AI成欺凌工具 美國高中假裸照氾濫

學校監管措施缺失 受害女生身心受創

隨着人工智能（AI）技術的發展和普及，智能手機上出現大量圖片處理應用程式，通過簡單操作就可以生成難以辨別的「深偽」（deepfake）照片。近期，美國高中校園「深偽」裸照氾濫，不少學生利用AI作為新的欺凌手段，製作女同學的假裸照，並通過聊天群組在校內散布，造成不良影響。專家指出，AI技術如今被濫用於性剝削，校園管理面臨挑戰。

▶美國高中校園「深偽」裸照氾濫，圖爲AI圖片生成器。 法新社

【大公報訊】今年2月，美國加州比佛利Vista中學有學生被揭利用AI生成技術，製作同學的「深偽」色情圖片，並在校內大肆傳播。

去年10月，美國新澤西州韋斯特菲爾德高中發生AI生成假裸照事件，其中包括一名叫弗蘭西斯卡的14歲女生。弗蘭西斯卡向校方報告，稱同班男生利用AI技術合成多名女同學的色情圖片，這些圖片隨後在Snapchat等通訊軟件的群組中傳播。校方初步調查後確認，一名高三男生通過Instagram獲取多名女生照片，使用AI換臉技術將她們頭像植入成人影像，製作出多張「深偽」色情圖片。

據學校和警方報告，美國多個州的男學生利用「裸體」或「脫衣」等AI程式，將女同學參加學校舞會等活動的真實照片，修改合成以假亂真的裸照，並通過Snapchat群組或Instagram等軟件散播，甚至在學校餐廳和校車上散布。

美國聯邦調查局（FBI）警告，傳播AI生成的兒童色情照同樣屬於刑事犯罪。兒童性剝削專家則指出，未經同意使用AI生成圖像來騷擾、羞辱和欺凌女性，會對她們的心理健康、聲譽、人身安全和職業生涯造成損害。

家長不滿學校處理方式

然而，一些學校在校園AI濫用問題上缺乏有效對策，未能充分保護學生。

去年11月，華盛頓州的伊薩夸高中有未成年女生成為假裸照的受害者，學校收到投訴之後沒有採取行動，直到學生家長報警事件才曝光。華盛頓警方一度疑惑，為何學校沒有上報這單可能是涉及未成年性犯罪的事件。

報告顯示，伊薩夸高中副校長在接受調查中表示，不清楚應該報告什麼。該高中隨後聲明稱，根據法律諮詢，無需向警方報告虛假圖像。校方「出於高度謹慎的考慮」，已向兒童保護服務機構報告，並為受害學生提供支持。

新澤西韋斯特菲爾德高中事件發生後，校長郵件告知家長事件「非常嚴重」，已刪除相關圖片以保證不再擴散，但未交代處理細節。此舉引起家長不滿。據受害者透露，校方在調查過程中區別對待男女生，對受害女生造成二次傷害。而且在5個月過去後，弗蘭西斯卡和其他受害女生的家長都表示，校區沒有公開就假圖事件表態，也沒有就使用AI實施監管措施。

▲新澤西州韋斯特菲爾德高中出現AI假圖事件。 美聯社

加州比佛利Vista中學在發現校內傳播的學生AI裸照後，派出校方人員和心理輔導員與學生溝通談心，隨後開除五名涉案學生。校方發布聲明稱，已聯繫警方並前往學校進行調查，同時敦促社區成員與學校分享信息，以確保學生「立即停止對人工智能的不當使用。」

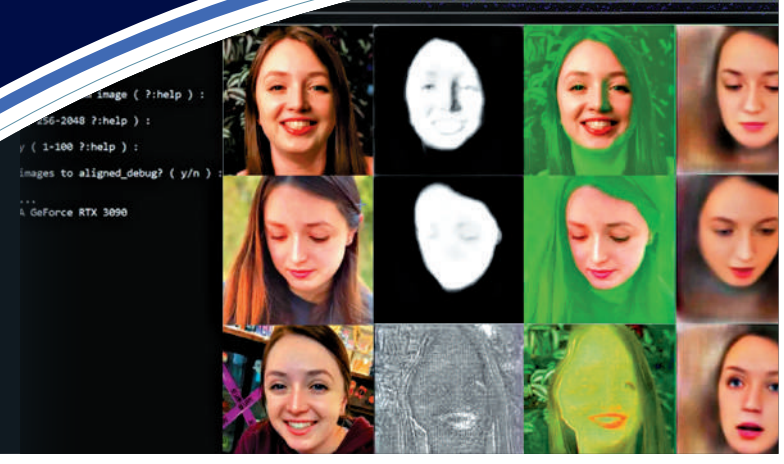
比佛利山聯合學區負責人布雷吉表示，根據加州教育法令，任何製作、傳播或擁有這類AI圖像的學生都將面臨嚴懲，包括開除學籍。學區對此類欺凌行為零容忍，並建議家長與孩子討論社交媒體和人工智能的危險，以及在群聊中傳播此類圖像的後果。

美國多州立法滯後

據報道，美國、加拿大、西班牙、巴西等多國學校都發生過類似的網絡技術欺凌事件。識別AI假圖技術公司Reality Defender的科爾曼表示，過去製造假照片需要在電腦上使用專業軟件操作，現在往往只需一部手機或者某個簡單的網站。儘管OpenAI的Dall-E和Adobe的Firefly等知名的圖片生成服務已設有防止生成色情圖片的門檻，但網絡上仍然充斥着各種「換臉」或「脫衣」程式。Snapchat、TikTok等公司承諾與政府部門合作，防止此類圖片的傳播，但收效甚微。

目前，弗吉尼亞州、加州、明尼蘇達州和紐約州已通過法令，禁止傳播偽造色情作品，並允許受害者控告製作者。新澤西州參議員布蘭尼克表示，正在研究是否增補相關立法，他強調：「在新澤西，這必須被視為一項嚴重的犯罪。」

▲韋斯特菲爾德中學的弗蘭西斯卡（左）成為假裸照的受害者。 網絡圖片



▶AI技術可用於製作「深偽」圖片或視頻。 網絡圖片

多國出現學生AI假裸照事件

美　國

2023年10月，新澤西州韋斯特菲爾德高中有男生利用AI技術將女同學照片加工成色情圖片，通過Snapchat群聊散布。

2023年10月，警方接到家長舉報，發現就讀華盛頓州伊薩夸高中的未成年女兒，成為AI假裸照受害者。校方未主動報警，詢問警方該如何處理，後上報兒童保護機構。

西班牙

2023年9月，阿爾門德拉萊霍市警方接到約20起家長投訴，稱孩子的AI假裸照在社交軟件與聊天群組流傳。調查發現多名13–15歲青少年利用AI生成數十名女童假裸照，並在社交軟件分享。警方擬起訴14歲以上的涉案人。

巴　西

2023年11月，里約熱內盧有學生利用AI技術製造20多名同學的假裸照。校方通知家長並報警，兒童保護警方介入調查。

加拿大

今年4月，安大略省一間中學有學生發現自己的照片被AI合成為裸照在群聊傳播。校方通知家長與警方介入調查。

全球六分之一青少年遭受網絡欺凌

【大公報訊】據BBC報道：今年3月，世界衛生組織發布報告指出，全球近六分之一的青少年曾遭遇網絡欺凌。這一問題自2018年以來日益嚴重。

報告調查了44個國家和地區超過27.9萬名青少年，結果顯示，自2018年以來，遭受網絡欺凌的學齡兒童人數增加。男孩遭受網絡欺凌的比例從12%上升到15%，女孩則從13%上升到16%。雖然總體學校欺凌趨勢保持穩定，但網絡欺凌事件，包括短訊、貼文、電郵，或者未經允許在網上分享視頻或照片等方式卻增加。

世界衛生組織歐洲區負責人克魯格博士對此表示擔憂。他指出，青少年每天上網時間長達6小時，即使微小的網絡欺凌和暴力變化也可能對數萬名年輕人的身心健康產生深遠影響。他呼籲將網絡欺凌視為重要的社會問題，保護孩子們免受線上和線下各種形式的暴力和傷害。

「網絡為學習和聯繫提供了難以置信的機會，同時也帶來了網絡欺凌等挑戰。」學齡兒童健康行為（HBSC）研究每四年開展一次，國際協調員喬安娜—英奇利博士說：「這就要求我們採取綜合策略，保護青少年的心理和情感健康。政府、學校和家庭必須通力合作，共同應對網絡風險，確保青少年在安全和支持性的環境中茁壯成長。」

◀世衛報告指六分之一青少年遭受網絡欺凌，圖為美國高中生。 美聯社

AI造假發展迅速 人類辨別手段有限

【大公報訊】據《衛報》報道：人工智能技術快速發展，造假問題日益嚴重。相較而言，辨別和防範AI造假的手段卻逐漸滯後，公眾識別「深偽」視頻和圖片面臨的挑戰越來越大。

根據Home Security Heroes去年的研究報告，40個常用Deepfake色情網站在前三季度上傳了超過14萬個虛假色情視頻，總量超過了前幾年的總和。早在2017年，Reddit論壇用戶就開始將女性名人的面孔嵌入色情影片中。近年來，AI造假的內容開始在開放的社交媒體平台上傳播，加劇了對公眾的影響。

由於AI技術飛速發展，尤其是Dall-E 3和Midjourney等AI繪圖程序讓造假更加容易，現有的辨別手段已經滯後。人工智能諮詢公司Faculty反虛假信息工作負責人斯皮爾斯說：「從技術角度看，這些程序的發展速度和步伐令人難以置信，也令人震驚。」「從拼寫錯誤的單詞、不協調的平面，或有皺紋的皮膚，我們有各種各樣的人工技術來識別虛假圖像。手是最典型的一種，然後眼睛也很好辨別。但我們時間已經不多了——AI繪圖程序在不斷進化」。

為區分平台上信息的真偽，OpenAI、Meta以及BBC、谷歌、微軟和索尼等公司加入的「內容來源和真實性聯盟」，決定為人工智能生成的內容添加水印和標籤。然而，這些措施並不能達到預期的效果。事實核查組織Logically的政府事務主管帕克指出，標籤和水印雖然有助於觀眾識別真假，但並不能完全消除誤導的可能性。此外，自動檢測器能夠查找和刪除AI造假內容，但其準確率約為70%，存在誤報的風險。

◀AI程式造假發展迅速，人類辨別手段有限。 法新社

日本兩大企業警告：AI失控或導致戰爭

【大公報訊】據美國《華爾街日報》報道：4月8日，日本最大的電訊商日本電信電話（NTT）和最暢銷《讀賣新聞》發表聯合聲明，敦促日本政府盡快出台法規，監管生成式人工智能的應用。聲明警告，如果AI發展不受限制，可能導致社會動盪甚至戰爭。

官方控股逾35.59%的NTT和最暢銷報章《讀賣新聞》在日本具有重要的政治影響力。兩家公司8日均在各自官網上發布了針對AI的宣言，並稱其聯合宣言是出於對公共言論的擔憂。

宣言指出，儘管生成式人工智能在提高生產力方面具有巨大潛力，但總體上對這項技術持謹慎態度，若盲目無限制的使用，恐對人類和社會帶來各種問題，「在最壞的情況下，民主和社會秩序可能崩潰，乃至引發戰爭」，必須在技術上和制度上對生成式AI施行平衡措施，控制其使用。

近年來，各大科技公司推出的AI產品接連引發種族歧視、傳播假信息等爭議，受到社會各界的批評。宣言指出，「AI模型可能教導人們如何製造武器或傳播歧視性思想」，在各國選舉中出現幾乎以假亂真的政界人士演講視頻，造成了極其惡劣的影響。

歐洲議會3月通過《人工智能法案》嚴格禁止「對人類安全造成不可接受風險的人工智能系統」。《華爾街日報》指出，日本兩大巨頭的宣言主旨和《人工智能法案》有諸多脗合，且一定程度上表明，即使美國是盟友，對美國公司領導研發AI計劃的疑慮聲音愈來愈強烈。