



## AI智慧協助巡網 宣戰兒少性剝削

### AI智慧協助巡網 宣戰兒少性剝削

專訪衛福部資訊處處長李建璋

撰稿：中國時報記者 林周義

藝人黃子佼持有未成年人性影像，今年4月遭北檢依違反《兒童及少年性剝削防制條例》起訴，立院也著手修法，授權主管機關運用科技主動巡查。由於網路性犯罪往往「道高一尺，魔高一丈」，被抄後又會重啟爐灶，犯罪速度極快，衛福部保護司和資訊處近期研議推動網路科技巡查，揪出有性剝削疑慮的影像。另計畫也規劃訓練語言模型，一旦在網路上發現「喝茶」、「茶資」、「茶溫」等常見性剝削術語，衛福部、刑警就能及時介入。

黃子佼繼捲入性騷風波後，又遭媒體爆料是非法網站的高級會員，持有7部兒少性影像，遭北檢依違反《兒童及少年性剝削防制條例》起訴。事件曝光後，多名藝人紛紛表態拒絕合作，衛福部也開出公部門第一槍，不僅不再與黃子佼合作，也下架與黃子佼相關的影片。風波越演越烈之際，原被政府封網的非法網站，又遭爆料重啟爐灶，甚至以黃子佼為噱頭，繼續販售性影像，引發各界譁然。



兒少性影像過去採取的是「被動式檢舉」，當民眾發現性剝削的文字、圖像，經檢舉後，政府才介入處理。為加強兒少保護，立法院5月針對《兒童及少年性剝削防制條例》修正案進行討論，重點包括加重刑罰、必要時可強制封網、建立主動巡查機制等，並於7月三讀通過。

面對防不勝防的性影像，上述條例修正案授權中央主管機關得運用科技設備方式，主動巡查網站有無涉及兒少性剝削，且網站業者不得規避，違者可處新台幣6萬~60萬元罰鍰。

因應修法新增內容，為加強網路安全的把關，衛福部資訊處李建璋處長表示，衛福部將規劃推動科技巡查，透過「爬蟲技術」辨識，一旦發現兒少性剝削內容，就會主動處理，必要時請警方介入，其中的關鍵在於讓人工智慧（AI）辨識出兒少性剝削內容。

在影像的辨識上，傳統的作法是提供大量兒少性剝削影像，訓練模型辨認，但李建璋認為此舉並不洽當，原因在於這會造成兒少的二次傷害。正因如此，大部分的國家都難以建立出模型。由於Google曾公開發表過影像辨識工具，衛福部正研議與Google合作以AI方式揪出兒少性影像。

「當你打開Google的影像搜尋，會搜不到兒少性剝削畫面，原因是Google可以自動辨認、過濾」。李建璋解釋，Google的模型並非以大量兒少性影像來訓練，而是讓電腦學會辨識正常的兒少影像，採取「消去法」，當影像脫離正常的範疇，就會認定為疑似性剝削內容。

李建璋指出，兒少性剝削並非只有本地網站有，還有很多國外的影像，在世界各地破獲後，就會產生識別碼。衛福部可透過國際組織的參與取得識別碼，但得仰賴Google才能辨識，以免性剝削內容流入台灣。目前衛福部已與Google台灣總部人員進行初步討論，Google認為，衛福部的想法與Google的文化、價值觀吻合，給了初步正面回應。

除了影像，網路上亦存在許多台灣專有、用於形容性剝削的術語。李建璋舉例，在台灣，性剝削被稱為「喝茶」，而性剝削費用被稱為「茶資」，性剝削對象的品質則被稱為「茶溫」等，這是本地獨有，而且會隨時代演變的用詞。有鑑於此，衛福部也規劃與民間資訊產業合作，訓練在地化的語言模型，希望讓AI辨識出這些文字內容，以利公部門採取行動。

「我們採取行動了，不會被動地等待」。李建璋表示，目前相關架構、方案都已在進行中，衛福部期盼透過主動巡查機制，讓兒少性剝削影像在台灣的網路環境中「不至於泛濫」。若能成功合作，揪出不法網站、俱樂部，這將成為領先於亞洲、歐美國家公部門的創舉。

李建璋也強調，衛福部能處理的範圍是真人影像，但隨著科技進步，未來性影像的處理上，挑戰也會越來越大。現行的人工智慧都是靠著過去的歷史資料來訓練，當生成式AI普及化，除了真人影像，卡通也可能成為兒少性剝削的溫床。他提醒民眾，必須對性剝削有警覺心，一旦搜到私密網站、受到私密網站的邀約，就應主動向衛福部通報，透過民間、政府部門的協力，才能杜絕性剝削。