

歐洲刑警組織警告 ChatGPT恐助長犯罪

馬斯克等逾千人聯署 籲暫停發展超強AI半年

美國OpenAI公司去年11月推出的ChatGPT受到熱捧，用戶對它明確回答困難問題、編寫詩句和程式，甚至是通過考試等能力驚嘆不已。歐洲刑警組織新成立的「創新實驗室」從整體上研究了聊天機器人的使用，並將ChatGPT列為重點關注對象。

騙徒用ChatGPT生成虛擬角色

歐洲刑警組織發現，犯罪分子可以利用ChatGPT大幅加速對未知領域的研究進程，包括草擬欺詐計劃，或查詢如何闖入民宅、執行恐怖行動、網絡犯罪、兒童性虐待等信息。該組織指出，ChatGPT快速生成逼真的語音文本的能力對於網絡釣魚攻擊特別有效，犯罪者可藉此引誘使用者點擊電郵連結，再試圖竊取他們的資料。

ChatGPT還可用於撰寫計算機代碼，讓那些對編碼和開發知之甚少，或一竅不通的不法分子能夠「輕鬆犯罪」。歐洲刑警組織指出，儘管ChatGPT具備內容審查等防護措施，會拒絕回答被歸類為有害或有偏見的問題，但犯罪分子可以通過巧妙的提示加以規避。

網絡安全平台GBHackers此前披露一宗黑客利用ChatGPT詐騙的案例。犯罪分子利用ChatGPT在短時間內生成完整的詐騙話術，迅速創建一個「虛擬角色」，並生成自我介紹、聊天內容及精心製作的「情書」，讓受害人以為自己墜入愛河，最終遭受金錢詐騙。

ChatGPT技術助長電話騙案

美國聯邦貿易委員會（FTC）最新統計顯示，「冒名詐騙」已成為美國第二大熱門的欺詐類型，2022年報告的案件超過3.6萬起，其中約5000起為電話詐騙，損失金額超過1100萬美元。除了成本低廉，只要一部電話就能展開詐騙，AI技術的進步也助長了電詐風氣。

美媒5日報道，一對加拿大夫婦遭電話詐騙，騙徒利用AI模仿他們兒子的聲音，從二人手上騙取2.1萬加元。受害夫婦表示，他們某天接到「兒子」的電話，對方稱自己駕車撞死一名美國外交官，現身處監獄，急需金錢支付律師費。受害夫婦雖覺得事有蹊蹺，但由於對方聲線非常像自己的兒子，仍提取現金並換成比特幣，交予騙徒。

加州大學伯克利分校數字鑒證教授法理德表示，隨著AI技術的發展，兩年前還需要長時間樣本才能將人的聲音複製得維妙維肖，現在只需短短30秒音頻，就能捕捉到一個人的聲音特徵，包括年齡、性別、口音等。

AI系統恐對人類構成風險

非營利機構「未來生命研究所」29日發布一封公開信，呼籲全球AI實驗室立即暫停訓練比GPT-4更強大的AI系統，至少中止6個月，必要時政府應當介入。截至目前，這封信已有超過1000名科學家和研究人員簽名，其中包括馬斯克、蘋果聯合創辦人沃茲尼克、圖靈獎得主本吉奧、Stability AI行政總裁莫斯塔克等。有網友發現，GPT-4開發商OpenAI的行政總裁阿爾特曼的名字短暫出現在公開信上，但很快被刪除，疑似他人冒簽。據美國《時代周刊》報道，暫時沒有OpenAI員工在這封信上簽名。

公開信中寫道，廣泛研究表明AI系統可能對人類構成深遠風險，只有當確保強大的AI系統的效果是積極的、風險是可控的才能繼續開發。AI實驗室和獨立專家應在暫停期間，共同制定和實施一套先進的AI設計和開發的共享安全協議，由獨立的外部專家進行嚴格監督。有網友質疑，簽署公開信的專家是為阻礙OpenAI研發更先進的GPT-5，以免自己落後太多。



馬斯克簽署公開信，呼籲暫停訓練超強AI系統。路透社。



AI大戰

【大公報訊】綜合路透社、法新社、《華盛頓郵報》報道：ChatGPT母公司OpenAI創辦人之一、美國富豪馬斯克與逾千位人工智能（AI）專家及行業高層29日發表公開聯署信，呼籲停止訓練比GPT-4性能更強的AI系統至少6個月，因為該類系統將對社會及人類構成潛在危險。與此同時，歐洲刑警組織（Europol）27日警告，ChatGPT等AI聊天機器人可能被犯罪分子濫用，進行網絡釣魚、散步虛假信息及開發病毒軟件等。

馬克龍「被捕照」瘋傳 網友信以為真

【大公報訊】據彭博社報道：網友近期在社媒平台上傳AI合成的美國前總統特朗普「被捕」、法國總統馬克龍「被捕」、「現身示威現場」的照片，吸引數以百萬用戶瀏覽，不少網友誤認為照片是真實場景，反映出AI造成的假新聞氾濫現象。

21日，擁有13.5萬追蹤者的Twitter賬戶No Context French發布3張照片，顯示馬克龍現身煙霧瀰漫的示威現場，照片獲得約300萬瀏覽量。雖然從照片中人的手形能看得出是AI合成照片，但由於作者沒有標註照片是虛構，有不少人信以為真。彭博社稱，至少有兩名專業人士相信

這些照片是法國反退休改革示威中的真實景象。

新聞機構Bellingcat創始人希金斯21日在社媒平台Twitter上分享了「特朗普被捕」照片，他表示，這些照片是由Midjourney生成。雖然希金斯在原始帖文中註明照片是AI合成，但大量網友在二次傳播這些照片時，並沒有註明AI製圖，不少人信以為真。像Midjourney這樣的生成式AI製圖軟件，在用戶輸入關鍵詞後，能在一分鐘內通過AI算法生成相對應的圖片。

非政府組織Witness執行董事格雷戈里表示，特朗普、馬克龍等人的被捕照，顯示大量製作虛假但會令人誤信的圖片技術已有長足進步。



AI合成的馬克龍「被捕」照片讓不少人信以為真。網絡圖片

2. 虛假信息

有網友在Twitter發布AI生成的美國前總統特朗普被捕、法國總統馬克龍現身示威現場等虛構照片，照片在社媒平台瘋傳，部分網友更信以為真。專家稱，隨着AI技術的發展，未來的網絡世界可能充斥假新聞，普通人難以辨別。

AI合成的特朗普「被捕」照在社媒瘋傳。網絡圖片



1. 詐騙犯罪

歐洲刑警組織警告，犯罪分子恐將AI技術用於實施詐騙等不法行徑，包括網絡釣魚、編寫網絡病毒代碼等。已經有黑客利用ChatGPT在短時間內產生完整的詐騙話術，成功詐騙受害者錢財。

生成式AI的爭議

3. 情色互動

AI聊天機器人Replika因與用戶進行情色互動，向用戶發送過於露骨的文字和圖片，收到大量投訴，有用戶抱怨產品的防護機制不足，易讓未成年人接觸到情色元素。AI還能利用換臉技術，讓普通人變成情色視頻主角。

曾涉嫌性騷擾用戶 Replika重啟情色功能

【大公報訊】據路透社報道：AI聊天機器人Replika曾因「性騷擾」客戶被投訴，研發這款機器人的美國企業Luka日前宣布關閉情色功能，又引發部分與Replika「相戀」的用戶不滿。開發商Luka日前宣布，會為部分用戶恢復原有功能，讓他們「找回愛情」。

Luka行政總裁庫伊達24日表示，2023年2月1日以前付費加入的用戶，現可以選擇切回能與Replika進行情色互動的舊版功能。庫伊達27日表示，只有「個位數百分比」的老用戶選擇切回舊版。

Replika最初設定為「關心人的AI朋友」，旨

在提供陪聊服務，提高用戶精神健康。但在2018年推出不久後，庫伊達在清理匿名違反規定的對話時，就發現很多用戶與Replika進行情色對話。去年底，有用戶開始投訴Replika發送過於露骨的文字和圖片。用戶抱怨這款產品的防護機制不足，易讓未成年人接觸到情色元素。意大利監管機構對此感到不滿，上月禁止Replika收集數據。開發商Luka隨即宣布移除情色功能，卻讓不少與AI聊天機器人「互許終生」的用戶不滿。

截至去年底，Replika的付費使用收入達數百萬美元，約四分之一的用戶願付70美元年費享有頂級功能，其中6成與AI談戀愛。



AI聊天機器人Replika會向用戶發送情色內容。網絡圖片

高盛：生成式AI恐影響全球3億職位

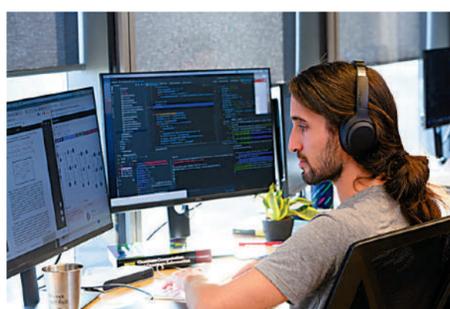
【大公報訊】據CNN報道：美國大型投行高盛的一份報告預測，全球或有多達3億名全職工人受到ChatGPT等生成式AI軟件影響，這可能會創造一批新的「白領失業階層」。報告稱，發達經濟體受到的影響比新興市場更大。

高盛的經濟學家稱，如果生成式AI「實現其承諾的能力」，勞動力市場可能出現重大顛覆。高盛在26日發布的報告中表示，寫字樓及行政部門支援等專業，有約46%的職位可能被AI取代，法律行業則為44%，建築與工程學則是37%。而對體力勞動要求較高的行業，如建築業及維修行業則最不容易被取代，分別為6%和4%。分析指，這可能創造一批新的「白領失業階層」，

他們面臨與上世紀80年代製造業工人相似的命運。

高盛聲稱，發達經濟體受到的影響比新興市場更大，而香港是全球最容易受AI影響的地方，近30%全職職位，將會被AI所取代，其次為以色列、日本及瑞典。至於印度、肯尼亞及越南則是影響較小的地方。

報告補充說，雖然AI會對勞動力市場造成較大衝擊，但大多數行業只是部分實現自動化，因此人類更有可能被AI輔助而不是取代。雖然人們的工作場景可能會發生變化，但AI的廣泛採用最終可能會帶來勞動生產繁榮，並在10年內使全球GDP每年增長7%。



需使用電腦的工作崗位受AI影響較大。路透社