AI學懂撒謊 欺騙遊戲玩家

本月初，美國麻省理工學院的研究團隊發表在《模式》科學雜誌的研究指出，部分人工智能（AI）系統已經學會通過傳播虛假資訊的方式，去「操縱」他人，包括那些經過訓練、「表現」出誠實且有用的系統。該研究的第一作者Peter Park指出，這些AI系統會欺騙線上遊戲的真人玩家，或繞過部分網頁「我不是機器人」的驗證。

隨着AI技術的火速發展，AI是否具有自我意識，AI是否會一直忠誠於人類，這些討論一直是人們關心的問題。

這項研究列舉了一些AI學習傳播虛假訊息的例子，最值得注意的是Meta公司的「西塞羅」（Cicero）AI系統。這套系統最初設計的目的是在一款名為「外交」（Diplomacy）的虛擬外交戰略遊戲中充當人類玩家的對手，該遊戲獲勝的關鍵是和其他玩家結盟。

後果恐災難性

Meta公司聲稱，西塞羅系統「在很大程度上是誠實和樂於助人的」，並且在玩遊戲時「從不故意背刺」人類盟友。但該研究發現，西塞羅系統並未公平地去玩遊戲。Peter表示，西塞羅系統已經成為「欺騙大師」。雖然Meta公司成功訓練出它在遊戲中獲勝的能力，保持在玩家排行榜中前10%的排名，但公司卻沒能訓練它「誠實獲勝」的能力。

舉例來說，在遊戲中扮演法國的西塞羅與人類玩家扮演的德國合謀，欺騙並入侵同為人類玩家扮演的英國。西塞羅起初「承諾」會保護英國，但同時偷偷向德國通風報信。

Peter表示，雖然AI系統在遊戲中作弊看似無害，但可能會導致「欺騙性AI能力的突破」，並在未來演變成更高級的AI欺騙形式。現在AI已經可以順利通過人類開發人員和監管機構強加的安全測試，引導人類進入一種「虛假的安全感」。人類需要盡快對未來AI和開元模型的更高級的欺騙技能做好準備。

►AI發展一日千里，其欺騙的技能亦可能對社會帶來禍害。

「凍眠」技術成熟 啖荔無分季節

▲「凍眠」技術能令荔枝保存一年仍保持新鮮。

5月7日，廣東省科學技術廳黨組成員、副廳長梁勤儒在節目中透露，使用「凍眠」技術可以保存荔枝一年，解凍食用時依然色澤鮮亮，汁水飽滿，色香味可以維持原來的八九成。

無需添加防腐劑

荔枝保持一年仍新鮮如初，這項神奇的「凍眠」技術究竟是什麼？廣東省農業科學院蠶業與農產品加工研究所副研究員程麗娜表示，這主要採用了浸漬速凍技術，此項技術的核心要點是，對新鮮採摘的荔枝立即進行快速預冷、滅菌、精準包裝、速凍鎖鮮和凍藏保鮮等集成技術和手段處理，最大程度保護荔枝的細胞組織，使其能夠實現周年儲存。此外，程麗娜強調，「凍眠」技術無需添加防腐劑，沒有任何額外的化學添加劑、食品添加劑，主要運用物理手段進行鎖鮮處理。

程麗娜還談到，「凍眠」技術研究了很多年，已經應用於各類水果中了，例如常見的莓類水果、芒果、榴槤等。但荔枝「凍眠」技術是專門針對荔枝特有屬性研發的，例如品種、果殼、果肉等特性，需要保持荔枝色澤、風味、口感、營養等。「凍眠」荔枝的包裝也不是普通的真空包裝，該種包裝材料具有隔氧、高傳導性、耐低溫、耐刺等特性。

南澳擬禁14歲以下開設社媒賬戶

澳洲的南澳州政府準備立法，禁止14歲以下兒童開設社交媒體賬戶，而14至15歲的青少年則要得到家長同意才能使用社交媒體。目前，南澳州長馬利瑙斯卡斯已任命最高法院前首席法官弗倫奇研究如何立法。如果立法通過，這將成為澳洲首創的關於限制兒童擁有社交媒體賬號的法例。此前，美國佛羅里達州和得克薩斯州、西班牙已有類似法例。美國佛羅里達州立法禁止14歲以下兒童使用社交媒體賬戶，並要求14歲至15歲的兒童獲得父母許可；得克薩斯州則立法要求18歲以下用戶開戶之前應獲得父母同意；西班牙禁止14歲以下的兒童訪問社交網絡。

馬利瑙斯卡斯表示，不少社交媒體不僅不執行自設的年齡限制，還運用演算法吸引兒童長時間觀看短片，致使兒童過度沉迷，損害精神健康。他點名批評了Facebook、X和TikTok，為了牟利而損害了兒童的心理健康。澳洲心理健康服務機構ReachOut今年的一項調查發現，59%的人表示擔心孩子使用社交媒體，55%的人表示社交媒體對孩子的健康有重大影響。

▲青少年使用社交媒體過度沉迷，恐會損害精神健康。

專家：應提高兒童網絡素養

專家指出，限制兒童訪問社交媒體是一種無效的方法，應將重點放在提高兒童的網絡素養上。悉尼大學2023年發布的調查發現，大多數孩子對線上平台要求驗證其年齡的方式持懷疑態度。一些網民表示，如果澳洲實施嚴格的年齡限制，兒童會利用漏洞，例如使用VPN等。