

後真相時代下的 網絡安全隱患

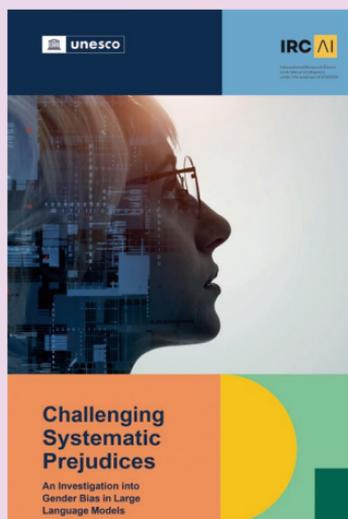
2023 年 11 月，聯合國教科文組織發布有關生成人工智慧技術助長性別暴力問題研究，指出封閉型或開放型人工智能均可生成網上騷擾範本，包括虛假帖文和圖片，繼而進行騷擾、起底、影像暴力等行為。研究亦反映多數應用程式開發目的令女性能夠處於網路安全，同時把責任歸於受害者保護自己免受網絡危害。

去年聯合國教科文組織和國際人工智慧研究中心共同發布《挑戰系統性偏見：大型語言模型中的性別偏見調查》亦指出，大型語言模型，即時下生成式人工智慧平台語言處理工具，如：OpenAI、ChatGPT，存在性別偏見傾向，加劇歧視、恐同和刻板定型。可見，要減低人工智慧的性別偏見，科技公司和平台開發者必須透過技術收集和管理多樣化、具包容的資料庫。

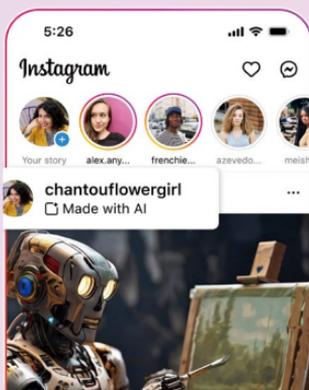
跨界別攜手 加強網上安全意識

國際科技巨頭 Meta 從帳戶選項入手，打擊性勒索和私密圖像濫用行為。未滿 16 歲或特定國家地區未滿 18 歲的青少年帳戶，將預設不公開。用戶現時無法於陌生訊息中傳送影像或語音訊息，直至傳送對象接受聊天要求。

Meta 官方網站亦表示成立內部專責團隊打擊性勒索，處理惡意網絡攻擊，以機器學習大規模偵測並移除這類帳戶，並以「AI 資訊」偵測以 AI 生成或修改而成的內容，用戶亦可以為帖文標註，為其他平台使用者提高資訊透明度。



Meta 由 2024 年 7 月改用「AI 資訊」標註（左），取代「以 AI 製作」（右）



就香港而言，Meta 自 2022 年與風雨蘭合作，推出 StopNCII.org 香港版，進一步幫助識別、調查和回應牽涉影像罪行內容。StopNCII 運作原理為：私密影像轉換成代碼和產生數碼指紋，繼而令參與公司平台，包括社交媒體企業與色情網站，探測和移除分享到其平台上的私密影像。原有影像會留在當事人的裝置中而不會上傳，並不會與 StopNCII 或參與的技術平台共享。



StopNCII.org 由隸屬英國非營利組織 SWGfL 的「復仇式色情求助熱線 (Revenge Porn Helpline, RPH)」所營運。目前成功從網上移除超過 200,000 張未經同意下被散布的私密影像 (Non-Consensual Intimate Images, NCII)，影像下架率高達 9 成以上。

科技平台推動媒體素養教育

2021 年，Meta 與各界攜手合作，於香港推動 We Think Digital 計劃，提供相關教材及推廣活動，培養公民正確使用數碼科技的素養和技巧。自 2019 年起已於全球 14 個國家及地區推行，包括台灣、日本、新加坡、印尼、越南、阿根廷等。超過 200 萬名來自世界各地的年輕人、老師，長者及家庭接受培訓及參與計劃。

慎防數碼足跡 (出自 We Think Digital)

- 注意貼文照片、影片資訊 (例如：登機證、背景、天氣、日期、用戶標註)
- 留意個人 IP 地址、存取 GPS 定位的程式、軟件、插件
- 定期清除瀏覽記錄、搜尋記錄、登入記錄
- 選擇不自動登入、拒絕讓瀏覽器記住登入資訊
- 使用無痕模式，於瀏覽器申請帳號或填寫表格，避免輸入資料被記錄

We Think Digital
Instagram 帳戶



WETHINKDIGITALHK

結語

影像性暴力並非少眾議題，而是與每位市民息息相關的社會問題。影像暴力不分受害者的性別、年齡或性傾向，任何人都有機會成為受害者。

在智能電話普及、社交媒體發達的環境下，影像傳播速度之快，相信大部份香港人，總有身邊的人，不論男女，作為影像性暴力的施害者、受害者或旁觀者。社會對影像性暴力議題容易掉以輕心，我相信一般人目睹實體的性暴力事件發生，都認為需要給以阻止或報警，但許多時候我們面對影像性暴力傳播，可能只是無可奈何，更對影像性暴力的貼上負面標籤，令不少受害人蒙受二次傷害，甚至放棄報警求助。這不僅令受害者無法及時獲得適當支援，更可能縱容施暴者繼續惡行。

因此，預防教育的重要性不言而喻，既要提高市民保護私隱的重要性，提高對不當行為的警覺性，亦要建立支持性的社會環境，阻止這些影像的散播，讓受害者感到被理解和支持，勇於尋求協助。不論是受害者的支援、公眾教育、懲介遺法者，阻止相關影像傳播都需要跨界別的協作，包括社福界、教育界、立法者、執法者、資訊科技界的共同努力。



香港社會服務聯會
總主任 (家庭及社區服務)
黃和平先生