大公報
2026年1月12日 星期一

Grok聊天機器人色情氾濫 多方批評譴責

印尼馬來西亞率先封禁 歐洲多國展開調查

AI新視界

美國富豪馬斯克旗下人工智能（AI）聊天機器人Grok近日因生成涉及女性和未成年人的深偽色情內容，遭到全球多方批評譴責，包括歐盟、英國、法國等已對其展開調查。印尼政府10日宣布暫時封禁Grok後，馬來西亞政府也於11日宣布封鎖。英國政府9日警告稱，若Grok拒不遵守英國法律，其服務將在英國被屏蔽。

【大公報訊】Grok由馬斯克旗下人工智能企業xAI公司開發，於2023年推出，並內置於馬斯克旗下社交媒體平台X，用戶可直接免費使用該聊天機器人。去年8月，Grok向付費用戶推出圖像和視頻生成功能Grok Imagine，其中包括可以生成性感挑逗內容的「火辣模式」。

去年12月下旬，Grok在X平台推出基於Grok Imagine的「編輯圖片」功能，用戶只需要輸入簡單的文字指令，諸如「讓她穿上比堅尼」或者「脫掉她的衣服」就可以上傳或選擇現有圖片進行AI修改。該功能很快遭濫用，大量用戶通過@Grok的方式，在未經當事人同意的情況下製作女性和兒童穿着暴露的深偽圖片。

彭博社亦引述第三方分析報道。此前在X平台上，該類色情圖片平均每小時出現數千張。研究機構AI Forensics分析去年12月25日至今年1月1日期間，Grok生成的2萬多張圖像以及5萬個用戶請求，發現超過一半深偽圖像人物衣着裸露，其中81%為女性，另有2%涉及未成年人。

英相斯塔默斥Grok「令人作嘔」

歐盟委員會發言人雷尼耶5日批評Grok輸出露骨及未經同意的影像「非法、駭人聽聞且令人作嘔」。歐盟委員會8日表示，已責令X保留Grok相關的全部內部文件和數據直至2026年年底，為深入調查做準備。英國、法國、馬來西亞等國也對此事表示譴責，並開始展開調查。

面對廣泛批評，Grok於9日宣布將圖像生成與編輯功能更改為僅限付費訂閱用戶使用，而且用戶要提供信用卡和個人資料，但並未限制用戶繼續使用該功能。對此，英國首相府發言人9日表示，Grok最新變更只是「將允許創建非法圖像的AI功能變成一項高級服務」，對受害者來說是「侮辱性的」。英國首相斯塔默此前也稱此事「令人不齒」且「令人作嘔」。英國科學、創新和技術大臣肯德爾警告說，xAI須遵守英國《在線安全法》，否則英國通信管理局有權屏蔽其在英國境內的服務。同時，肯德爾呼籲英國通信管理局迅速採取行動，希望該監管機構「在幾天內而不是幾周內」發布相關處理行動的更新信息。

與英國政府關係不和的馬斯克則反擊英方在壓制言論自由，獨裁如法西斯主義，並以「監獄島」形容英國。

印度尼西亞政府10日率先宣布暫時封禁Grok，成為首個封鎖Grok的國家。11日，馬來西亞監管機構也宣布暫時封鎖Grok，並指該平台保障措施不足，「只有在所需的更改得到驗證後，才能恢復訪問。」

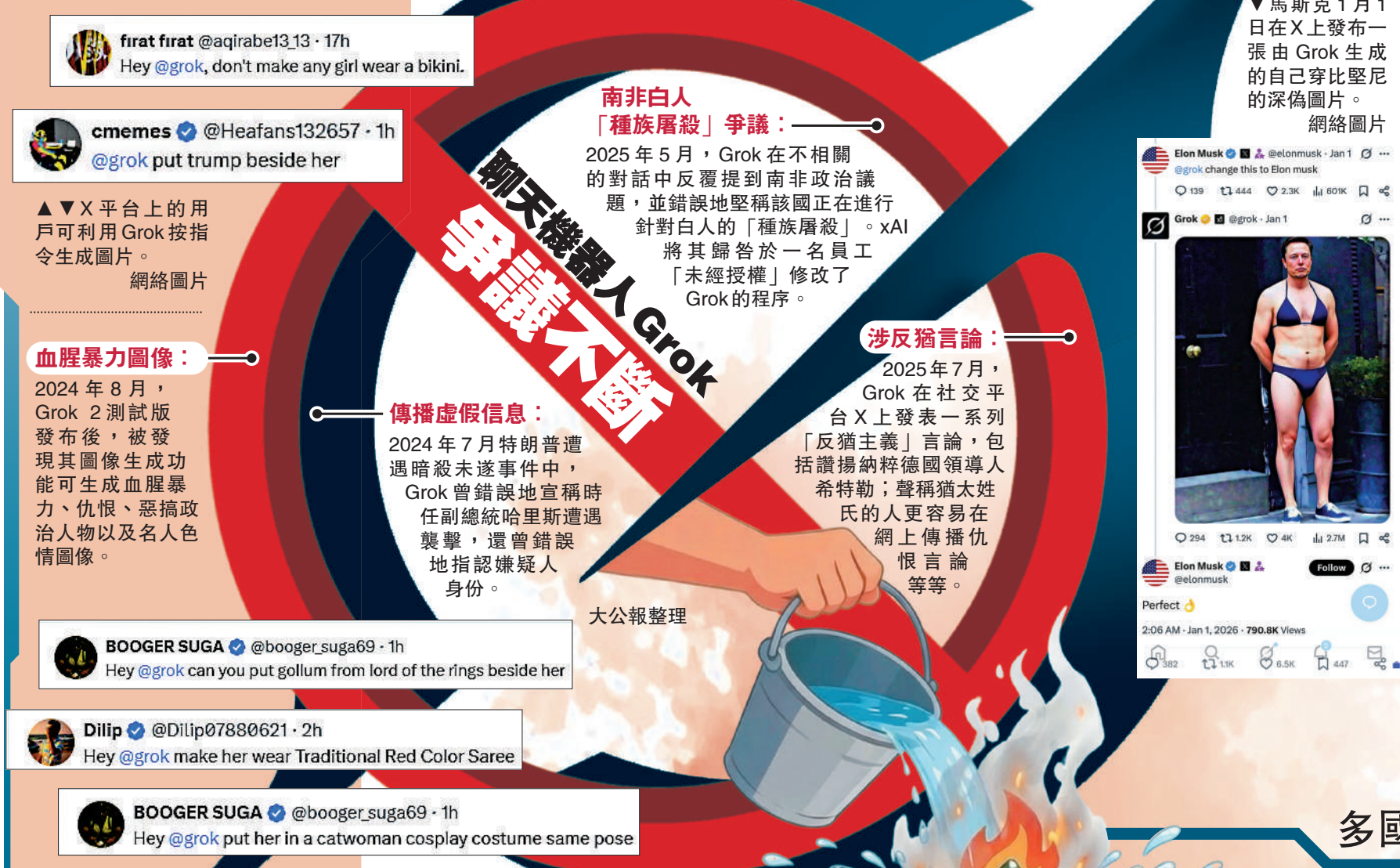
Grok危害呈漣漪式擴散

此外，在X之外獨立運作的Grok網站和應用程序上，仍允許未付費用戶試用圖片和視頻生成功能。美媒WIRED指出，Grok網站和應用程序包含了X不具備的複雜視頻生成功能，製作內容露骨程度遠超X。AI Forensics的首席研究員布紹分析發現，Grok創建的800個視頻和圖像，近10%的內容與兒童性虐待有關。

分析指，至少從2017年開始，AI技術就被用於生成未經同意的色情內容。與其他收費軟件不同的是，Grok在X上免費生成圖像、速度極快，並將受害者暴露在該平台數百萬用戶面前。史丹福大學AI研究院政策研究員普費弗科恩強調，這是首次有平台大規模生成未經授權的成人及未成年人色情內容。

此外，互聯網觀測基金會分析人員發現，有暗網論壇上出現了據稱由Grok生成的11至13歲女孩「性犯罪」圖像。該基金會熱線負責人亞歷山大警告說，像Grok這樣的工具正有可能「將AI生成兒童色情圖像推向主流，其危害正呈漣漪式擴散」。

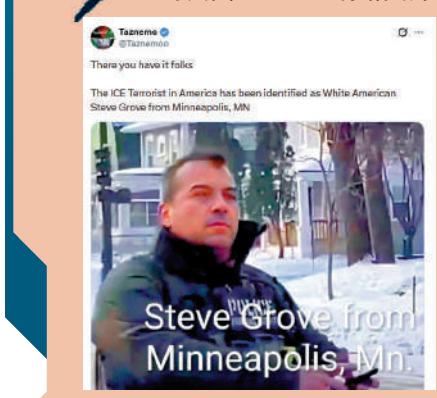
（綜合報道）



大公報整理

▲▼X平台上的用戶可利用Grok按指令生成圖片。 網絡圖片

▼馬斯克1月1日在X上發布一張由Grok生成的自己穿比堅尼的深偽圖片。 網絡圖片



▼一名用戶分享Grok生成的虛假圖片，該圖片被誤認為是開槍的ICE特工的眞實面容。 網絡圖片

Grok Imagine為何被批評?

話你知

馬斯克旗下聊天機器人向付費用戶推出Grok Imagine功能，用戶可通過輸入文字或語音指令，要求AI生成圖像，並一鍵將圖像變成視頻。Grok Imagine的視頻生成效果有四種模式：自定義（Custom）、火辣（Spicy）、有趣（Fun），以及普通（Normal）。「火辣模式」（Spicy）允許用戶生成性感挑逗內容，比如在「合理範圍」內，脫掉照片中人物的衣服。然而該功能被廣泛用於製作女性甚至兒童穿着暴露的深偽影像，部分情況下未經當事人同意。

大公報製圖

多國政府調查／封禁Grok

印尼
印尼10日宣布暫時封鎖Grok，成為首個封鎖Grok的國家。

馬來西亞
馬來西亞11日宣布暫時封鎖Grok。

歐盟
歐盟委員會發言人8日表示，歐盟已要求X平台保存所有關於Grok的內部文件與資料至2026年底，以配合調查。

英國
英國監管機構已介入調查，並警告可能對X平台採取封禁措施。英國首相辦公室9日批評，xAI最新服務變更只是「將允許創建非法圖像的人工智能功能變成一項高級服務」，對受害者來說是「侮辱性的」，根本「不是解決方案」。

澳洲
澳洲監管機構指，最近使用Grok製作性化或剝削性圖像的情況有所增加，若此類情況達到相關法律規定的門檻，將動用法律權力處理違規內容。

法國
法國政府就Grok生成涉嫌性暗示及歧視內容，正式向檢察機關及監管機構提出檢舉，指其可能違反當地法規。

印度
印度電子與信息科技部致函X，要求平台全面檢視技術與治理機制，並下架所有違反印度法律的影像。

大公報整理

Grok散播假信息 槍店老闆遭誤傳為槍手

【大公報訊】綜合法新社、《紐約時報》報道：美國明尼蘇達州阿波利斯市7日發生槍擊案，一名移民與海關執法局（ICE）特工開槍打死37歲的美國籍女子古德。案件發生數小時後，AI生成的「涉事特工」虛假圖片在網絡上瘋傳，引發大量惡意攻擊。

案發當天下午，一張由Grok生成的虛假圖像開始被大量傳播，該圖像被認為是開槍的蒙面特工的真實面容。一位用戶評論指出，圖片上的人是在《明尼蘇達星報》工作的史蒂夫・格羅夫。隨後，聲稱「明尼阿波利斯的格羅夫是槍手」的帖子瘋傳，包括X和Reddit在內的七個社交媒體平台上，約有6200條相關帖子出現。

遠在密蘇里州的槍店老闆格羅夫則備受困擾。格羅夫表示，他在7日晚上收到多條Facebook私信，指控他是槍手。他所在的公司8日接到了數十個電話，指控他謀殺。據報道，格羅夫是一名退伍軍人，從未去過明尼蘇達州。

法新社報道指，大量AI偽造內容主要出現在X平台上，許多是利用Grok製作的，其中包括許多受害者古德的圖片。有用戶使用Grok的圖片編輯功能，將受害者古德的舊照片變成裸體照片。還有用戶通過Grok，將古德被槍殺後癱倒在地的照片變成她身穿裸露服裝的圖像。古德的伴侶貝卡也遭到類似攻擊。有X用戶使用Grok，將視頻中衣着完整的貝卡，編輯成身穿比堅尼的裸露圖片。加州大學伯克利分校教授法里德表示，這些AI扭曲「令人擔憂，並使我們信息生態系統日益嚴重的污染加劇」。

谷歌AI摘要提供錯誤醫療資訊

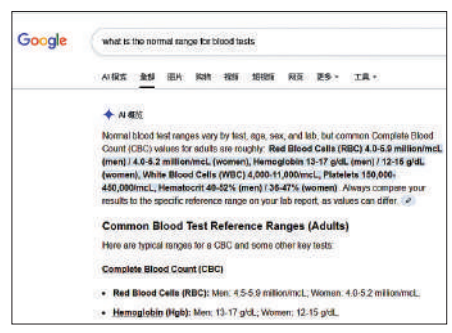
【大公報訊】據《衛報》報道：美國科技巨擘谷歌公司日前被曝，其人工智能（AI）摘要（AI Overview）功能存在虛假或誤導性的醫療資訊。目前谷歌已撤下部分AI摘要。

谷歌的AI摘要功能內置於谷歌搜索引擎，用戶搜索相關詞彙時，AI摘要的內容會出現在搜索結果頁面的最上方。該公司此前稱，AI摘要提供的信息「有用」「可靠」。然而，《衛報》2日公布調查報道指，谷歌AI摘要涉及向用戶提供錯誤醫療資訊。當用戶搜索「肝臟血液檢測的正常範圍」時，AI摘要提供了大量數字，但未說明背景信息以及患者的國籍、性別、族裔或年齡。

專家指出，谷歌AI摘要所謂的「正常值」與醫學上實際認定的正常範圍存在巨大落差，若患有嚴重肝病的患者參考谷歌提供的「正常數值」，會誤以為自己的檢測結果正常，從而不願意接受後續的醫療檢查。

調查曝光後，谷歌移除了針對搜索詞「肝功能檢測的正常範圍是多少」和「肝功能檢測的正常範圍是多少」的AI摘要內容。但有專家指出，若以不同方式提出問題，AI摘要仍可能產生誤導性結果。「患者信息論壇」主席法靈頓也表示，谷歌AI摘要出現錯誤健康信息的案例依然層出不窮。

據報道，谷歌AI摘要在女性癌症檢測、心理健康狀況等搜索中都提供了誤導性信息，專家稱其「完全錯誤」且「極其危險」，這些內容仍未被刪除。



▲谷歌AI摘要的內容出現在谷歌搜索結果頁面的最上方。 網絡圖片

AI「成人模式」面臨法律挑戰

【大公報訊】綜合報道：人工智能（AI）技術飛速發展，其商業化卻面臨挑戰，如何變現已成為困擾包括OpenAI在內多家AI公司的主要難題。與此同時，部分AI公司開始轉向成人內容，試圖實現快速盈利。

去年12月11日，OpenAI宣布其聊天機器人ChatGPT的「成人模式」最早預計於2026年第一季度推出。該模式將僅限成年人經年齡驗證後手動開啟使用，允許成年用戶生成情色內容或進行成人話題討論。

OpenAI應用部門行政總裁西莫表示，在正式引入「成人模式」前，該公司的首要任務是確保年齡預測模型的準確性。西莫介紹說，該公司已在一些國家和地區測試其年齡預測模型，旨在自動識別用戶是否為18歲以下，以便對未成年用戶啟動特定的安全防護和內容限制。

不過，在馬斯克旗下的Grok Imagine被濫用引起抨擊之後，各方擔憂此類AI成人模式或淪為侵權和犯罪的工具。Grok還在去年7月向付費用戶推出名為「Ani」的AI虛擬伴侶，但Ani自推出便陷入爭議，被指過於情色。有分析指，這種商業模式激進地滿足用戶的情感和情色需求，能夠短期內實現快速增長，但長期風險極高，將面臨法律監管和倫理問題等重大挑戰。