

## 如何透過泰勒絲被「深度偽造」(Deepfake) 詐騙的經歷， 幫助學生審視人工智慧道德規範

駐芝加哥辦事處教育組

當 17 歲的 Alex Bordeaux 正在社交媒體上瀏覽泰勒絲的最新服裝時，畫面上竟然跑出來這位流行音樂偶像令人不安的色情圖片。高中高年級生 Bordeaux 已經在喬治亞州高中 (Dawson County High School in Dawsonville, Ga.) 上過新聞學和校刊課程並學習了人工智慧，因此她立即判斷出自己正在查看的是「深度偽造」——這是指未經本人許可，以不正當的管道並利用人工智慧製造出「完美複製某人聲音或肖像」來操縱此人的視頻、音頻或照片。Bordeaux 感到憤怒和困惑的是，有太多人完全相信社群媒體上的消息或圖片的真實性，她說：「真是太瘋狂了，大多數民眾非常容易在網路上搜尋到這些詐騙視頻並立馬相信這些發佈的真實性且分享，很多和我同齡的孩子對所聽到和看到的並不會做再次確認的動作。」

專家說：「並非所有學生都具備 Bordeaux 的學習背景，他們不知道這種利用最新科技人工智慧所製作出虛擬影像和影片的主要目的是為了影響公眾輿論與欺騙他人錢財——如播放 S 泰勒絲的『深度偽造』，還有一些錯誤訊息的傳播。『深度偽造』對社會大眾所產生的負面作用是極大的。」

Kate Ruane 主任 (民主與科技自由表達計畫—推廣數位權益的非營利組織) 表示「深度偽造」是一個大問題，它們混淆我們的信息世界已經到了相當驚人的程度，目前學校需要優先考慮增加認識人工智慧技術的教學，因為學生們每天都沈浸在網路裡面，人工智慧也正在大規模地發展，這種「深度偽造」傳播假訊息造成無辜者的傷害將會繼續發生。

「深度偽造」的傷害將會繼續發生，因為人工智慧正在大規模地發展，泰勒絲「深度偽造」圖像大都已經被撤下，但它們並不是在吸引了大量瀏覽之前就被撤下了。根據科技新聞網站 The Verge 報導，在 X (Twitter 的前身) 的用戶帳戶因違反 X 政策而被暫停之前，他所發佈的一張圖片一分享就馬上獲得了超過 4 千 5 百萬的瀏覽、2 萬

4 千的轉發和十萬的點讚次數，而該貼文在該平台上已經上熱搜了大約 17 個小時。根據華爾街日報（The Wall Street Journal）報導，該問題嚴重到 X 不得不暫時封鎖所有對泰勒絲的搜尋。

高中生 Bordeaux 知道泰勒絲並不是第一個被「深度偽造」詐騙傷害的名人，前總統 Barack Obama、特斯拉執行長馬斯克和教宗方濟各（Pope Francis）最近也都成為被詐騙傷害的目標。甚至紐澤西一所高中的男學生竟製作並散佈同班女同學的「深度偽造」色情圖片，在總統初選期役間紐漢西流傳拜登總統的「深度偽造」機器錄音通話。Bordeaux 說：「因為人工智慧正在大規模地發展，使用它的人越多，它就越容易被濫用，我認為人工智慧的道德規範非常重要。」

CSforALL（一個推廣電腦科學教育的非營利組織）共同創辦人及執行董事 Leigh Ann DeLyser 表示「深度偽造」有很多顯著的負面影響，需要教師引導學生去思考有關「深度偽造」技術的關鍵問題，和討論開發人員以及政策制定者該如何減輕「深度偽造」所造成負面影響的相關議題，例如「深度偽造」有什麼好處？深度偽造面臨哪些爭議？Leigh Ann DeLyser 提到「深度偽造」確實有存在的爭議，網路社會應該或能夠制定哪些規範？例如直接給「深度偽造」貼上標籤或是在使用某人圖像之前必需先獲得本人的許可？

Bordeaux's 的新聞學老師 Pam Amendola（接受國際科技教育協會關於如何教授人工智慧的培訓）表示她的許多學生感到憤怒，尤其是「Swifties」（泰勒絲粉絲並認為是流行音樂明星的代表的綽號），這些「Swifties」也想到如果發生「深度偽造」的主角不是「擁有類似泰勒絲名聲」的人，他們該如何為自己辯護呢？Amendola 補充說，這個問題可以提供教師一個機會，去提醒學生注意自己的數位足跡，並讓學生了解他們發佈在網路上的信息是有可能會被扭曲或被有心人士利用在犯罪的目的。這次的泰勒絲「深度偽造」詐騙讓 17 歲的 Bordeaux 深刻理解人工智慧技術的發展將會更加複雜，她說：「我非常擔心我的圖像出現在網路上，因為不懷善意的人可以通過多種方式操縱「深度偽造」，讓圖像看起來很可怕。」

### 修改數位相片電腦程式（Photoshop）的類固醇

民主與科技中心（the Center for Democracy and Technology）的

Ruane 說，很多人發現網路上泰勒絲的畫面立刻就知道那不是真實，Ruane 建議教師可以問學生：「詐騙者是怎麼做出『深度偽造』的？關於圖像、和圖像中的人哪些細節讓你知道這是『深度偽造』？什麼直覺讓你的內心做出這是『深度偽造』的結論？」

泰勒絲「深度偽造」詐騙出現時，紐約州公立學校 Newburgh Free Academy 的 Elizabeth Thomas-Capello 電腦課還正跟學生們討論哪些部分是人工智慧尚無法做出的，全班同學正試圖辨識一系列面孔中哪些是人工智慧生成的，哪些是真實的，但最後我們仍然無法成功地辨識出網路上流傳的泰勒絲「深度偽造」是詐騙。

Safinah Arshad Ali（麻省理工學院的研究助理致力於教授人工智慧）將「深度偽造」描述為「修改數位相片電腦程式（Photoshop）的類固醇」，她要求學生「全面性地思考『深度偽造』詐騙的來源？誰是『深度偽造』製作者？發佈這『深度偽造』目的為何？」

Bordeaux's 的新聞學老師 Pam Amendola 也提醒她的學生「科技正在發展到令人難以相信親眼所見的一切，我們必須仔細判斷在網路上所瀏覽內容背後的真實性——無論是流行歌星的照片或是總統候選人的訊息，因為正處於歷史性的階段，我們需要保持一點懷疑的態度，否則就會成為被利用的棋子。」

Thomas-Capello 說「類似泰勒絲的『深度偽造』詐騙肯定會繼續發生——但這事件為未來的電腦科學家提供了一個機會，讓他們能夠更深入研究他們正在學習創造的技術背後的道德規範。」

Thomas-Capello 又補充說「我想強調是 Taylor Swift 的『深度偽造』詐騙是我們社會正在發生的事件，這類事件對我們社會是有相當震撼的影響，身為學生的你們是可以主宰事件的正或負面的影響結果，你們是能夠創造美好科技的人，並且確保發生類似泰勒絲的『深度偽造』詐騙將會越來越困難。」她補充道：「我們真的不知道人工智慧未來走向，我們的學生也才剛開始學習，但如果我們訓練學生利用人工智慧的科技來從事有益社會人群的目的地，那麼人工智慧將是一項非常完美的科技。」

撰稿人/譯稿人：駐芝加哥辦事處教育組

資料來源:2024年1月30日” How Taylor Swift’ s Experience With ‘Deepfakes’  
Can Help Students Examine AI Ethics” , by Alyson Klein ,Assistant Editor,  
Education Week & Lauraine Langreo, Staff Writer, Education Week.  
[https://www.edweek.org/technology/how-taylor-swifts-experience-with-  
deepfakes-can-help-students-examine-ai-ethics/2024/01](https://www.edweek.org/technology/how-taylor-swifts-experience-with-deepfakes-can-help-students-examine-ai-ethics/2024/01)

