AI以假亂真 2024年美國大選恐被操縱

ChatGPT創始人籲加強監管人工智能

【大公報訊】綜合《華爾街日報》、彭博社、美聯社報道：人工智能(AI)聊天機器人ChatGPT爆紅，引發全球關注。ChatGPT母公司OpenAI的行政總裁阿爾特曼(Sam Altman)16日首次出席美國國會聽證會。他在會上發出警告，聊天機器人能夠提供互動式虛假資訊，可能會影響明年的總統選舉，並呼籲政府加強對AI技術的監管。此外，阿爾特曼還提到有關AI影響就業以及過度集中於少數企業等問題，還建議政府成立專責機構加強規範。

▲美國共和黨全國委員會用AI製作競選廣告抨擊總統拜登。 網絡圖片

當天的聽證會在參議院司法委員會下屬的隱私、技術和法律委員會舉行，兩黨議員就AI的廣泛影響和監管措施向阿爾特曼提問。此次聽證會由美國參議員布盧門撒爾開場，他播放了由ChatGPT撰寫、並通過AI聲音程序朗讀的發言錄音。隨後他本人說，對這段音頻如此逼真的效果感到驚訝，儘管ChatGPT此次撰寫的發言稿內容與其本人立場一致，但仍可被利用，產生可怕的後果。

共和黨參議員霍利指出，生成式AI發展規模之大「可能會招致類似原子彈的嚴重後果」。阿爾特曼則表示，AI將以不可預測的方式改變社會，甚至給世界帶來「嚴重危害」，因此監管至關重要。

阿爾特曼建議美國政府考慮實施3項計劃，政府可成立監管機構，向人工智能公司發放牌照，如不符合相應安全標準，則吊銷牌照；建立一套安全準則規範AI模型；要求獨立專家學者對AI模型進行審核。同時，他支持設立一個國際組織爲AI制定標準，並舉出管制核武已有先例。

對於AI對人類就業的影響，阿爾特曼表示，「當前的一些工作確實會被AI取代，但同時也會創造新的就業機會」，政府需與企業合作以把控該影響。

AI可精準鎖定目標選民

目前，生成式AI技術對於選舉的影響能力已逐漸浮現。美聯社報道指出，該技術可以更加精準地向選民發放大量競選電子郵件、文本或圖片，甚至偽造圖片或者宣傳資料，事後幫助進行選情分析，再利用數據和算法來自動完成任務，如在社交媒體上精準鎖定目標選民或追蹤捐款人。

另外，AI可以通過深僞(deepfake)技術，給視頻中的人物「換臉」以及模仿特定人物的聲音，用來混淆視聽、誹謗競選對手甚至煽動暴力，例如呼籲選民在錯誤的日期投票；讓政治對手發表憑空捏造的爭議性言論，例如承認犯罪或者表達種族主義觀點；甚至可以生成假新聞，謊稱某人退出選舉。

上月，美國共和黨全國委員會在總統拜登宣布參選後發布了一則競選廣告，片頭出現「如果我們有史以來最軟弱的總統再次當選怎麼辦」的字樣，接着出現一系列AI生成的圖像，如經濟崩潰讓美國的店舖被封鎖、士兵和裝甲軍車街頭巡邏、紋身的罪犯和大批移民，並寫道，如果拜登2024年再次當選，這就是AI對國家未來的展望。此外，特朗普上個月就封口費案前往紐約曼哈頓出庭之際，網絡上流傳着由AI生成的特寫照片以及特朗普拒捕的照片，逼真程度讓網友難以辨別。

恐引發下次金融危機

此外，國會成員也對AI過度集中於少數大企業手中提出擔憂，科技巨擘對大眾生活的影響力與日俱增。阿爾特曼對此回應說，AI技術目前掌握在少數企業手中，他相信未來會有更多競爭。

鑒於AI有望用更少人工完成更多工作，銀行和另一些金融機構已經在各種功能中使用AI。美國證券交易委員會（SEC）主席根斯勒在聽證會上，對於AI的系統性風險發出警告，認爲下一次金融危機可能出現企業對AI的使用上。他說，數據聚合器和AI平台可能成爲未來金融系統「脆弱性」的主要組成部分。

AI如何影響或擾亂選舉？

更有針對性地進行競選宣傳

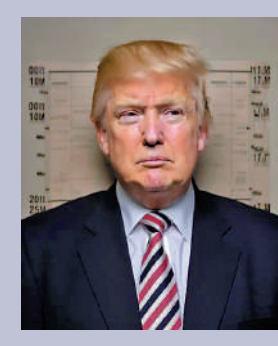
- 生成式人工智能可以更有針對性地迅速產生大量的競選電子郵件、文本或圖片，向選民發放，甚至偽造圖片或者宣傳資料。

通過算法精準鎖定目標選民

- 通過AI協助，可以更加精準地投放政治廣告和進行選情分析，利用數據和算法來自動完成任務，如在社交媒體上精準鎖定目標選民或追蹤捐助者。

僞造音頻視頻抹黑對手

- 通過深偽(deepfake)技術，AI已經可以實現給視頻中的人物「換臉」，或生成某個名人的聲音。相關技術可以用來混淆視聽、誹謗競選對手甚至煽動暴力，例如呼籲選民在錯誤的日期投票；讓政治對手發表憑空捏造的爭議性言論，例如承認犯罪或者表達種族主義觀點；甚至可以生成假新聞，謊稱某人退出選舉。 來源：美聯社



◀▲由AI生成的特朗普疑犯照片在網上流傳。 網絡圖片

▲前美國總統特朗普12日分享了一段由AI處理過並扭曲了的CNN主持人庫珀反應的視頻。 網絡圖片

惡意機器人程式成主要威脅

▲日本東京10日舉辦第七屆人工智能博覽會。 法新社

【大公報訊】據《獨立報》報道：根據最新研究，去年近一半的全球互聯網流量都來自於機器人程式，其佔比比去年增長了5.1%，一些機器人甚至還會模仿人類，逃避安全軟件的檢測。其中，惡意機器人(Bad Bot)程式的網絡流量已佔據整個互聯網網流量的三成，成為網絡面臨的主要威脅。

機器人（程式）流量，指的是任何非人類訪問網站或應用的流量，而惡意機器人技術可用於盜取賬戶數據、入侵電子郵件、暴力破解核發動網絡襲擊等不法活動。有些「高級」的機器人甚至模仿人類行為，以避免被安全軟件檢測到。

網絡安全公司Imperva的數據顯示，2022年自動程序和惡意網絡活動顯著增加，人類流量的比例降至八年來的最低水平。該公司指出，惡意機器人的活動達到自2013年有紀錄以來的最高水平，佔全網所有流量的30.2%。

研究人員稱，由於OpenAI的ChatGPT和谷歌的Bard等生成式人工智能（AI）工具的出現，今年機器人的網絡活動預計會進一步增加。Imperva的高級副總裁特里貝斯（Karl Triebes）說：「自2013年以來，機器人發展迅速，但隨着生成式AI的出現，該技術將在未來10年以更快、更令人擔憂的速度發展。」特里貝斯還說，網絡犯罪分子或許會進一步利用這一技術，因此相關的威脅將會變得更嚴峻。

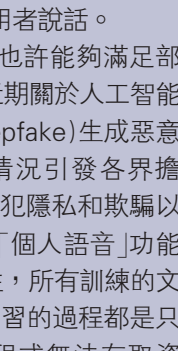
另外，報告還指出機器人的使用在戰爭中有所增加，2022年初針對烏克蘭網絡應用程序的自動攻擊激增了145%，可能旨在破壞該國的關鍵基礎設施。

蘋果AI可模仿用戶聲音 惹隱私爭議

【大公報訊】據CNN報道：美國蘋果公司16日宣布了一系列語音、認知、視覺等相關輔助功能，其中一項新功能名為「個人語音」，利用AI技術，可在15分鐘的訓練後學會並模仿出機主的聲音。但是，該功能也引起了關於對隱私的討論。

該功能名為「個人語音」（Personal Voice），用於幫助失去言語能力的使用者，如罹患肌萎縮性脊髓側索硬化症（ALS）或其他會逐漸喪失說話能力的使用者。使用者可以通過朗讀一系列隨機文本提示，在蘋果手機或平板電腦上錄製15分鐘的音頻，就可以利用該功能以創造出獨一無二的「個人定製語音」。一個名為Live Speech的相關功能可使用「合成語音」在電話通話、FaceTime對話和面對面對話時朗讀用戶鍵入的文本，替使用者說話。

雖然該工具也許能夠滿足部分技術真實的需求，但近期關於人工智能（AI）通過「深偽技術」(deepfake)生成惡意的虛假視頻以及音頻的相關情況引發各界擔憂，蘋果的該項功能也面臨着侵犯隱私和欺騙以及誤導公眾的風險。蘋果表示，「個人語音」功能會保護用戶信息的私密性和安全性，所有訓練的文本都是即時隨機生成，而且機器學習的過程都是只在相關設備裏進行，第三方應用程式無法存取資料，加上蘋果自帶的臉部辨識等保護，減少盜用風險。




▶蘋果將推出可合成用戶聲音的新功能，引發爭議。 網絡圖片

▼OpenAI行政總裁阿爾特曼16日出席美國國會聽證會，探討AI風險及監管。 美聯社

世衛：使用AI監測健康須謹慎

【大公報訊】據路透社報道：世界衛生組織（WHO）警告說，在醫療保健領域中使用人工智能（AI）時，要防止偏見和錯誤信息。

世衛16日呼籲在將AI用於公共醫療方面時要謹慎，稱AI用於決策的數據可能會有偏差或被濫用。世衛還稱，對AI的潛力表示憧憬，但對如何使用AI來更好地獲取改善健康信息、作為決策支持工具以及協助診斷護理表示擔憂。

OpenAI公司旗下的AI聊天機器人ChatGPT去年年底面世以來，AI以更快的速度滲透各個行業以及人們的生活中，然而，AI生成的假資訊讓人難以辨別，引發各界的擔憂。世衛組織在一份聲明中說，用於訓練AI的數據可能有偏差，並產生誤導性或不準確的資訊，而且這些模型可能被濫用以產生虛假資訊。世衛說，「必須」評估使用大型語言模型工具（LLMs）（如ChatGPT）的風險，以保護和促進人類福祉並保護公共健康。

來自英國、美國、澳洲、哥斯達黎加和馬來西亞的衛生專家，上周在《英國醫學期刊》（BMJ）發表聯合署名文章，表示AI的發展可能對人類健康產生負面影響表示擔憂，專家們當時舉例說，使用AI技術的脈搏血氧儀可能會錯誤讀取皮膚較黑患者的血氧水平，導致他們缺氧時難以得到及時治療。

▶一名塞浦路斯高中生使用由ChatGPT驅動的機器人。 路透社