deepfake影片場景逼真 假信息更易攻陷目標

AI造假削民眾感知 拒信真相助長謊言

香港文匯報訊 香港坊間近期出現偽冒特區政府高官誘騙人投資的短片，政府嚴正澄清全屬偽造。伴隨人工智能（AI）技術進步，製作場景逼真卻從未真實發生過的「深度偽造」（deepfake）影片更為簡單，讓人難以辨別。《金融時報》分析稱，深偽技術的發展增加了人們「信以為真」的可能性，反過來人們也會把真相誤認為謊言，正所謂「假作真時真亦假」，人們會更易陷入充斥謊言的網絡世界。


◆AI技術進步下，製作場景逼真的「深偽」影片更讓人難以辨別真假。 網上圖片


◆「深偽」影片的移花接木技術，右圖由「深偽」技術製作。 網上圖片

早在2017年，形容透過謊言鞏固目標群體的偏見、從而掩蓋真相的「後真相」（Post-truth）概念，便被《牛津詞典》評為年度詞彙。自此之後，從烏克蘭總統澤連斯基跳肚皮舞相片、到美國前總統特朗普表示自己有勇氣退出巴黎氣候協定的影片，網絡虛假訊息愈來愈多。AI技術還創造出眾多引人注目、易於理解的深偽影片，在網絡上迅速擴散。

辨識假消息或許並非一項熟能生巧的技能。期刊《實驗心理學雜誌》一項最新研究指出，劍橋大學研發幾款幫助人們辨識假新聞的小遊戲並進行實驗，透過對參與實驗的志願者調查發現，他們玩完遊戲後，更有可能辨認出假新聞。然而與此同時，志願者們也更容易將真實的新聞報道當作假新聞，這意味他們分辨真假的能力並未提升，反而傾向將所有信息都視作假象。

公開撒謊新藉口「錄音是偽造」

報道也指出，AI深偽技術的興起，或會為公開撒謊提供新的藉口。例如2016年美國總統大選前夕，美媒曝光一段特朗普私下頻繁侮辱女性的錄音，特朗普為避免選情受挫，當即就錄音內容致歉。然而如今若發生類似事件，特朗普等政客完全可以辯稱「錄音是AI偽造」，屆時即使有充分理據反駁特朗普的說法，他的支持者也會對他深信不疑。

「洪水戰術」阻斷可信消息源

報道還分析，極易製造的假消息可以批量生產，造假者透過假扮不同消息源，在同一時間向網絡用戶不斷灌輸內容相近的假消息，有機會達到以假亂真的效果。在假消息「洪水戰術」下，社媒上往往充斥攻擊他人、散播謠言、煽動情緒的謊言，以及毫無意義的廢話，網絡用戶更難從中找到可信消息源，也會失去閱讀新聞的興趣。

《金融時報》最後提醒，「如果我們看到足夠多的深偽影片，我們可能也會忽視身邊的謠言和暴力行為。保持懷疑是好事，但過度質疑會讓最簡潔明瞭的事實，也能在人群中掀起爭議。不論如何，人們要明白深偽技術正在快速發展，很多不法分子準備利用它。」

多國爆deepfake音頻干擾選舉

香港文匯報訊 人工智能（AI）不僅製作「深度偽造」（deepfake）影片，音頻也是重災區。《金融時報》報道，在英國、印度和尼日利亞等國，近期都出現試圖用「深偽」音頻影響選舉的事件。專家稱AI音頻軟件眾多，且免費提供多數服務，製作「深偽」音頻非常容易。部分軟件製造商正採取措施，嘗試暗中標記「深偽」音頻供人識別。

報道指出，包括ElevenLabs和Resemble AI等AI軟件，都能平價製作「深偽」音頻。新聞網站評估軟件NewsGuard發現一個TikTok賬號上，有多段用AI生成、模擬美國前總統奧巴馬聲音的「深偽」音頻傳播謠言，相關音頻已播放超過一億次。調查發現其使用的ElevenLabs軟件可以免費使用基本功能，訂閱也只需每月最低1美元（約7.81港元）、最高330美元（約2,578港元）。

多間AI研發企業都設法打擊虛假信息，微軟發布聲明，呼籲用戶舉報任何濫用微軟名下AI音頻工具的行為，若利用AI模擬他人聲音，製作者需得到對方批准。ElevenLabs也製作一款偵測軟件，用於識別音頻是否由其系統製作。Resemble還嘗試在音頻中插入聽不見的特殊標籤，用特定軟件分析音頻波紋即可甄別。

誤信「杜魯多」爲投資平台背書 加國男被騙8.6萬元

香港文匯報訊 人工智能(AI)「深偽」影片經常選用名人照片，讓觀眾稍有不慎便上當受騙。加拿大CTV電視網絡報道，安大略省一名男子看到「加國總理杜魯多」和「億萬富豪馬斯克」支持一個投資平台的影片，並未察覺其中有詐，結果被詐騙1.1萬美元（約8.6萬港元），才發現是「深偽」騙局。


◆杜魯多的影像被不法之徒利用詐騙。 美聯社

受騙男子約翰稱，他看到影片中的「杜魯多」宣稱在該平台投資可以賺錢，自己還以為是投資機遇，便依照影片指示聯繫投資平台。對方鼓勵他從250美元（約1,954港元）開始，其後不斷聲稱他的收益已經翻倍，催促他加碼，「平台聲稱我的收益已經累積到4.6萬美元（約36萬港元），我想取錢時，卻被告知錢被凍結了，要再支付6,000美元（約4.7萬港元）才能取出。」

約翰才反應過來自己上當受騙，「但那段影片，我真的以為它是真的。我快要退休了，拿的是最低工資，我很傷心，根本不知道這是個騙局。」

「岸田」發表性騷擾言論

日本社媒去年11月也流傳一段首相岸田文雄的「深偽」影片，畫面是「岸田」出現在標註有「日本電視台突發新聞」的畫面中，卻在發表性騷擾言論。25歲的影片製作者表示，他用一款免費AI軟件，耗費不到一個小時就製成影片，承認自己只是想開玩笑。日本電視台批評做法不可接受，強調會追究影片製作者責任。

社媒瘋傳Taylor Swift露骨照 歌迷發起保護行動


◆Taylor Swift獲「粉絲」在社媒發布照片力撐。 資料圖片

香港文匯報訊 人工智能（AI）深偽內容層出不窮，美國著名女歌星Taylor Swift也被針對，她的多張虛假AI圖片周三（1月24日）起在X等社媒瘋傳，露骨的內容被大幅轉載。大批粉絲紛紛在社媒發布偶像的照片，希望用真實的照片淹沒假圖。多個社媒已陸續將傳播假圖的賬號封禁，但仍有漏網之魚暗中行動。

英媒《每日郵報》報道，事件源於Taylor Swift公開與美式足球聯盟（NFL）球星Travis Kelce的戀情，有人趁機用AI製圖，生成以球場為背景、Taylor Swift與多名虛構「隊員」合影的露骨照片，在X、Instagram和Reddit等社媒和網絡論壇大肆散播。「粉絲」們趕忙召集發起行動，配合偶像的真實照片，將「保護Taylor Swift」標籤推上熱門。但也有人混入其中，藉標籤繼續傳播假照片。

「保護Taylor Swift」標籤推上熱門

報道指出，X翌日封禁了多個幕後推手的賬號，但部分圖片仍然在平台上流傳，且帖文瀏覽量達到數千甚至數萬，還有始作俑者用代理IP地址掩蓋真實位置。消息人士稱Taylor Swift本人和親友都對假照片事件非常憤怒，會考慮採取法律行動。

美聯社上月的一項獨立研究顯示，全球社媒網絡去年合共發布超過14.3萬張「深偽」影片，數目超過往年總和，顯示「深偽」技術散播假消息問題愈演愈烈。

deepfake偽造哈馬斯殘殺平民


◆「伊斯蘭國」2015年推平民落樓的視頻被謠傳是哈馬斯所為。 網上圖片

香港文匯報訊 網絡專家提醒民眾，即使不利用人工智能（AI）技術，「深度偽造」（deepfake）影片製造者依靠一些「移花接木」、「指鹿為馬」的技巧，依然能散播謠言。如今配合更先進的AI技術，類似影片會更顯以假亂真，需要觀眾謹慎甄別。

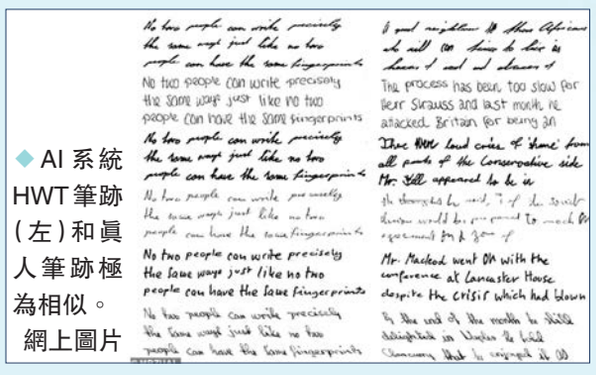
希拉里影片被「移花接木」

指鹿為馬是社媒上常見的傳謠途徑，多數是將一段陳舊且毫不相關的影片，與時下熱點事件相關聯。例如多個社媒上月曾流傳一段影片，聲稱是巴勒斯坦武裝組織哈馬斯的成員，將平民從加沙一幢建築的屋頂推落「處決」。事實上這些影片確有其事，但真正的暴行發生在2015年的伊拉克，施暴者是極端組織「伊斯蘭國」，與加沙和哈馬斯無關。

移花接木則是將毫不相關的影片剪輯拼接，再配以誤導性說明，達到傳謠效果。2016年美國總統大選期間，社媒上一度瘋傳一段荷里活影星狄維莊遜，當面向時任民主黨總統候選人希拉里唱一首辱罵她的歌曲。這段影片實則是用狄維莊遜與另一名搖滾歌手開玩笑的片段，與希拉里露出尷尬表情的片段拼接而成，相關事件並未發生。

關注深偽問題的記者格羅薩斯訪問了上述影片作者，出乎意料的是，作者本意只是惡搞，也沒有添加虛假說明，並非有意傳謠，「他（影片作者）意識到情況令人不安：評論如潮水般湧入。他驚訝『這些人難道以為這是真的？』但事實正是如此。」

模仿筆跡恐增欺詐虛假文件罪案


◆AI系統HWT筆跡（左）和真人筆跡極為相似。 網上圖片

香港文匯報訊 ChatGPT等人工智能（AI）工具可以起草信件甚至提供法律建議，卻只能以電腦文字的形式進行。然而科學家創造了一種可以模仿人類筆跡的AI，這可能預示着有關欺詐和虛假文件的新問題。

事實證明由AI寫出的字跡，與人手所寫的幾乎沒有區別。由阿布扎比扎耶德人工智能大學科學家團隊開發的AI系統HWT，曾接受模仿筆跡訓練，結果成功模仿6名不同人類作家的筆跡樣本。有專家表示，HWT產生的筆跡，看起來比其他現有的AI更真實。

團隊在研究中，向100人展示了HWT和另外兩種手寫生成技術的假文本，並詢問他們更喜歡哪一種，結果顯示81%參與者更喜歡HWT，更重要的是參與者根本無法區分模仿的筆跡和真實的筆跡。

研究團隊表示在訓練中使用了「視覺變壓器」，這是一種專為電腦視覺任務設計的神經網絡，能夠理解影像中物理上彼此遠離的部分是統一連結的。其中一名研究員法赫德表示，「為了模仿某人的筆跡風格，我們需查看整個文本，只有這樣我們才能開始理解這人如何連接字符、如何連接字母或間隔單詞。」