連接變童癖網絡 Ig淪兒童色情溫床

縱容性意味標籤 演算法助長買賣雙方對接

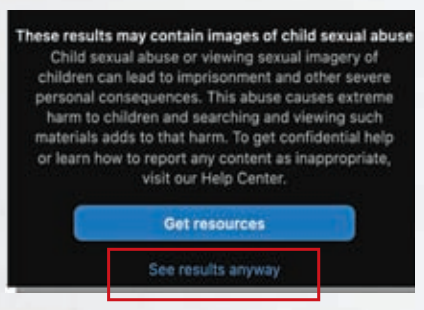
香港文匯報訊 《華爾街日報》7日發布調查顯示，科企Meta名下熱門社媒Instagram（Ig）當中，連接着一個規模龐大的變童癖賬號網絡。這些賬號公開「招兵買馬」，委託製作和購買大量兒童色情內容。Ig不但未有及時阻止違規行為，其演算法反而推動變童癖賬號在特定標籤下活躍，協助買家與賣家對接。報告質疑Meta名下社媒嚴重缺乏監管，讓這些備受青少年喜愛的平台成為兒童色情的溫床。

Ig擁有逾13億用戶，其中不少是青少年和年輕人。《華爾街》今次與史丹福大學和馬薩諸塞大學阿默斯特分校的網絡專家合作，發現Ig上的兒童剝削問題尤其嚴重。與執法部門合作的非牟利組織「國家失蹤與受虐兒童中心」去年合共收到全美3,190萬兒童色情報告，其中多數來自互聯網公司，單是關於Meta名下平台的報告佔比就達85%，來自Ig的報告多達500萬份。

賣方發布「菜單」安排兒童見買家

標籤是Ig的核心功能，但調查發現，變童癖賬號的慣用標籤並未被平台識別封禁，例如將「變童癖」（pedophile）與「妓女」（whore）合併而成的#pedowhore，或是意味未成年性行為的#preteensex等。這些標籤與宣傳出售兒童色情內容的賬號關聯，許多賬戶聲稱由兒童自行管理，使用帶有明顯性意味的措辭。稍微隱匿的#mnsfw（「Minors not suit for work」縮寫，即「未成年人不適合工作」）等標籤，同樣會指向數以千計兒童色情相關帖文。在媒體輿論壓力之下，相關標籤才被下架。

調查發現，出售兒童色情內容的Ig賬號通常不會公開發布內容，改為發布「菜單」，或歡迎買家委託製作特定內容。史丹福大學網絡觀察室研究人員發現，部分菜單會推銷兒童自殘影片，以及「未成年人與動物發生性行為圖像」等。甚至只要價格合適，兒童可以親自與買家「見面」。



◆Ig用戶早前在警告彈窗點擊「不論如何都要查看結果」（紅框示），便能瀏覽違規內容。 網上圖片

彈窗警告「做樣」賬號恐數十萬個

在多數情況下，Ig還允許用戶直接檢索可能涉及違法內容的標籤，用戶會收到一個彈窗警告稱，相關結果可能包含兒童性虐待圖片，提示製作或交易這些素材會對兒童造成「極度傷害」。然而彈窗不會禁止用戶查看，用戶只需點擊「不論如何都要查看結果」，照樣可以瀏覽違規內容。

史丹福大學網絡觀察室合共在Ig找到405個「原創」兒童色情賣家、即聲稱由兒童自行管理的賬號，有賣家聲稱自己只有12歲。調查還發現在Ig上有大量賬號收集支持變童癖的表情貼圖，或引導用戶討論如何接觸兒童。參與Ig兒童安全倡議的的部分現任和前任Meta員工估計，這些變童癖賬號數目恐多達數十萬個。

Meta回應《華爾街》時承認，公司運營平台存在問題，已成立內部工作小組跟進處理。Meta聲稱過去兩年公司已切斷Ig與27個變童癖網絡的鏈接，過濾數以千計兒童色情標籤。平台會限制系統推薦用戶檢索關鍵詞，盡量避免變童癖賬號經系統推薦聯絡互動。

史丹福大學網絡觀察室斯塔默斯2018年辭去Meta首席安全官一職，他指出今次研究訪問權限有限，都能發現龐大的變童癖網絡，「這應該給Meta敲響警鐘，我希望公司重新聘用足夠人手負責審查。」

美雙標「賊喊捉賊」無視自身社媒問題

香港文匯報訊 美西方過去常以TikTok所謂「剝削兒童」為藉口，企圖對TikTok實施種種禁制，但今次調查並無發現TikTok存在與Ig等西方社媒平台相同問題，而美西方卻反而對Ig成變童溫床視而不見，明顯是「賊喊捉賊」和雙標。

Instagram（Ig）的演算法與多數社媒平台一樣，基於行為模式向用戶推薦他們可能感興趣內容，然而若平台缺乏監管，這些演算法也會讓危險群體聚集。史丹福大學網絡觀察室調查發現，Ig依賴以「標籤」為主的演算法，平台監管亦不到位，這讓其相較其他社媒在打擊兒童色情問題上表現最差。

Ig打擊兒童色情成效包尾

研究團隊在社媒上追蹤一個變童癖用戶網絡，發現他們在Ig有405個賬號交易兒童色情材料，在Twitter上有128個，數目為Ig的三分之一。研究團隊隨後將這些賬號上報國家失蹤與受虐兒童中心，由該中心與執法部門合作要求平台跟進。然而一個月後，被舉報的Ig賬號仍有至少31個「賣家」和28個「買家」活躍，至於活躍的Twitter賬號只剩22個。

研究團隊指出，未發現Twitter有推薦過這些違規賬號，刪除賬號的速度也遠勝Ig。至於其他受歡迎社媒，聊天應用程式Snapchat主要用於好友互相交流，很難建立類似網絡。影音分享平台TikTok則沒有發現兒童色情內容有所增加。

史丹福大學網絡觀察室首席技術專家泰爾曾在Ig母公司Meta從事網絡安全工作，他認為Ig的演算法設計不當，「對於那些熱度快速增長的討論議題，你必須監督其主題是安全的，但Ig並未做到。」

舉報無法審核 投訴不獲回覆

香港文匯報訊 社媒平台Instagram（Ig）被揭其演算法助長兒童色情問題蔓延，對於呼籲打擊兒童剝削的用戶卻並不友好。《華爾街日報》提到，有用戶只是偶爾訪問了一個變童癖用戶主頁，自己的賬號就被平台默認與變童癖賬號相關聯。還有用戶抱怨他們屢次投訴變童癖賬號，卻並未得到平台回覆。

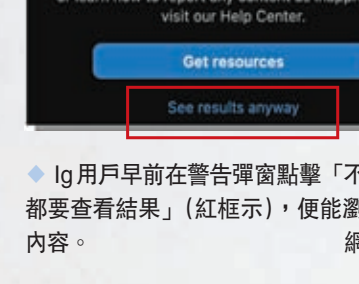
用戶舉報反被認作變童賬號

加拿大用戶亞當斯是兩個孩子的母親，她在Ig上倡議打擊兒童剝削，有不少同為人母的粉絲。亞當斯提到今年2月，有粉絲私訊舉報一個帶有「亂倫幼童」標籤的賬號，她僅點擊進入該賬號主頁數秒就自行退出，並向平台舉報。然而不久後，亞當斯的賬號竟也被平台演算法認作與「亂倫幼童」相關，「我的粉絲慌忙告訴我，當他們查看我的Ig主頁時，平台會向他們推送其他『亂倫幼童』的賬號。」

多名兒童安全倡議人士研究Ig去年收到的數十份投訴，發現有時用戶舉報裸體兒童圖像在平台流傳，卻等待數月沒有得到回覆。一名兒童權益活躍分子稱，他今年初發現一個聲稱銷售兒童色情內容的賬號，其中有帖文配有衣着暴露的少女，以及「這個孩子已經為你準備好了」的文字，帶有明顯的性暗示。然而他舉報該賬號時，卻收到平台自動回覆稱「我們收到的舉報數量過多，團隊無法審核。」

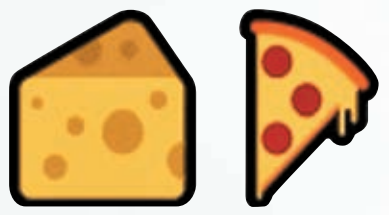
這名兒童權益活躍分子稱，他堅持繼續舉報該賬號，平台卻回應「我們的審核團隊發現該賬號的帖文不違反我們的社區準則」，建議他屏蔽該賬號，就可以不再瀏覽相關內容。

Ig母公司Meta發言人承認，公司確實收到一些舉報後沒有採取行動。發言人解釋，Meta審查了平台的處理方式，發現由於系統故障，多數用戶舉報兒童色情內容都未獲受理，也有部分員工未有正確執行平台審查指引。Meta稱公司會修復相關故障，重新培訓審核團隊成員。



◆倡議打擊兒童剝削的亞當斯，其賬號竟被認作與「亂倫幼童」相關。 網上圖片

變童癖用暗號避審查 利用表情符號諧音替換



◆變童癖使用「芝士薄餅」表情符號暗示身份。 網上圖片

香港文匯報訊 Instagram（Ig）對兒童色情內容打擊不力，讓不少變童癖用戶有機可乘。研究團隊指出，許多用戶會在Ig利用表情符號、諧音替換、倒轉數字等暗號逃避審查。加上Ig只會封禁賬號，用戶只需重新註冊賬號就可捲土重來。

map暗指「喜歡未成年人的人」

馬薩諸塞大學網絡救援實驗室主任萊文指出，用戶逃避Ig審查的方法五花八門。有人會用「地圖」（map）的表情符號，暗指「喜歡未成年人的人」（minor-attracted person），亦有人會使用「芝士薄餅」（cheese pizza），替代首字母相同的「兒童色情」（child pornography）。還有人會備註「喜歡生活中的小事物」，暗示變童癖身份。

萊文提到，部分出售兒童色情素材的Ig用戶用「31歲」配上一個反向箭咀，就可以避開審查、暗示涉及的兒童為13歲。還有賣家會將素材中兒童面部隱去，要求購買完整圖像或影片的買家額外加價。

Ig的演算法有時還會「幫倒忙」，例如Ig在檢索中，屏蔽了一款被用於傳輸兒童色情內容的加密文件傳輸服務名稱，然而平台的自動填充功能居然會建議用戶添加「男孩」等詞彙，再進行關聯檢索。平台甚至會主動推薦一些經常被檢索的「暗號」，戀童癖用戶自然可以「順藤摸瓜」。

加首宗deepfake製兒童色情片案定罪

香港文匯報訊（特約記者 成小智 多倫多報道） 加拿大魁北克省法院判處一名61歲男子使用人工智能（AI）製作最少7個兒童色情合成視頻罪名成立，並因他承認擁有超過54.5萬個涉及兒童遭性侵的圖像或視頻電腦檔案，合共判處他入獄8年。

61歲男子判囚8年

法官加尼翁在裁決書中指出，這是加拿大首宗使用AI深度偽造技術（deepfake）製作合成兒童色情產品的案件，更表明擔心這種兒童色情產品氾濫導致更多兒童成為受害者。

判囚8年的拉魯什被控採用的AI偽造技術屬於「移花接木」類型，他利用AI在社交媒體或其他途徑找到的兒童照片做手腳，將這些兒童的臉部疊加到曾遭性侵的兒童身體上面。加尼翁法官擔心犯罪分子利用這項技術催生一個新兒童色情製品市場，並指犯罪分子可以通過煽動對兒童的性犯罪幻想，把無辜兒童置於危險之中。在裁決書中，加尼翁表示AI偽造技術落在犯罪分子手上令人不寒而慄，足以傷害到每個社區的兒童，因為社交媒體上任何簡單的兒童視頻片段，或一段在公共場所拍攝的兒童視頻都可能被盜用。拉魯什承認曾把其中一些包含兒童遭性虐待的圖像或視頻電腦檔案，提供給其他人。這些圖像中包括一名女童在長達7年內被性侵的一系列照片，事發時受害人處於7至14歲之間。在拉魯什的電腦上，警方找到受害女童的照片，還有她的真實姓名、居住的城鎮和就讀學校名稱。