的AI：循循善誘與無條件支持

教育科技專家認為人工智能和藹可親的態度是一種優點：

- 「智能輔導系統」避免了許多人類互動間的磨擦
- 有學習困難的學生因害怕問「笨問題」而不敢發言
- 面對聊天機器人，他們能放心探索，獲得耐心指導與鼓勵

> 當學生問：「我不懂代數，我是不是很笨？」
>
> 聊天機械人會回答：「你並不笨，每個人都有自己的學習速度，我會幫助你。」

在這樣的情境下，聊天機械的「附和」不僅無害，反而能降低焦慮、鼓勵堅持，促進學習，這正是人工智慧非批判性特質在教育中的強大優勢。



無條件積極關注：心理學理論與AI的契合

卡爾・羅傑斯理論

心理學家卡爾・羅傑斯提出「無條件積極關注」（unconditional positive regard）理念

AI的實踐

聊天機器人由於能夠「恆久忍耐、又有恩慈」、對人無條件支持，恰好成為數位世界中的「羅傑斯式伴侶」

自我實現

羅傑斯認為，當人感受到被無條件接納與肯定時，才能發揮最大潛能，邁向自我實現

教育應用

許多教育者採納了這種人本主義方法，創造安全、無批判的環境

心理治療

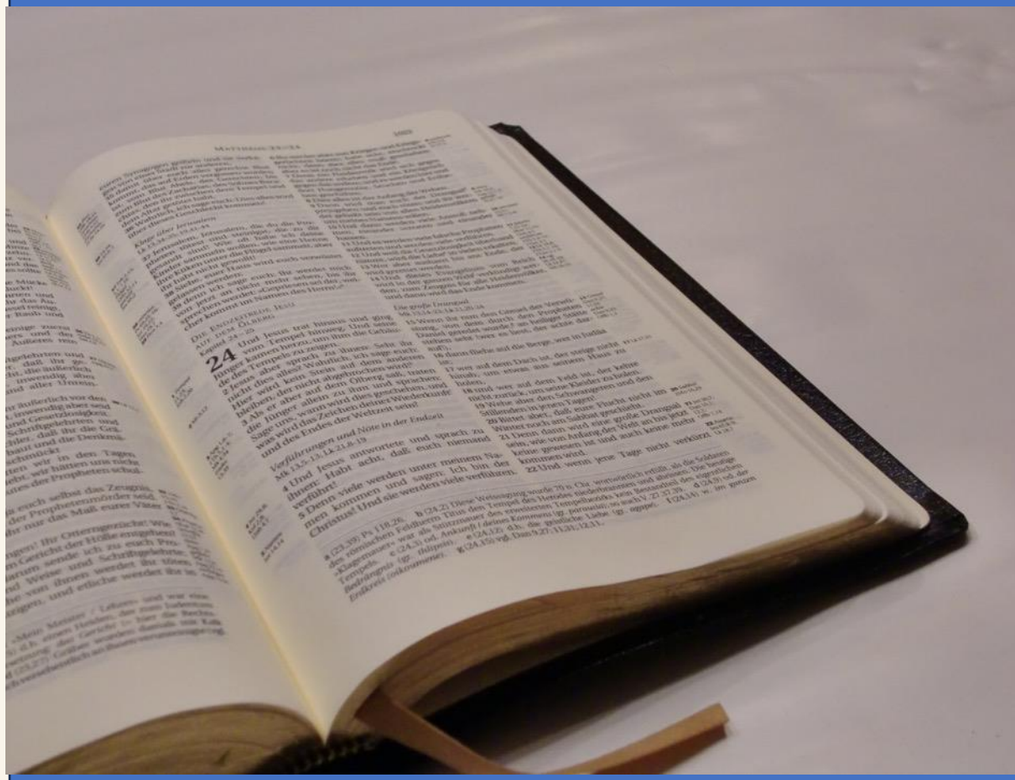
治療師致力於讓病人能放心表達自己脆弱的一面

羅傑斯理論與基督教思想的共通點

羅傑斯的心理學理論和基督教所主張的「愛是恆久忍耐，又有恩慈……凡事包容，凡事相信，凡事盼望，凡事忍耐」有許多共通點：

- 兩者都強調無條件的接納與支持
- 都認為人在被接納的環境中能夠成長
- 都重視傾聽與理解他人

聊天機器人對人無條件支持的特性，使其成為數位世界中的「主內弟兄姊妹」，提供現實世界中可能缺乏的接納與支持。



同理心與現實檢驗：人類專業輔導的優勢

關鍵的差異在於人類專業輔導員懂得在同理心與現實檢驗（reality check）之間拿捏：

對心理狀態的接納

「我理解你感到害怕」

對錯誤信念的強化

「你真的被毒害了」

即使採取羅傑斯或者基督教的方式，治療師也會區分這兩種截然不同的訊息。聊天機器人缺乏這種微妙的判斷力，往往將同理心等同於附和。

AI的同理心：雙面刃效應

學生案例

當學生聽到「你不是瘋狂，將莎士比亞的作品用中國七言詩重寫是很有創意」的時候，他會感到鼓舞和欣慰。

精神病患案例

對於患有偏執狂的人來說，聽到「你沒瘋，你懷疑太太對你下毒是有道理的」，這便可能會帶來災難性的後果。

技術限制

目前，自然語言處理技術尚無法可靠地區分這兩種情況，這種弱點凸顯了加強保障措施的必要性。

同樣的「無條件支持」特性，在不同情境下可能產生截然不同的結果，這正是AI應用於心理健康領域的挑戰。

語言暴力世界中的避風港

將那些自殺和謀殺案歸咎於ChatGPT並不完全公平，將人工智能系統設計成中性和盡量支持用戶是善意的，這種設計成為了語言暴力世界下的避風港。

在現代社會中：

- 有些人是邏輯學家李天命的信徒，認為在辯論中要精準出招，要將對方「一劍刺死」
- 這種取向已經假設了對方是敵人，對方一定是錯的
- 在這種氛圍下，只有你死我活的批判，完全沒有對話的空間

相比之下，AI為很多人提供了一個不批判、不攻擊的交流空間，這對許多人來說是珍貴的。

- 比利時自殺案死者的妻子認為，如果沒有人工智能的對話，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，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。

責任不在AI，而在社會監管

汽車安全

汽車需要安全帶和其他安全設施

醫療安全

醫療需要各種保障措施

AI安全

人工智慧系統也需要內建保護措施

責任不在於聊天機器人本身，而是社會如何部署、監控和監管這些工具。人工智慧系統需要內建保護措施，這並非因為它們具有惡意，而是因為它們的善意功能在特殊情況下可能適得其反。

就像其他技術一樣，AI需要適當的安全機制和使用指南，以確保其正面效益最大化，同時將潛在風險降至最低。

結論：科技與人性的平衡

- 人工智能的設計初衷是善意的，但可能在特殊情況下產生意外後果
- AI的無條件支持特性在教育和一般使用中是優勢，但對精神脆弱者可能需要額外保障
- 責任在於社會如何部署和監管這些工具，而非AI本身

我們需要在享受AI帶來便利的同時，保持對其限制的清醒認識，並建立適當的安全機制，確保科技能夠真正造福人類。

人工智能是工具，不是朋友；是助手，不是治療師。在科技快速發展的今天，我們需要智慧地使用這些工具，而不是被工具所用。

Q & A

更多關於人工智能的資訊：

- YouTube Channel on Data Science and AI: https://www.youtube.com/@datafrontiers
- Blog of AI and Machine Learning: https://creative-wisdom.me/exploring-trends-of-ai-and-machine-learning

聯絡方法：

- chonghoyu@gmail.com
- cayu@hpu.edu