

AI「產生幻覺」致假消息氾濫

舊詞獲新解 入選年度熱話

【大公報訊】綜合《衛報》、美聯社、彭博社報道：人工智能（AI）全球爆紅，英國劍橋詞典15日將「Hallucinate」，即「產生幻覺」一詞評選為本年年度詞彙。劍橋詞典稱，隨着AI的發展，該詞於今年獲得了新的含義：AI在處理信息時亦可能會「產生幻覺」，從而「編造」出一系列看似具有說服力的錯誤信息，恐導致假消息氾濫。

「Hallucinate」的原始詞義是「看到、聽到、感覺到或聞到一些不存在的東西」的體驗，通常是由於健康狀況或服用藥物所產生。該詞今年獲得了全新的含義，即指代ChatGPT等AI系統產生「幻覺」，生成虛假信息的情況。

劍橋詞典在其官網稱，之所以選擇「產生幻覺」一詞作為年度詞彙，是因為新含義「觸及了人們談論AI的核心」。ChatGPT等生成式AI，是一種「強大」的工具，但「遠不完美」，人們仍在學習如何安全有效地與AI互動，而這意味著必須要先意識到AI的潛在優勢和其當前的弱點。

另外，劍橋詞典亦正式收錄了一系列與AI相關的詞彙條目進入詞典，包括大型語言模型（LLM）、生成式AI（GenAI）和GPT（生成式預先訓練轉換器的縮寫）等。

ChatGPT現「幻覺」幾率約3%

ChatGPT等聊天機器人爆紅，但不斷被外界發現，其實AI系統也會

「一本正經」胡說八道，可能會對原文增刪、竄改，甚至杜撰內容，這過程通常被稱為AI「幻覺」。例如，今年2月，谷歌發布AI聊天機器人Bard之際，其宣傳視頻中言之鑿鑿地聲稱，韋布望遠鏡拍攝了首張太陽系外

行星的照片。然而該說法實際是錯誤的，第一批系外行星的照片實際是由歐洲南方天文台的甚大望遠鏡（VLT）所拍攝。這個杜撰錯誤，一度讓谷歌市值蒸發1000億美元。

據統計，目前流行的AI模型，都有可能產生「幻覺」，其中ChatGPT產生「幻覺」的幾率約為3%，而谷歌的用於社交平台AI的Palm-Chat模型則擁有高達27.2%的發生AI「幻覺」幾率，為主要AI模型中最差。

需要批判性思維

目前，AI產生「幻覺」的原因目前仍未清楚，因為其背後的大型語言模型擁有海量數據，訓練過程極其複雜，難以確定具體的出錯環節。有研究人員認為，最主要的問題出在訓練AI算法的數據質量上。過時、低質量的訓練數據在訓練過程中，可能誤導AI模型。

隨着越來越多的領域都開始在工作上使用AI，這可能會促進假信息的傳播。專家指出，當人們利用AI技術尋找可能影響其健康、投票意向和其他潛在敏感話題的答案時，AI「幻覺」更會帶來風險。

The infographic features a central illustration of a human head with a brain, surrounded by various text boxes and images. It includes sections on what AI hallucinations are, how they happen, and their consequences. It also highlights a specific error made by Google's Bard AI. A large callout at the bottom left reads 'ChatGPT: Optimizing Language Models for Dialogue'.

何謂AI「幻覺」？
AI在回答相應問題時，輸出看似有理有據，實則虛假或荒謬的答案。

為何會產生「幻覺」？
造成AI「幻覺」的原因暫時未知。AI的大型語言模型是通過龐大數據庫訓練出來的，用來針對用戶提示做出響應，而這當中涉及多個過程，有些過程是系統自動的，有些過程則受到人為干預的影響，因此難以追溯是哪一個部分導致「幻覺」產生。

AI「幻覺」的後果？
AI給出的回答看似有理有據，但如果用戶對於答案內容涉及的相關領域不了解，可能就會接收了大量的假信息。

如何應對AI「幻覺」？
保證AI模型訓練數據的時效性與高質量；根據人類反饋，改進AI的算法；引入自動糾錯機制等措施，可以緩解AI「幻覺」的出現。

ChatGPT: Optimizing Language Models for Dialogue

AI預測天氣 90%結果勝過傳統方法



▲人工智能天氣預報系統GraphCast提前9天預測到了颶風「李」登陸加拿大。

預報中心和多國氣象局目前使用的傳統方法，則是使用超級電腦根據大氣物理學進行計算，完成相關預測則需要幾個小時。

GraphCast使用一種圖像神經網絡（GNN）的機器學習架構，以ECMWF天氣預報系統過去40多年的歷史資料作為訓練模型。ECMWF機器學習協調員錢特里表示，經過訓練的GraphCast運行成本非常低，能源消耗可能降低一千倍。

DeepMind研究人員表示，「標誌着天氣預報領域的一個轉折點」，提到了9月北大西洋的颶風「李」，作為成功預測的範例。GraphCast能夠在颶風登陸加拿大新斯科舍省前9天作出正確預測，而傳統方法僅能提前6天。不過，研究人員強調，暫時AI系統還不能完全取代傳統方法，但可以作為輔佐使用。

TikTok在澳洲成求職利器



▲希爾頓澳洲要求求職者上傳TikTok影片，替代傳統簡歷。

估，然後與入圍的候選人建立聯繫並邀請其參加面試。

希爾頓澳洲地區人力資源總監霍

格表示，該公司選擇TikTok試點是為了吸引從小接觸互聯網的Z世代（1990年代末出生）員工。她補充稱，當酒店需要具有良好人際交往能力、能夠處理賓客關係，以及其他方面事物的人時，僅從紙面上「不知道他們是否有能力」。

希爾頓委託Randstad公司進行的一項調查顯示，大多數澳洲Z世代求職者正在轉向TikTok、Instagram和Snapchat等平台。該研究發現，每四名35歲以下的調查對象中，就有一人通過社交媒体平台成功找到了工作。對於不使用TikTok的申請者，希爾頓表示仍將接受傳統的書面簡歷。

塑膠包裝污染環境 紐約州起訴百事

【大公報訊】綜合路透社及《紐約時報》報道：美國紐約州檢方周三對飲料巨頭百事公司提出起訴，指控百事製造的塑膠垃圾危害環境，並抨擊百事就環境問題對公眾發表「誤導性」聲明。

紐約州檢察長詹姆斯（Letitia James）在紐約州最高法院提起民事訴訟，尋求法院就百事公司對紐約水牛河造成「公害」一案做出裁決。詹姆斯請求法院對百事進行罰款並要求其賠償損失，要求百事停止銷售未標明對環境有害警告的一次性塑料產品，還要求該公司為生產產品在布法羅地區造成的污染提供補救措施，減少百事公司塑料產品進入河流的數量。

起訴書指出，檢方認為，百事生產



▲美國紐約州水牛河上漂浮大量的塑膠廢物。