臺灣警察專科學校警專學報，民國 114.6
第八卷第五期：59-82 頁

生成式人工智慧有關兒少性剝削製品問題

Child Sexual Exploitation Material Created by Generative AI: Issues and Challenges

韋愛梅[1]

Ai-Mei Wei

摘要

生成式人工智慧的發展為數位內容創作帶來了重大變革，特別是影像與音訊的生成技術。然而隨著生成式人工智慧技術的成熟，社會亦面臨新的倫理與法律挑戰，人工智慧生成的兒少性剝削製品問題即為其一。近年來不斷出現人工智慧生成的兒少性剝削製品，包括利用深偽技術及生成模型生成的高度擬真的性圖像或影像等，嚴重侵害兒少的身心健康及隱私權。根據美國國家失蹤與受虐兒少援助中心及多項國際研究顯示，人工智慧生成技術不僅大幅增加兒少性剝削圖像的生產數量，更增加警察執法部門辨識和追蹤被害者及加害者的難度。本研究旨在分析生成式人工智慧技術被濫用的現況及相關問題，並進一步探討國際社會於法制層面與實務層面的回應措施與因應策略，期能作為我國未來法制發展與政策規劃之參考。

關鍵字：生成式人工智慧、兒少性剝削製品、深偽、網路安全

Abstract

The rapid advancement of generative AI technology has significantly transformed digital content creation, particularly in image and audio generation. However, alongside technological maturation, society is increasingly confronted with emergent ethical and legal challenges, notably the proliferation of generative AI-produced child sexual exploitation material. Recent years have witnessed a surge in highly realistic AI-generated child sexual exploitation material, facilitated by deepfake technology and generative models, resulting in substantial harm to minors' physical and psychological well-being and a serious violation of their privacy rights. According to reports by the

[1] 臺灣警察專科學校行政警察科副教授。衷心感謝二位匿名審查委員的寶貴意見。

National Center for Missing & Exploited Children (NCMEC) and various international studies, generative AI technologies have significantly escalated the volume of child sexual exploitation material production, complicating law enforcement agencies' efforts to identify and track victims and offenders. This paper examines the current risks and challenges posed by the misuse of generative AI technology, explores international legal frameworks and practical response strategies addressing this issue, and provides recommendations for informing future legal developments and policy planning in the national context.

Key words: Generative Artificial Intelligence, Child Sexual Exploitation Material, Deepfake, Cybersecurity

壹、前言

生成式人工智慧（Generative Artificial Intelligence, GAI）的發展為創意產業和數位內容創作帶來革命性的變革，該技術不僅能夠生成高質量的影像、音頻和影片，還能模擬出逼真的人物和場景，應用範圍從電影製作、廣告設計到娛樂內容的自動生成。然而，隨著該技術的發展成熟，也出現了觸及道德、法律和社會的問題，人工智慧生成的兒少性剝削製品即為其一。根據美國國家失蹤與受虐兒少援助中心（National Center for Missing & Exploited Children, NCMEC）2024 年公布的年度報告指出，網路兒少性剝削案件近期呈現遽增趨勢以及也開始出現人工智慧生成的兒少性剝削圖像與影音（NCMEC, 2024）。NCMEC 從 2023 年開始接受關於人工智慧生成的兒少性虐待或性剝削製品的檢舉案，當年即收到近 5,000 條的檢舉，內容多為人工智慧根據真實兒少的相片所製成，加害者甚至會利用人工智慧生成的兒少性剝削內容對受害兒少及其家庭勒索（張瑞邦，2024；The Guardian, 2024）。為探討此日益嚴重的議題，本文先從生成式人工智慧技術濫用創造出兒少性剝削製品問題及現況進行探討，並分析人工智慧生成兒少性剝削製品可能帶來的困境與挑戰，及以他山之石瞭解聯合國、歐盟等國家現行相關法制及防制作為，最後提出可供改善我國問題的相關建議。

貳、現況與問題探討

一、生成式人工智慧的風險與濫用

2022 年 11 月美國 OpenAI 實驗室開發推出人工智慧聊天機器人程式 ChatGPT（Chat Generative Pre-trained Transformer），藉由文字問答，回答人類的提問或指令（楊舒凱，2023）。2023 年被視為生成式人工智慧元年，生成式人工智慧在 OpenAI 開發的 ChatGPT 引領下，掀起全球熱潮，引起各界廣泛的討論，被視為是人工智慧之一項重大突破（艾貝斯網路 iBEST，2023）；接續 2023 年 3 月 14 日發布 ChatGPT-4 生成式預訓練變換模型（Generative Pre-trained Transformer 4, GPT-4），功能更為強大，具備更強大的生成能力、可讀取圖片和理解、多語言支持、高度精準的翻譯功能、深度學習能力、語境理解及推理能力等功能，不僅為人工智慧領域帶來更多創新和變革，也為人類生活和工作帶來更多便利（張麗卿，2024；Rae Yu，2023）。生成式人工智慧是一種顛覆性技術，有可能極大地改善我們的生活，但也有可能被濫用和造成全球危害（陳婉潔，2024；顧振豪，2023）。

生成式人工智慧是一種電腦程式，旨在創建類似於人類製作（human-made）之新內容；其大量蒐集、學習與產出之資料，可能涉及智慧財產權、人權或業務機密之侵害，且其生成結果，因受限於所學習資料之品質與數量，有可能真偽難辨或創造不存在之資訊（行政院，未註明），也有可能會生成不符合社會價值或公序良俗的暴力、色情等內容而引發爭議（靜宜大學教育發展中心，未註明）。有人說生成式人工智慧將打開潘朵拉的訴訟盒子，如果機制無法面對餵給模型的資料從何而來、模型產出對原始資料持有者之間，將有何連結又會有什麼回饋等問題，這類型的爭議將持續發生（張東文，2022；陳冠瑋，2023）。

二、深偽影像與線上兒少性剝削危機

生成式人工智慧有許多積極的用途，但生成式人工智慧的技術和應用，也助長了兒少性剝削問題的發生並使之更加嚴重。美國司法部剝削及猥褻兒少司司長葛洛奇（Steve Grocki）表示，透過使用人工智慧而創造出性露骨的兒少影像，是格外邪惡的網路剝削。英國愛丁堡大學全球兒少安全研究機構 Childlight 推估，2023 年全球有 12.6%、超過 3 億的兒少成為文字或影像的網路受害者，網路傷害的態樣包括不受歡迎的性對話，例如未經同意的色情簡訊、不受歡迎的性問題或是性邀約；犯罪掠奪者還會以性影像保密為由以及以濫用人工智慧深偽技術產生擬真的

性圖像，進行性勒索或金錢勒索（Edinburgh University, 2024）。2024 年 1 月音樂天后泰勒絲（Taylor Swift），成為深偽色情的受害者，她的深偽裸照在社群媒體瘋傳，雖然不實影像在 17 小時後下架，但已超過 4,700 萬人次的閱覽（劉淑琴，2024）。泰勒絲及許多知名人物的被害，更凸顯此生成式人工智慧新興科技的潛在風險。

英國網路觀察基金會（Internet Watch Foundation, IWF）2023 年 10 月的人工智慧生成兒少性虐待報告指出，人工智慧技術被用於製作兒少性虐待製品（Child Sexual Abuse Material, CSAM）問題日益嚴重。一個月內暗網論壇上有 20,254 張人工智慧生成的圖像，其中 3,000 多張是描繪兒少遭受性虐待的圖像（IWF, 2024）。生成式人工智慧合成的內容和未經同意散布的真實圖像，尤其針對女性、女孩和 LGBTQI+族群的犯罪傷害事件，近年來在全球激增（Romero-Moreno, 2024）。在網路已為生活日常的同時，兒少色情影像議題也因網路的作用下為全球帶來空前的危機，堪稱「兒少性剝削大流行」（張子午，2024）。

美國民主與技術中心（Center for Democracy and Technology）在 2024 年 9 月 26 日發布的一篇報告指出，美國高中正在成為色情深度偽造的污水池，未經同意的私密圖像在學校中令人震驚地普遍存在。該中心的民意調查發現，在上一學年有 15%的高中生自陳聽過深度偽造或人工智慧生成的圖像，且該圖像以露骨性或親密的方式描繪學校中的學生。生成式人工智慧工具加大了學生成為被害者和加害者的範圍（Laird et al., 2024）。

參、人工智慧生成的兒少性剝削製品類型及影響

一、人工智慧生成的兒少性剝削製品類型

依據聯合國「關於買賣兒童、兒童賣淫和兒童色情問題之任擇議定書」第 2 條，「兒少性剝削製品」係指以任何方式呈現兒少進行真實或技術合成之露骨性活動或主要為性目的而裸露與兒少性相關之部位；另依據我國兒童及少年性剝削防制條例第 2 條第 1 項第 3 款，「兒少性剝削製品」包括：兒少性影像、與性相關而客觀上足以引起性慾或羞恥之圖畫、語音或其他物品；至於「性影像」的概念，則是依據刑法第 10 條第 8 項之定義，包含性交行為、性器接觸、客觀上足以引起性慾或羞恥的身體隱私處，或以身體或器物接觸性器或身體隱私處或是其他與性相關在客觀上足以引起性慾或羞恥行為的影像或電磁紀錄；與兒少相關的性影

像，呈現的形式可能有靜態影像，例如照片、圖畫；也可能有動態影像，例如影片。

綜上概念可知，兒少性剝削製品涵蓋與兒少性有關的文字、圖畫、聲音及影像等內容，而聯合國之兒少性剝削製品，主要指兒少之性影（圖）像，且不論是真實兒少或是技術合成的虛擬兒少，皆包含在兒少性剝削製品的範圍之內。

而由人工智慧生成的兒少性剝削製品，指的是利用人工智慧技術生成、變造或操縱涉及兒少性剝削的內容，包括圖像、影片、聲音、文字及動畫等。透過人工智慧生成的模擬兒少遭受性剝削的照片、聲音或影像，也被稱為虛構的兒少色情製品、偽色情製品或幻想圖像（Krishna et al.,2024），這些製品中的兒少雖「非真實存在」，但由於其高度逼真的特徵，可能導致與真實兒少性剝削內容類似的社會危害；另外以人工智慧技術合成例如深偽技術所「虛構」出的兒少性剝削製品，卻是「真實存在」的兒少且遭受實際傷害。

本文聚焦在兒少影像性剝削製品的探討，將人工智慧生成之內容分類如下。

（一） 以文本或圖像生成的虛擬兒少性剝削製品

生成式人工智慧可以通過學習一個大量的圖像數據集，透過生成式對抗網路[2]（Generative Adversarial Network, GAN）等技術，生成與原始圖像相似的新圖像。例如 Midjourney 繪圖工具，可以文本生成圖片或是以圖生圖，亦可修復圖像、風格轉換、圖像合成等。在人工智慧生成圖像部分，目前市面上已開發的網站或繪圖軟體多達數十種，例如 MyEdit、Midjourney、Unidream、DALL· E3、Stable Diffusion、NovelAI…等，人工智慧繪圖除可產出各種超現實的奇幻想像外，也可以生成十分擬真如真人般的圖像。人工智慧繪圖具有文生圖、圖生圖等功能，也多可以直接在行動裝置上操作，只要輸入中文或英文的文字描述並挑選喜好的風格，像是卡通、3D 等，短短幾秒就可以快速生成人工智慧的繪圖作品（Chen，2024）。

當使用者利用文本描述與「兒少」及「性相關」的文字時，即可由電腦生成虛擬兒少性剝削圖像；或是使用者利用上傳的圖像，透過生成式人工智慧平台內建的模板，修改生成新的兒少性剝削圖像。史丹佛數位儲存庫（Stanford Digital

[2] 生成對抗網路是一種非監督式學習，通過兩個神經網路相互博弈的方式進行學習。該方法由伊恩・古德費洛等人於 2014 年提出。生成對抗網路由一個生成網路與一個判別網路組成。兩個網路相互對抗、不斷調整參數，最終目的是使判別網路無法判斷生成網路的輸出結果是否真實。生成對抗網路常用於生成以假亂真的圖片。

Repository）一份報告顯示，流行的人工智慧繪圖生成器的基礎中隱藏著數千張兒少性虐待圖像，這些相同的圖像使人工智慧更容易產生高度擬真的兒少性圖像（Thiel, 2023）。以人工智慧生成的兒少性剝削圖像製品，最大的爭議在於人工智慧結合了來自受性剝削和非受性剝削兒少圖像的大型數據庫，生成不存在但可能與實際兒少相似的、新的和逼真的性化圖像。

2024 年 5 月美國第一宗針對完全以人工智慧生成的圖像技術製作兒少性虐待製品的聯邦指控 Anderegg 案，威斯康辛州 Steven Anderegg 使用 Stable Diffusion 以文本生成圖像方式，創造數千張青春期前未成年人的擬真圖像，其中許多圖像描繪了全裸或半裸的未成年人展示或觸摸自己的生殖器，或與其他男性發生性關係。Anderegg 因涉嫌製作、散布和持有從事露骨性行為的人工智慧生成未成年人圖像，以及將類似的人工智慧生成的露骨色情圖像轉傳給未成年人而遭到刑事犯罪指控（U.S. Department of Justice, 2024）。

（二） 結合深偽技術變造合成的兒少性剝削製品

根據歐盟人工智慧法（EU Artificial Intelligence Act, AIA）第 1 章一般規定第 3 條定義（60），「深偽」（Deepfake）是指「運用人工智慧技術所生成或經由演算操控之圖像、音訊或視訊內容，其形式仿擬現實中存在之人物、物體、地點、實體或事件，並足以誤導人們信以為真，或誤認其具有真實性。」[3]（EU AI Act, 2024）。深偽一詞最早於 2017 年出現在 Reddit 網路社交平台，是「深度學習」（Deep learning）和「偽造」（fake）的結合體，是專門用來替換影片中人臉來製作色情影片的軟體名稱，最初僅限於色情內容，但當程式碼被分享並用於廣泛創作時，這種做法引發了道德問題。研究顯示 2019 年至 2023 年間，人工智慧處理的照片驚人地增加了 550%，顯示其可近性及存在濫用的風險（Romero-Moreno, 2024）。深度學習先驅吳恩達即曾發推文稱，DeepNude[4] 是「人工智慧最令人噁心的應用之一」（雷峰網，2020）。

利用深度生成技術對影像、聲音、文字等多媒體內容進行編輯等操作生成之內容統稱為 Deepfake（陳駿丞，2022），目前較常見的 Deepfake 軟體、應用程式或

[3] 原文為： ‘deep fake’ means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.

[4] DeepNude 運用柏克萊大學開發的開源演算法 pix2pix，為生成式對抗神經網路(GAN)，係以人工智慧運作的人體圖像合成技術，只要花短短數秒的時間，就能將有穿衣服的女性圖像，偽造成逼真的裸照圖像。

網站有 MyEdit、Reface、faceswapper、Remaker、Artguru、DeepAR 等。雖然深偽技術在娛樂、教育等領域有許多合法且創新的應用，但其濫用，尤其是用於生成兒少性剝削製品或其它虛假色情內容，已引發嚴重的法律和道德問題，並以以下幾種方式呈現：

1. 未經審查的人工智慧工具產生兒少性剝削圖像的方法，是結合成人色情內容和兒少的良性照片兩個不同的線上圖像，並從中學習創造新的圖像。
2. 以兒少的真實照片生成兒少性姿勢的新圖像，以真實兒少的非性圖像在視覺上進行修改以包含性內容，例如將一個原本拿著玩具的兒少修改握著成人生殖器的性圖像。
3. 將一個人的臉疊加在另一個人的身體上，透過將網路蒐集到其他未受害兒少的照片與受性剝削兒少的照片疊加起來，生成創建新的兒少性剝削圖像。
4. 生成式人工智慧具有改圖功能，使用者若將真實兒少受害者圖像上傳，藉由人工智慧修改生成看似虛構的動漫人物或非逼真的兒少圖像。

逼真的虛擬兒少性剝削圖像，使得辨別真假變得更加困難，過去的「眼見為憑」、「有圖有真相」，在現今人工智慧的年代可能已不再適用。隨著人工智慧生成的兒少性剝削製品的不斷增加，警察及執法部門需要耗費更多的時間來分析圖像以識別不存在的兒少。還有上述人工智慧深偽技術的誤用或濫用下，使原本兒少的良性照片被合成虛構成不實的性剝削圖像；而更令人擔憂的是當人工智慧修改生成看似虛構的動漫人物或非逼真的兒少圖像，可能掩蓋真實的犯罪，將會造成執法單位的誤判或是延滯救援。

二、生成式人工智慧對兒少及其家庭造成的影響

虛擬兒少性剝削製品隨著生成式人工智慧技術的發展而速迅擴散，使兒少與家庭面臨一種無形卻極具破壞性的風險。儘管這類內容不涉及真實兒少的直接接觸，但研究指出，若人工智慧生成圖像涉及兒少之肖像、姓名或身分特徵，仍可能對兒少當事人及其家庭造成深遠的心理與社會影響。

（一） 兒少的心理創傷與社交風險

雖屬虛擬生成，但若人工智慧性剝削內容涉及真實兒少的圖像或身分特徵時，其造成的心理創傷與真實性暴力相近。兒少在發現自身圖像遭不當使用時，可能出現羞恥、焦慮、自責與創傷後壓力症候群（Posttraumatic Stress Disorder, PTSD）等反應，尤其當圖像被擴散至社交圈或學校時，其自尊心、自我認同與社

交功能皆可能遭受重創（廖美蓮，2020）。生成式人工智慧的普及，也增加了兒少成為非自願數位受害者的風險。

（二） 家庭隱私與安全感的流失

當兒少圖像遭濫用為虛擬性剝削內容時，整個家庭將面臨隱私權受侵害的恐懼與焦慮。生成式人工智慧可能從社群媒體或家庭分享的圖像中擷取素材，進行深偽合成或性化處理，使家庭對日常分享與社交行為產生強烈戒備心理，甚至進一步選擇社會隔離，限制兒少網路活動。

（三） 社會標籤化與關係孤立

即使並未遭受實體性剝削，被描繪於虛擬兒少性剝削製品中的兒少及其家庭，仍可能因社群輿論與網路傳播遭到社會標籤與「二度加害」。青少年階段正處於自我認同建構期，對評價高度敏感，若遭遇公開羞辱或群體排斥，將可能導致嚴重社交孤立、自我否定甚至自傷行為（Amlani, 2024）。

（四） 對數位公民權與教育參與的侵蝕

虛擬性剝削事件對受害兒少與其家庭在數位環境中的信任感造成破壞，可能導致其不願再參與線上學習、社群互動或公開表達，進而限制其數位公民權的實踐。例如部分家庭為避免再度遭受傷害，選擇限制子女的網路活動與公開曝光，反而剝奪其在數位時代中應享有的平等參與權與教育接受權（UNICEF, 2021）。

（五） 家庭動力與心理壓力的重構

面對人工智慧生成虛擬性剝削所引發的事件，家庭經常處於情緒高張的情境中。父母或照顧者可能出現自責、憤怒、羞辱感，並轉而對受害兒少進行過度保護或限制，導致親子互信關係受損。兄弟姊妹亦可能因事件波及產生次級心理創傷，整體家庭氛圍陷入緊張與壓抑之中。

肆、防制人工智慧生成兒少性剝削製品的相關規範與作為

在未有相機以前，兒少性剝削內容多以繪畫、文字方式呈現，隨著視覺技術的進步，兒少性剝削呈現方式也隨之改變。現代常見的形式包括照片、影片、視頻和線上直播。過去兒少性剝削製作成本高、流通慢，但在相機、攝影機、錄放影機等技術發明並廣為使用後，兒少性剝削圖像數量大大增加，且因容易複製，成為成本不高、卻有高獲利的商品。20 世紀 80 年代和 90 年代初，兒少性剝削圖像變得容易取得，是否涉及兒少遭受性剝削及性虐待的不法問題也受到社會大眾

更多的關注，國際間紛紛透過立法及判決見解回應此問題（Palomares, 2024）。面對數位或人工智慧時代的來臨，國際間關於兒少性剝削製品之相關法律規範及防治策略如下。

一、聯合國有關兒少保護、防止性剝削之相關規範

（一）兒童權利公約

聯合國《兒童權利公約》（Convention on the Rights of the Child, CRC）是保障兒童[5]應有基本人權的國際法律，強調兒童因身心尚未成熟，因此其出生前與出生後均需獲得特別之保護及照顧，包括適當之法律保護。該公約第 34 條提及應保護兒童免於所有形式之性剝削及性虐待，包括防止：（a）引誘或強迫兒童從事非法之性活動；（b）剝削利用兒童從事賣淫或其他非法之性行為；（c）剝削利用兒童從事色情表演或作為色情之題材。公約內容雖有提及兒童色情，但對兒童色情並未加以定義（衛生福利部社會及家庭署，2014）。

（二）關於買賣兒童、兒童賣淫和兒童色情問題之兒童權利公約任擇議定書

聯合國 2000 年生效之《關於買賣兒童、兒童賣淫和兒童色情問題之兒童權利公約任擇議定書》（Optional Protocol on the sale of children, child prostitution, and child pornography, OPSC），在第 2 條（c）中將「兒童色情」定義為「係指以任何方式呈現兒童進行真實或技術合成之露骨性活動或主要為性目的而裸露與兒童性相關之部位。」；另外，在議定書第 3 條第（c）款規定，要求各締約國對於「製作、經銷、散布、進口、出口、期約、出售或持有第 2 條所界定之兒童色情」，締約國應以刑事法規作出充分規定，不論該等罪行是在國內或跨國所實行[6]（立法院，2017）。

（三）兒童權利公約第 25 號一般性意見—關於與數位環境有關的兒童權利

兒童權利公約第 25 號一般性意見聚焦於保障兒童在數位環境中的隱私、安全和平等，強調防止剝削、加強數位素養教育、完善法律框架，並推動國際合作保護兒童權益。該意見中強調：締約國應採取立法和行政措施，保護兒童在數位環境中免受暴力侵害，包括開展定期審查、更新和執行強有力的立法、監管和體制

---

[5] 依據兒童權利公約第 1 條規定，「兒童」係指未滿 18 歲之人；我國法律定義，「兒少」亦指未滿 18 歲之人。

[6] 關於買賣兒童、兒童賣淫和兒童色情問題之兒少權利公約任擇議定書電子資料，請參閱立法院人權公約專區 https://www.ly.gov.tw/Pages/Detail.aspx?nodeid=5966&pid=55849

框架，保護兒童在數位環境中免受公認和新出現的一切形式暴力的風險。這類風險包括身體或精神暴力、傷害或凌虐、忽視或虐待、剝削和虐待，包括性剝削和性虐待、販運兒童、性別暴力、網路攻擊、網路打擊和資訊戰。締約國應根據兒童不斷發展的能力，採取安全和保護措施（兒童少年權益網，2021）。

從上述聯合國有關兒少保護的相關文件可知，兒少色情已被兒少性剝削一詞所取代，兒少遭受性剝削不僅是真實兒少、也包括以技術合成虛擬的兒少，只要是基於性目的、裸露兒少有關性部位的任何方式，包括數位環境中，國家即有責任要提供兒少的保護，並以刑事罰法處理。

近年來，聯合國對於人工智慧生成的兒少性剝削影像問題給予高度關注，並建議全球採取一套以兒少權利為核心的應對方式。聯合國人權事務高級專員辦事處在 2024 年的報告中強調，人工智慧生成的兒少性剝削製品正成為一個全球性數位危機。報告指出，這些技術不僅對兒少的心理和社會安全構成威脅，還對法律究責帶來挑戰。聯合國呼籲各國採取「兒少權利為核心」的法律框架來管控人工智慧生成內容，建議各國更新法律以包括人工智慧生成的兒少性剝削製品，加強科技與法律的跨領域合作，並開發有效的技術工具以預防此類影像的生成和擴散，並確保所有措施能有效保護兒少的身心安全（OHCHR, 2024）。

二、歐盟人工智慧法案在兒少保護的相關規範

歐盟法律的變化是在 OpenAI 於 2022 年 11 月推出 ChatGPT 之後發生的。歐盟的官員當時意識到，現有立法缺乏解決新興生成人工智慧技術的先進功能以及使用受版權保護材料的風險所需的細節。歐洲理事會在 2024 年 5 月 21 日批准全球首部監管人工智慧法案（EU AI Act），強調在創新技術的同時，信任、透明度和問責制的重要性。基於風險管理，法案為人工智慧技術制定了全面的規則，將人工智慧系統分為：不可接受風險、高風險、有限風險和極低風險四種類別，針對不同風險類別採取相應的監管策略。由於法案對生成式人工智慧系統施加了嚴格的限制，歐盟將其稱為「通用」人工智慧，另外還包括尊重歐盟版權法的要求、模型訓練方式的透明度揭露、例行測試和充分的網路安全保護（巫敏慧，2024；彭思遠，2024；葉奇鑫、粘元馨，2025）。

鑑於兒少是易受傷害的脆弱群體應受到保護，免受利用與年齡相關的脆弱性的人工智慧系統的影響，歐盟人工智慧法序言第 28 條，提到人工智慧的許多有益用途之外，它也可能被濫用，並為操縱、剝削和社會控制實踐提供新穎而強大的

工具。這種做法特別有害和濫用，應予以禁止，因為它們違背了尊重人的尊嚴、自由、平等、民主和法治的聯盟價值觀以及《歐盟基本權利憲章》所載的基本權利，包括不受歧視的權利、獲取數據的權利保護以及隱私和兒少權利（EU AI Act, 2024）。歐盟人工智慧法案為防止生成式人工智慧可能造成兒少性剝削等危害兒少權益的情形發生，明確引用兒童權利公約及第 25 號一般性意見，明訂政策和作為對人工智慧系統開發業者兒少保護責任的依據。

三、美國人工智慧行政命令及風險管理框架

美國前總統拜登於 2023 年 10 月 31 日簽署「安全、可靠、值得信賴的開發及使用人工智慧行政命令」（Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI），要求美國公私部門遵循安全、保密、可信賴的方式發展及使用人工智慧，藉由該行政命令以及透過 8 項具體行動目標，指引公私部門開發及實施人工智慧安全措施，預防和避免人工智慧潛在風險及濫用（巫敏慧，2024）。但川普總統上任後為維持美國於人工智慧領域之領導地位，於 2025 年 1 月 23 日撤銷該項行政命令，以降低人工智慧等前瞻科技研究、發展之障礙（國家資通安全研究院，2025）；但放鬆監管政策可能使得高風險的人工智慧系統缺乏足夠的風險控管機制，增加科技誤用與風險的可能性，造成對隱私、安全、公民權利的潛在威脅；故目前美國加州、德州等多個州已積極立法，針對人工智慧生成兒童性剝削內容進行規範，以填補聯邦法律的空白（NCSL, 2024）。

另外，2023 年 1 月 26 日由美國國家標準與技術研究院（National Institute of Standards and Technology, NIST）發布的人工智慧風險管理框架（Artificial Intelligence Risk Management Framework, AI RMF 1.0），為自願性框架，提供相關資源協助組織與個人管理人工智慧風險，並促進可信賴的人工智慧（Trustworthy AI）之設計、開發與使用（陳箴，2024；NIST, 2023），該風險管理框架提供了一個系統性的方法來識別、評估和減輕人工智慧技術對不同利益相關者可能帶來的風險。全美 50 州檢察長 2023 年 9 月 5 日聯名致函國會參眾兩院兩黨領袖「為保護我國兒少免受 AI 危害，我們正與時間賽跑。事實上，城牆已被攻破；現在是行動的時候了」。「AI 正在顛覆一切，對執法界和保護兒少來說也是如此。壞人在擺脫正義束縛方面不斷進化，我們也必須隨之進化」，呼籲國會成立專家委員會研究人工智慧可能如何透過色情內容剝削兒少；並敦促國會儘速立法，將兒少性虐待定義擴大到人工智慧生成圖像，以保護兒少免受人工智慧危害。據世界日報報導

開發免費 ChatGPT 的 OpenAI 執行長奧特曼（Sam Altman）也在 2023 年 5 月參院聽證會上表示要減輕人工智慧系統日益強大的風險，政府的干預至關重要。奧特曼提議成立全美或全球機構，針對人工智慧系統頒發執照，確保遵守安全規範。美國國家標準與技術研究院（NIST）又於 2024 年 7 月 26 日提出「人工智慧風險管理框架：生成式人工智慧概況」（Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile），補充 2023 年 1 月發布的 AI 風險管理框架，提出生成式人工智慧特有或加劇的 12 項主要風險，協助組織識別生成式人工智慧可能引發的風險，並提出風險管理策略（許嘉芳，2024；NIST, 2024）。

四、我國有關性影像及兒少保護法制

隨著生成式人工智慧技術迅速發展，並在性別暴力情境中被日益濫用，我國近年亦加速強化數位性影像犯罪與兒少保護相關法律體系。自 2023 年起政府陸續修訂多部法律，針對生成影像、非自願性影像散布以及網路性暴力等新型態犯罪予以處罰。其中，《刑法》新增「妨害性隱私及不實性影像」章節，特別針對深偽（deepfake）與人工智慧生成性影像加以規範，第 319 條之 4 所設之「電腦合成製作不實性影像罪」，即明文禁止透過電腦合成或其他科技手段製作虛構性影像以侵害他人權益，回應了虛擬影像對人格權與性自主權的威脅（林東茂，2024；張麗卿，2024）。

此外，為保護女性、兒少等弱勢群體，政府亦修訂《犯罪被害人權益保障法》、《性侵害犯罪防治法》、《家庭暴力防治法》及《兒童及少年性剝削防制條例》等法，納入網路性私密影像事件的救濟與保護機制；其中包括賦予網路平台業者移除不當影像、主動通報警方之法定義務，強化數位平台之共治責任。

在專門保護未成年人的《兒童及少年性剝削防制條例》中，第 2 條第 3 款明確界定兒少性剝削製品，涵蓋「拍攝、製造、重製、持有、散布、播送、交付、公然陳列、販賣或支付對價觀覽兒童或少年之性影像，與性相關而客觀上足以引起性慾或羞恥之圖畫、語音或其他物品」。根據立法說明，此處所指之「圖畫」不限於真實影像，亦包含素描、漫畫、繪畫等描繪未滿 18 歲兒少進行露骨性行為的色情創作。對於所描繪兒少是否為真實人物，並非本條所問之重點，顯示條例已有涵蓋虛擬性剝削製品的立法意圖。然而，2024 年 2 月由網路內容防護機構 iWIN 通報要求社群平台下架一款成人遊戲，引發高度爭議。iWIN 認為該遊戲中的虛擬

角色呈現「幼態」、「兒少」等性化設定，可能構成違反該條例之兒少性剝削製品，要求平台依法移除。不過，此舉亦引來二次元創作者質疑其對創作自由的干預，並要求主管機關明確回應人工智慧或虛擬圖像之法律適用界限（曾子軒，2024）。針對該事件，主管機關衛生福利部於同年 4 月與法務部、民意代表及民間團體召開協調會議後做出函釋，若內容屬於兒少性影像、以真人為原型繪製之色情圖畫，或以人工智慧生成之擬真色情圖像，仍應依兒童及少年性剝削防制條例處理；至於動漫、二次元創作，若內容涉及暴力、性虐待、人獸性交，且欠缺藝術性、醫學性或教育性價值者，則應適用刑法第 235 條與兒童及少年福利與權益保障法第 46 條之相關規定（衛生福利部，2024；蕭宥辰，2024）。

雖然主管機關已透過行政函釋釐清部分虛擬性圖像之法律適用範圍，實務爭議暫時趨緩，但此事件突顯出我國對於人工智慧生成擬真兒少性剝削製品雖已有處罰機制，然在「虛構 vs. 擬真」的界線、「創作自由 vs. 兒少保護」的平衡，以及執法認定標準上，仍缺乏社會共識與明確的法律一致性基準。在人工智慧技術快速發展下，如何同步調整法制框架、確立技術中立原則與兒少保護優先的規範準則，將是未來法律實務與政策制定的重要課題。

五、網路兒少性剝削製品的相關防制作為

（一）兒少保護機構與網路業者合作

美國國家失蹤與受虐兒少援助中心（NCMEC）於 2022 年底推出 Take It Down 工具，以打擊兒少性剝削並幫助兒少從網路上刪除他們的色情圖像。Take It Down 是 Google 一項免費服務，允許來自世界各地的使用者提交報告，以說明刪除描繪 18 歲以下兒少的線上裸體、部分裸體或色情照片和視頻。業者 Microsoft 使用 PhotoDNA 技術保護用戶隱私及防止兒少色情影像的擴散，PhotoDNA 使用人工智慧演算法創建已知兒少色情影像的數位指紋（Digital Fingerprint），使平台能夠在不影響加密的情況下識別和刪除非法內容。OpenAI 也宣布從圖像生成工具 DALL-E 中移除可能被用戶用以創建色情圖像的內容（魏權龍，2023）。

Google 為致力打擊線上兒少性剝削內容，與非政府組織和業界團體攜手合作，透過多項計畫分享 Google 在技術方面的專業知識，同時也開發及提供相關工具，協助機構組織打擊線上兒少性剝削內容。當有人試圖搜尋兒少性剝削內容時，Google 搜尋不會傳回含有兒少的圖像，避免讓兒少與色情內容產生關聯（Google 平台, nd）。Google 訂定的服務條款規定，要求使用者不得利用任何 Google 平台/

服務儲存或分享這類內容。Google 也利用自動化偵測工具及指派經過專業訓練的人員進行審查；以及開放第三方和使用者提出檢舉，並通報給美國國家失蹤兒少與受虐兒少援助中心，2024 年 Google 在平台上發現有兒少性虐待內容向 NCMEC 提交回報的內容總數高達 535 萬 7,587 則，以及提供 NCMEC 對非法內容負有責任的使用者及未成年受害者的身分資訊、非法內容或是其他有幫助的相關脈絡資訊 117 萬 6,026 則（Google, nd）。

（二）業者遵循規範自行開發建置安全及審查機制

歐盟制定人工智慧法案目的是希望在促進採用人工智慧與確保人民有權利使用負責任、合乎道德和可信賴的人工智慧之間取得平衡。該法案採取風險管理策略，要求大多數的人工智慧系統需於 2026 年上半年以前開始遵守人工智慧法案；被禁止使用的人工智慧系統，必須在人工智慧法案生效的 6 個月後逐步淘汰，生成式人工智慧也納入通用人工智慧的規則管理（安侯建業，2024）。本文在網路蒐集資料之際，發現許多生成式人工智慧的業者都已加入使用的提醒及告知相關可能的法律責任，例如 Google 在使用者帳戶中若發現兒少性虐待內容時，除向 NCMEC 提交報告外，會採取停用或限制服務存取權的處置，2024 年因違規而受到處置的帳戶總數有 64 萬 2,959 個（Google, nd）。

（三）我國透過 iWIN 兒少網路防護機構及網路科技巡查防制兒少性剝削

為防止兒少接觸有害其身心發展之網際網路內容，我國委託民間團體成立 iWIN 防護機構，辦理兒少使用網際網路行為觀察、申訴機制之建立及執行，與推動網際網路平臺提供者建立自律機制等事項。iWIN 訂有「網路有害兒少身心健康內容防護層級例示框架」做為業者自律規範參考，於受理民眾檢舉網際網路內容有害兒少身心發展時，參考該例示框架通知業者自律移除，或限制兒少接取，倘網際網路內容疑似違反《兒童及少年性剝削防制條例》時，除立即請業者移除並保存證據外，亦同步將違法部分交由警政單位啟動犯罪偵查，或移由直轄市、縣（市）主管機關依法裁罰（衛生福利部，2021）。為加強兒少保護，2024 年 7 月新修之正《兒少及少年性剝削防制條例》，除了對持有兒少性剝削影像加重處罰外，授權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，必要時可強制封網。目前主管機關已規劃透過「爬蟲技術」辨識進行科技巡查，一旦發現兒少性剝削內容，就會主動處理（林周義，2024）。

伍、生成式人工智慧下的新興困境與挑戰

隨著生成式人工智慧技術的迅速演進，兒少性剝削問題的形式與風險出現擴散與變化（Wolbers et al.,2025）。相較於傳統以實體受害者為基礎製作與傳播的圖像，人工智慧生成的虛擬圖像可高度擬真且不涉及實際拍攝，卻仍呈現兒少被性剝削之情境，形成更加隱匿、難以追查、且傳播規模龐大的剝削樣態。特別是在 COVID-19 疫情期間，網路使用時間顯著增加，加劇網路性剝削問題的嚴重性（Christensen et al., 2021），而生成式人工智慧內容的擴散則為全球執法與法律制度帶來前所未有的挑戰。

此外，端對端加密（End-to-End Encryption, E2EE[7]）、匿名性與跨境平台的使用，亦使相關圖像的傳播更加難以監控與追查。歐洲刑警組織（European Union Agency for Law Enforcement Cooperation, Europol）2024 年網路組織犯罪威脅評估報告指出，這些技術環境成為犯罪者傳播兒少性剝削內容的重要途徑，使現有的辨識、執法與保護機制失靈。人工智慧生成的虛擬兒少性剝削製品引發的困境與的挑戰，以下分就技術層面、法律與治理困境以及受害者保護與犯罪風險三個層面探討。

一、技術層面的新挑戰

（一）技術門檻降低助長內容氾濫

生成式人工智慧的普及，使得非專業使用者也能輕易產出擬真度極高的兒少性圖像。相較於過去需仰賴繪圖與影像編修專業技能，現今使用者可透過現成平台進行圖像生成與修改，大幅擴大潛在製作者的範圍，導致非法內容產出與流通量劇增，對內容平台與執法機關皆造成龐大壓力。

（二）真假難辨造成鑑識困難

生成圖像與真實影像幾可亂真，難以透過肉眼或常規技術辨識其真偽。當無法確定圖像是否涉及真實受害兒少，將直接影響執法機關是否能介入調查與展開救援，也導致受害者身分確認與案件處理程序的延遲。

（三）複合性手法加劇內容危害

[7] 端對端加密（E2EE）是一種傳訊類型，它使訊息對所有人保持私密，包括訊息服務。使用 E2EE 時，訊息僅在傳送訊息的人和接收訊息的人面前以解密形式出現。傳訊者是對話的一個「端」，而接收者是另一個「端」；因此得名「端對端」。端對端加密，所以中間的任何人都不可能解密訊息。參考自 https://www.cloudflare.com/zh-tw/learning/privacy/what-is-end-to-end-encryption/

人工智慧技術除可生成全新影像外，亦能進行深偽操作，包括將兒少臉孔嫁接至成人色情影像，或以真實照片為基礎進行性化改造。這類內容混合虛構與真實元素，可能對被描繪或擬真之兒少造成嚴重的心理創傷、名譽損害與社會排斥。

二、法律與治理困境

（一）法律適用模糊與規範真空

目前尚有不少國家對生成式人工智慧所產出的虛擬兒少性剝削內容無明確法律規範。有部分國家僅針對含有實體受害者的圖像進行規制，但對於純粹虛構內容是否構成犯罪仍存爭議，導致執法機關難以適用現有法條，亦讓加害者得以規避法律責任。

（二）跨境法律落差削弱執法力道

人工智慧生成內容多透過暗網與加密平台進行跨境流通，而各國對此類內容的界定與處罰標準差異甚大，欠缺統一協調合作機制。即使部分國家已強化立法，但若缺乏跨國司法互助與資訊共享協議，仍無法有效進行追查、引渡或內容下架等執法行動。

三、受害者保護與犯罪誘發風險

（一）匿名與加密阻礙追查與識別

E2EE 平台與匿名網路環境使犯罪者可不具名地傳播兒少性剝削圖像，無須揭露身分或留下可溯源紀錄。即使圖像被發現，其來源、涉案者與是否為真實受害兒少的判別仍極為困難，嚴重限制執法介入的可行性。

（二）虛擬圖像可能誘發實體犯罪

人工智慧生成內容可能強化特定使用者對兒少的性幻想與慾望，甚至成為實際施害的前奏（Amlani, 2024）。若社會與法律對此類內容採取寬鬆態度，將可能促成加害行為的合理化與正常化。

（三）受害者身分辨識困難削弱救援機制

即便圖像為虛構，若以真實兒少照片為基礎進行性化生成，當事人仍可能承受極大心理與名譽壓力。NCMEC（2023）指出，此類變造影像難以透過現行技術比對辨識，使得相關機構無法即時啟動保護與救援程序，造成實質傷害的延續與擴大。

生成式人工智慧所帶來的兒少性剝削圖像問題，不僅在技術面構成挑戰，更涉及法律規範與受害者保護體系的根本性衝擊。其高擬真、低門檻與跨境傳播的特性，已超越傳統法治國家所能處理的邊界；尤其是在「真假難辨」、「無明確受害人」與「匿名難追溯」的多重特徵之下，既有的執法與社會回應系統正面臨失靈風險。為有效因應此一新興數位性剝削形式，亟需從國際合作、法律修正、平台責任與受害者支援等多面向進行系統性改革。惟有建立兼具即時性與技術能力的跨國防線，才能防止虛擬性暴力最終演化為真實且無法挽回的兒少傷害。

陸、結論與建議

生成式人工智慧的發展正以驚人的速度融入日常生活，且其應用日益多元化，已成為不可逆的科技趨勢。然而，隨著生成式人工智慧的進步，該技術被惡意應用於兒少性剝削的案例也層出不窮，無論是真實的兒少受害者還是虛擬生成的內容，都對社會倫理、兒少保護和犯罪防治形成挑戰。人工智慧所生成的兒少性剝削製品，雖未必直接涉及真實兒少受害，但其可能結合深偽技術或實際影像，造成模糊性與真偽難辨，給執法機構和受害兒少帶來複雜且深遠的影響。目前國際間對人工智慧生成兒少性剝削製品問題高度重視嚴陣以待，我國對此問題也應有更多的關注。

雖然警察、執法單位對真實兒少性剝削製品的偵查已疲於應付，但面對生成式人工智慧可能帶來的兒少保護挑戰，立法者和執法者在規範技術應用的同時，也需謹慎把握其中的平衡，除了改善法律規範以外，亦應考慮保護創作自由和科技創新的空間，透過倫理教育和內容分級制度來引導社會大眾對此類技術的正當使用，減少其負面影響。此外，生成式人工智慧平台與技術提供者也應承擔相應的社會責任，建立防範機制防止其技術被濫用於不法用途。

生成式人工智慧的應用在拓展創作可能的同時，也在兒少保護領域帶來前所未有的威脅，要求多方的協同努力來因應。期待透過完善的法律監管、嚴格的技術防範、深入的社會教育以及強化的跨國合作，能夠逐步解決生成式人工智慧在兒少性剝削議題上的挑戰。只有在多方共同努力下，科技進步與兒少權益保護之間才能實現良好的平衡，營造安全、健康的數位環境。

本文提出以下建議：

一、法律規範與更新

隨著人工智慧生成技術日益發展，應更積極更新並完善法律，以防止其濫用來創造兒少性剝削內容。目前法律仍集中於處理涉及真實受害者的兒少性剝削製品，對於人工智慧生成的內容，由於法律邊界模糊，建議增加針對虛擬兒少性剝削內容範圍的明確規範，並依其性質進行分類處罰，例如針對營利性創作、廣泛散播或商業性剝削的案件加重處罰，確保對此類案件的適當處理與懲治。

二、人工智慧平台開發業者的責任

人工智慧技術的濫用風險要求生成式人工智慧工具的開發者建立責任機制，確保技術不被用於生成非法內容。建議業者在平台上開發篩檢與數位指紋識別技術，能夠及時攔截並過濾不當內容，防止兒少性剝削製品的生成和傳播。例如，Microsoft 的 PhotoDNA 和 Google 的 Content Safety API 等技術已能識別非法兒少內容。若各人工智慧平台能夠建立類似的技術規範，並且與執法機構合作，其技術不僅將發揮保護功能，也可有效防止兒少受害者因人工智慧技術而面臨二次傷害。

三、公私部門的合作與技術創新

建議政府部門與企業合作，建置風險防範工具，透過自動篩檢技術、違規圖像數據庫的構建，來監控人工智慧工具的生成內容。像 Google、Microsoft 這樣的科技企業，已經和美國國家失蹤與受虐兒少援助中心合作，用來辨別和快速移除兒少不當性影像。政府可設立第三方技術審查機構，對生成式人工智慧工具進行監管及安全認證，並參考國際上的人工智慧風險管理框架進行標準化規範，確保此類技術的安全性和可靠性，防止被用於不當用途。

四、公眾教育及防範意識的提升

公眾教育對預防生成式人工智慧的濫用具有關鍵意義。家庭、學校、媒體、業者和政府應加強對技術濫用風險的認識，尤其要讓家長、教師、兒少以及科技業者瞭解人工智慧生成兒少性剝削內容的社會危害，避免無意間生成或傳播此類內容。學校在課程中可增加數位倫理教育，強調負責任的科技使用，並在社會宣傳中提升兒少保護意識，確保兒少免於任何形式的性剝削威脅。

五、加強跨國合作與情資分享

鑒於生成式人工智慧所帶來的犯罪問題已突破國界，各國政府和執法機構應加強合作，建立系統化的跨國聯合行動和情資分享機制。網路的匿名性與加密技術讓兒少性剝削問題的偵查面臨巨大挑戰，生成式人工智慧所創造的大量虛擬性剝削圖像使得執法人員難以區分真實受害者，建議政府相關單位應投入資源研究防制策略及新的偵查技術與工具，以遏制人工智慧大量生成兒少性剝削製品，落實有效兒少保護的國家責任。

參考文獻

一、中文部分

Chen, Michelle（2024）。**精選 14 大最強 AI 繪圖網站，四步驟免費線上 AI 生成圖片！CyberLink**，9 月 20 日。https://tw.cyberlink.com/blog/photo-editing-tips/2345/best-ai-image-generators

Google（未註明）。**防制網路兒少性虐待內容措施**。Google。2025 年 1 月 20 日檢索，取自 https://transparencyreport.google.com/child-sexual-abuse-material/reporting

Google 中文平台（未註明）。**打擊線上兒少性剝削內容**。Google。2025 年 1 月 20 日檢索，取自 https://protectingchildren.google/intl/zh-TW/#alliances-and-programs

Rae Yu（2023），**顛覆性的變革：ChatGPT-4 七大功能大解析**。Tenten，3 月 15 日。https://tenten.co/learning/chatgpt4-introduction/

立法院（2017）。**關於買賣兒少、兒少賣淫和兒少色情問題之兒少權利公約任擇議定書**。立法院人權公約專區。7 月 25 日。https://www.ly.gov.tw/Pages/Detail.aspx?nodeid=5966&pid=55849

安侯建業（2024）。**解密歐盟人工智慧法案**。KPMG，10 月 7 日。https://kpmg.com/tw/zh/home/insights/2024/04/decoding-the-eu-artificial-intelligence-act.html

艾貝斯網路 iBEST（2023）。**什麼是生成式 AI？5 分鐘帶你了解生成式人工智慧**。艾貝斯網路有限公司，3 月 23 日。https://www.ibest.com.tw/news-detail/aigc/

行政院（未註明）。**行政院及所屬機關（構）使用生成式 AI 參考指引**。國家科學

及技術委員會。檢索日期 2024 年 10 月 2 日，取自 https://www.nstc.gov.tw/folksonomy/list/c79bf57b-dc94-4aff-8d14-3262b5559cfc?l=ch

巫敏慧（2024）。國際生成式人工智慧監理趨勢研析。**臺灣經濟研究月刊，47(8)**，52-58。https://doi.org/10.29656/TERM.202408_47(8).0008

兒童少年權益網（2021）。**關於與數位環境有關的兒童權利的第 25 號一般性意見**，4 月 21 日。https://www.cylaw.org.tw/about/crc/28/535

林周義（2024）。衛福部攜手 Google AI 揪兒少性影像。**中時新聞網**，7 月 22 日。https://www.chinatimes.com/newspapers/20240722000508-260114?chdtv

林東茂（2024）。**刑法分則**（4 版）。一品文化。

國家資通安全研究院（2025）。**國際資安政策法制觀測週報第 64 期**，國家資通安全研究院，2 月 7 日。https://reurl.cc/EVW2gA

張子午（2024）。**兒少性剝削犯罪同恐怖攻擊等級——與美國情資合作，台灣的偵查學習和挑戰**。報導者，6 月 3 日 https://www.twreporter.org/a/child-and-youth-sexual-exploitation-ncmec-cybertipline-reports-to-taiwan-cib

張東文（2022）。淺談人工智慧的倫理議題。**明新學報，45(1)**，26-42。https://www.airitilibrary.com/Article/Detail?DocID=1728760X-N202301040003-00003

張瑞邦（2024）。**AI 生成「兒少性虐待製品」案件遽增，美國執法機關如何有效打擊網路兒少性剝削**？關鍵評論，4 月 25 日。 https://www.thenewslens.com/article/201915

張麗卿（2024）。生成式 AI 對刑事實務之挑戰 (上)。**月旦法學雜誌，350**，51~67。

許嘉芳（2024）。**美國國家標準暨技術研究院發布「人工智慧風險管理框架：生成式 AI 概況」**。資策會科技法律研究所，10 月。https://stli.iii.org.tw/article-detail.aspx?no=64&tp=1&d=9263

陳冠瑋（2023）。**生成式 AI 的法律迷霧：ChatGPT 等工具好方便，但相關風險你承受得了嗎**？換日線，2 月 1 日。 https://crossing.cw.com.tw/article/17324

陳婉潔（2024）。**生成式 AI 發展所面臨的倫理挑戰與未來競爭格局**。科技網，6 月 7 日。DigiTimes。https://www.digitimes.com.tw/tech/dt/n/shwnws.asp?id=0000692189_9RZ2KL987SB6DZL2ZYDWT

陳箴（2024）。**美國國家標準與技術研究院公布人工智慧風險管理框架（AI RMF 1.0）**。資策會科技法律研究所，4 月。https://stli.iii.org.tw/article-detail.aspx?tp=1&d=8974&no=64

陳駿丞（2022）。**深偽偵測—在深偽內容（Deepfake）氾濫的時代，眼見仍為憑嗎？**中央研究院中研院訊，4 月 7 日。https://newsletter.sinica.edu.tw/27065/

彭思遠（2024）。人工智慧的規範與發展。**臺灣經濟研究月刊，47(1)**，89-94。https://doi.org/10.29656/TERM.202401_47(1).0013

曾子軒（2024）。宮崎駿都恐審查不過？動漫圈炎上「iWIN」事件一次看懂。**遠見雜誌**，2 月 8 日。https://www.gvm.com.tw/article/110095

楊舒凱（2023）。**運用 ChatGPT 撰寫作業的風險**。臺灣學術倫理教育資源中心，3 月 17 日。https://ethics.moe.edu.tw/files/resource/knowledge/knowledge_06.pdf

葉奇鑫、粘元馨（2025）。由歐盟人工智慧法初探我國人工智慧基本法制定。**電腦稽核，(51)，32-52**。https://www.airitilibrary.com/Article/Detail?DocID=2073090X-N202503260010-00002

雷峰網（2020）。一鍵脫衣 AI 再次驚現！Deepfake 正在盜用你的自拍照，製造大量色情圖片。**科技新報**，10 月 25 日。https://technews.tw/2020/10/25/deepfake-deepnude/

廖美蓮（2020）。網路性誘拐—兒少影像性剝削實務挑戰。**當代社會工作學刊（11）**，1-32。

劉淑琴（2024 年）。泰勒絲深偽不雅照網路瘋傳 白宮促立法應對。**中央通訊社**，1 月 27 日。https://www.cna.com.tw/news/amov/202401270155.aspx

衛生福利部（2021）。**如何遏止數位、網路技術增加 兒少性剝削案件數量及風險**。立法院第 10 屆第 4 會期社會福利及衛生環境委員會第 20 次全體委員會議書面報告。衛生福利部，11 月 17 日。https://www.mohw.gov.tw/mp-1.html

衛生福利部（2024）。**有關虛擬兒童及少年圖畫涉及兒童及少年性剝削防制條例規定疑義與處理原則案**。衛生福利部，5 月 14 日衛部護字第 1130014287 號函釋。

衛生福利部社會及家庭署（2014）。**兒少權利公約-中文版**。CRC 資訊網，11 月 20 日。https://crc.sfaa.gov.tw/PublishCRC/CommonPage?folderid=29

蕭宥辰（2024）。iWIN 管太寬爭議落幕？協調會達 3 共識 保守團體堅持虛擬兒童色情入法。**三立新聞網**，4 月 11 日。 https://ynews.page.link/MvBTn

靜宜大學教育發展中心（未註明）。**生成式人工智慧（GEN-AI）使用指引與規範**。靜宜大學教育發展中心。檢索日期 2024 年 12 月 20 日。https://tdcenter.pu.edu.tw/p/412-1061-4972.php?Lang=zh-tw

魏權龍（2023）。**AI 技術加速新型犯罪與認知作戰**。財團法人海峽交流基金會，7 月 26 日。https://www.sef.org.tw/article-1-204-14016

顧振豪（2023）。生成式人工智慧與法律的和諧共舞。**當代法律，(18)**，47-52。https://www.airitilibrary.com/Article/Detail?DocID=P20240521001-N202409200006-00004

二、外文部分

Amlani. N (2024.7.11). *AI Exploitation and Child Sexual Abuse: The Need for Safety by Design*. https://www.justsecurity.org/97065/ai-anti-child-sexual-exploitation/

Christensen, L. S., Moritz, D., & Pearson, A. (2021). Psychological perspectives of virtual child sexual abuse material.*Sexuality & Culture, 25*(4), 1353–1365. https://doi.org/10.1007/s12119-021-09820-1.

Edinburgh University (2024.5.28). *Scale of online harm to children revealed in g lobal study*. https://www.ed.ac.uk/news/2024/scale-of-online-harm-to-children-revealed-in-globa

EU AI Act (2024). The EU AI Act. https://www.euaiact.com/

Europol (2024.7.26).*Internet Organised Crime Threat Assessment (IOCTA) 2024*. https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-threat-assessment-iocta-2024

Internet Watch Foundation (2024). 2024 Update: *Understanding the Rapid Evolution of AI-Generated Child Abuse Imagery*. https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/

Krishna, S., Dubrosa, F., & Milanaik, R. (2024). Rising threats of AI-driven child sexual abuse material.*Pediatrics,153*(2), e2023063954.

Laird, E., Dwyer, M., Woelfel C. (2024.9.26). Report –*In Deep Trouble: Surfacing*

Tech-Powered Sexual Harassment in K-12 Schools. Center for Democracy & Technology https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/

National Center for Missing & Exploited Children (2024.3.11). *Generative AI CSAM is CSAM*. https://www.missingkids.org/blog/2024/generative-ai-csam-is-csam

NCSL(2024.11.22). *Deceptive Audio or Visual Media ('Deepfakes'). 2024 Legislation* https://www.ncsl.org/technology-and-communication/deceptive-audio-or-visual-media-deepfakes-2024-legislation

NIST (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

NIST (2024).*Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile.* https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf

Office of the United Nations High Commissioner for Human Rights, OHCHR (2024.1.11). *Children facing higher risk of sexual abuse and exploitation from misuse of technologies: Special Rapporteur*. https://www.ohchr.org/en/press-releases/2024/10/children-facing-higher-risk-sexual-abuse-and-exploitation-misuse

Palomares, N. (2024). Child Pornography: A Growing Threat for Forensic Science in the Digital Age. Themis: Research Journal of *Justice Studies and Forensic Science*, *12*(1), 3.

Romero Moreno, F. (2024). Generative AI and deepfakes: a human rights approach to tackling harmful content. International Review of Law, *Computers & Technology*, *38*(3), 297–326. https://doi.org/10.1080/13600869.2024.2324540