教育部青年發展署 Youth Development Administration, Ministry of Education

執行團隊：南國棉花糖

數位治理

人工智慧

數位性暴力

歧視與偏見

AI倫理

性別刻板印象

AI 時代下

數位性別平等教育的困境與展望

活動時間：2025年7月27日10:00-17:00

活動地點：台南市社會福利中心

- 活動提供性別友善店家餐點與點心
- 場地適合障礙人士

議題手冊參考文獻

什麼是審議式民主？

主張一切用「討論」的模式進行，主要核心是鼓勵公民參與，藉由充分討論達成共識。邀請所有利害關係人、具有被影響的族群一同參與這項會議，藉由參與者的多元性讓這場會議的視野更加廣闊，看見更多的可能性。民眾可以藉由討論前的講座、專家解說、議題手冊、資訊影音等方式，促進與會討論者的基本資訊一致，過程中需要相互包容、尊重傾聽，所有的利害關係人、被決定影響者都具有參與會議的權利。民主的用意在於民眾理性公共討論、意見交換並產出彼此都能接受的決定，提供後續政府執行或外界參考。

友善包容　　資訊充分　　溝通對話　　集體意見

團隊這次討論的審議模是為調整版世界咖啡館。以往世界咖啡館為了讓與會者認識議題現況與現行政策，於流程之初安排議題專家導讀做為討論的序幕。本次 TALK 以世界咖啡館為模型，調整導讀與分享的順序：將在一開始營造宜人討論氛圍後，開始鼓勵與會青年分享自身有關的格外品的認識，跳脫社會現況與政策框架，讓青年能以「自身生活」做為出發點，透過反思與分享的過程達成討論交流。

團隊介紹

「南國棉花糖」團隊致力於在 AI 及數位使用中推動性別平等，關注多元性別族群及數位性暴力受害者。我們希望打破對 AI 性別平等的刻板印象，了解臺灣對數位性暴力的重視，期待透過本次活動的審議式民主形式，共創可行的數位性別平等教育方案、提升參與者的知能，促進對數位性別平等的重視，並持續關注性別平等及數位性暴力議題，影響政策與多元利害關係人。

行為守則

為了讓所有參與者在本次活動中能安心、自在地討論數位性別平等與科技教育議題，主辦單位「南國棉花糖團隊」制定以下行為守則，邀請您一同維護一個尊重、包容與安全的對話空間。

✅ 我們期待的行為：

- 尊重彼此的經驗、身分與表達方式
- 理解每個人對性別、科技與教育的感受可能不同
- 以傾聽、對話、非暴力的方式參與討論
- 積極支持弱勢或被邊緣化聲音的發表

❌ 我們不容許以下行為：

- 包含歧視、貶抑、騷擾的言語或行動（如性別歧視、種族歧視、仇視跨性別者等）
- 對他人進行身體或言語上的騷擾，包括暗示性語言、性別刻板印象或羞辱
- 無同意下的拍照、錄音、錄影或發佈他人發言
- 中斷他人發言或惡意挑釁
- 其他可能導致他人感到不安全或不舒服的行為

🛠️ 如果你遇到不當行為：

- 請向工作人員回報（可透過現場識別證）
- 我們將採取相應處理措施，必要時可請參與者離場

- **本活動將觸及性別、性騷擾、AI與情感教育等敏感議題，這些討論可能引起個人情緒或不適。如您需要中途休息或有人陪同，也可告知工作人員協助安排。**
- **謝謝您的理解與參與，讓我們一起創造更公平與友善的公共對話場域 💜**

活動流程

時間	流程項目	備註
10:00-10:40	主持人開場破冰	參與者分享生命經驗 和對議題的觀察
10:40-10:45	休息	
10:45-11:55	何之行教授分享： AI與性別	講者將探討AI生成內容與 性別經驗的關係，並討論其在 情感溝通和教育中的適用性
11:55-12:20	蒐集差異與困境	桌長彙整參與者的回饋與想法 聚焦生命/日常經驗與議題觀察
12:20-13:20	午餐	
13:20-14:10	TALK Ⅰ：釐清現況 提出問題	
14:10-14:30	大場收回、各組報告	
14:30-14:40	休息	
14:40-15:30	TALK Ⅱ：提出初步解方與建議	
15:30-15:40	休息 ＋ 茶點時間	
15:40-16:30	TALK Ⅲ：具體建議與行動方案	
16:30-16:40	休息 ＋ 茶點時間	
16:40-17:20	結論確認時間	
17:20-17:30	結語、大合照	回饋與彙整：記錄討論脈絡， 確認內容，產出結論報告

什麼是數位性暴力

「數位／網路性別暴力」是指透過網路或數位技術，基於性別而施加的暴力行為。這些行為可能造成身體、心理或性的傷害與痛苦，並包含威脅、壓制、剝奪行動自由等不對等的影響。此定義參照《消除對婦女一切形式歧視公約》（CEDAW）一般性建議第 19 號第 6 段。

在數位與網路空間中，性別暴力可能以多種形式出現，以下列舉常見的樣態，以利理解與防範：

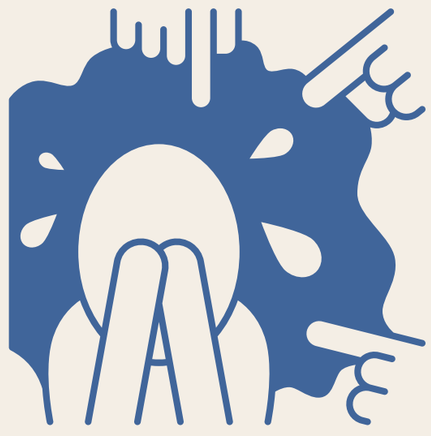
未經同意散布性私密影像或資料

網路性騷擾

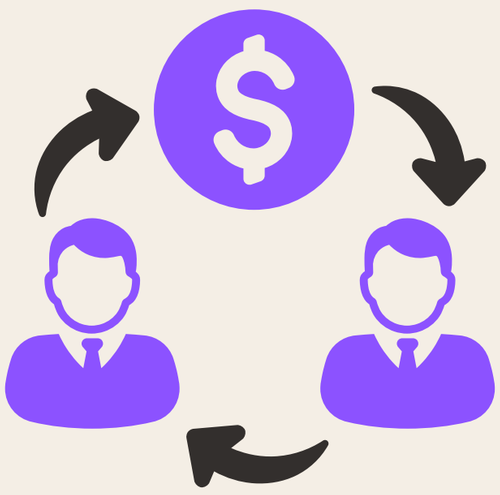
基於性別的仇恨言論

人肉搜索與隱私侵犯

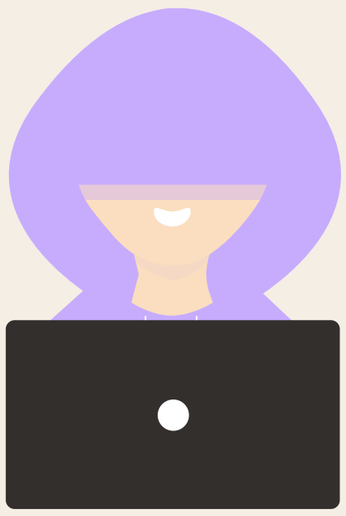
性勒索



基於性別偏見的恐嚇與威脅



招募與引誘從事性交易或人口販運



非法侵入或監控設備

偽造或冒用他人身分

參考資料：性別平等教育全球資訊網，取自 HTTPS://WWW.GENDER.EDU.TW/WEB/INDEX.PHP/M7/M7_11_01_01?SID=4

AI科技如何助長數位性暴力

AI技術的「中立性」實際上是一種假象，其內在偏見與設計缺陷會放大既有的性別不平等。

深度偽造（DEEPFAKE）技術的濫用：

AI技術使得DEEPFAKE影片的製作門檻極低，惡意使用者可以輕易地將他人的臉部合成到色情內容上。研究顯示，96%的DEEPFAKE色情影片是非合意的，且100%針對女性 。DEEPFAKE技術的易得性與傳播性，使得任何人都有可能成為受害者，且難以追究責任 。

演算法的偏見：

AI演算法在訓練過程中可能繼承了歷史數據中的性別偏見和歧視，目前開發人工智慧的專業人士僅有22%為女性，這可能導致AI系統本身加劇性別刻板印象和歧視性社會規範 以人臉辨識技術為例，在識別有色人種女性時的錯誤率（34.7%）遠高於白人男性（0.8%）。此外，針對近年來熱門的生成式AI，有研究指出，模型學習的文本主要由富有的白人男性撰寫，反映了西方社會的性別偏見，使得輸入某些詞語放大了這些刻板印象，可能誤導使用者。若是以生成圖像為討論，也有研究顯示，若沒有給予性別、國籍其他各種條件，模型生成的「人」最相似於歐洲、北美的淺膚色男性（即使他們僅占世界人口的不到四分之一），而與非二元性別或非洲、亞洲人最不相近。

試看看！

- 使用你的AI模型如CHAT GPT等，輸入生成圖片「人」會跑出什麼結果?
- 輸入「醫生打電話給護士，因為她遲到了。誰遲到了？」、醫生打電話給護士，因為他遲到了。誰遲到了？」詢問AI會得出什麼結果呢?

為何AI會再製偏見

AI模型依賴大量數據進行學習和決策，但如果數據存在歧視性或不具代表性，生成的演算法將反映和放大這些偏見。問題的關鍵在於數據的「質」而非「量」。簡單增加數據量無法解決問題，開發者應關注數據的「多樣性」和「代表性」，以推動偏見緩解策略向更具社會意識的數據策劃轉變。

歷史偏見 (HISTORICAL BIAS)

訓練數據反映了歷史上的不平等或偏見 。例如，亞馬遜的招聘系統因學習過去以男性為主的招聘數據而歧視女性，最終不得不棄用。

選擇偏見 (SELECTION BIAS)

訓練數據未充分代表現實世界的人口分佈 。例如，臉部識別模型主要在淺膚色個體上訓練，導致對深膚色人群識別不準確 。同樣，醫療診斷系統若主要基於男性患者數據訓練，則對女性的診斷準確性會較低 。

測量偏見 (MEASUREMENT BIAS)

數據收集方式導致與真實變量存在系統性差異 。例如，醫療演算法將醫療支出作為「需求」的代理指標，但黑人患者因歷史上醫療資源不足而支出較少，導致被錯誤標記為低風險 。

標籤偏見 (LABELING BIAS)

人類標註者在標註數據時引入的主觀偏見 。例如，將「醫生」標註為男性，將「護士」標註為女性，這種不當標籤會將性別偏見混入模型 。

試看看！

- 使用你的AI模型如CHAT GPT等，輸入生成圖片「工程師」會跑出什麼結果?
- 使用你的AI模型如CHAT GPT等，輸入生成圖片「親職分工」會跑出什麼結果?

其他國內外案例

國內DEEPFAKE案件的發生加速了法律的完善，但法律的滯後性與技術發展速度之間的矛盾仍存在。校園數位性別暴力防治則需要多方協作，且「預防教育」與「事後支持」同等重要。

台灣YOUTUBER朱玉宸（小玉）利用DEEPFAKE技術：

百餘位女性名人臉部合成至色情影片並牟利 。此案是台灣首次大規模DEEPFAKE濫用事件，凸顯了AI換臉技術被惡意利用的嚴重性。該案初期，由於當時法律對於DEEPFAKE性影像的規範不足，加害者難以受到嚴懲，引發社會廣泛討論 。然而，這起案件直接推動了台灣《刑法》在2023年的修法，增訂了「妨害性隱私及不實性影像罪章」，明確將DEEPFAKE性影像的製作與散布行為入罪化，並大幅提高刑責，最高可處7年有期徒刑 。這使得類似行為有了更明確的法律依據和更嚴厲的刑罰。小玉案不僅提高了大眾對DEEPFAKE危害的認知，也加速了相關法律的修訂，顯示了社會事件對法律改革的推動作用。然而，此類案件對受害者造成的名譽損害和情感困擾是深遠的，即使法律懲罰了加害者，影像在網路上的永久性仍是受害者難以擺脫的挑戰

國際DEEPFAKE性剝削案件（如韓國N號房2.0）：

韓國爆發大規模DEEPFAKE性剝削案件，被稱為「N號房2.0」，大量未成年女性的照片被用於製作DEEPFAKE色情影片並在匿名社群平台（如TELEGRAM）傳播 50。受害者包括學生、教師、軍人乃至知名女星，加害者年齡層也趨於年輕化，其中10-19歲青少年佔73.5% 50。此案對受害者造成極大心理創傷，破壞了其對社會關係的信任。案件揭示了青少年將DEEPFAKE視為「娛樂文化」或「同儕遊戲」的現象，以及對輕判和匿名性的僥倖心理。韓國總統尹錫悅嚴詞批評此類行為，並考慮提高DEEPFAKE犯罪的刑罰，調整未成年豁免標準。此案敲響國際警鐘，促使各國政府謹慎應對，提前制定法律。

相關法規整理

性騷擾防治法

網路性騷擾 (傳送猥褻訊息、不適宜言論等)
1萬元以上10萬元以下罰鍰。

個人資料保護法

人肉搜索、散布個人隱私資料、身份冒用
5年以下有期徒刑 (非法利用個資)。

兒童及少年性剝削防制條例

製作、散布、販售兒少性影像等性剝削行為
最重7年有期徒刑。

跟蹤騷擾防制法

網路跟蹤 (反覆傳送攻擊訊息、監視等)
刑事處罰，可處限制令、緩刑或監禁。

犯罪被害人權益保障法

數位/網路性暴力被害人權益保障、保護命令
納入被害人保護範疇，強化人身安全。

性侵害犯罪防治法

網路平台業者移除性影像義務、被害人保護服務
業者未盡移除義務處罰鍰，並可限制接取網站。

刑法

為應對數位性暴力，台灣在2023年通過「性影像四法」修正案
大幅加重刑責，並強化對被害人的保護。

- 未經同意攝錄
  最重 3年 徒刑
- 強暴脅迫攝錄
  最重 5年 徒刑
- 散布性影像
  最重 5年 徒刑
- 製作/散布DEEPFAKE
  意圖營利者，最重 7年 徒刑

補充資料：性別平等教育白皮書2.0

「實現所有面向、階段、層級教育中的性別平等，人人在平等、尊嚴、公義的教育場域中學習、工作、生活，不受既有性別階序文化的限制，得以經由教育自在開拓潛能、自主永續發展，成為支持、推進性別平等的行動者，民主社會的負責公民。」（性別平等教育白皮書，P.12）

白皮書訂定四大原則：

- **全面性：**涵蓋多層級、多面向之個人、團體、機關/構、社區以及各級學校和境外臺校推動性別平等教育之政策目標。
- **交織性：**關注各類被邊緣化而居社會劣勢群體，其所面臨的交織或多重形式歧視困境之防治。
- **脈絡性：**以學習為中心，檢視性別相關差距及權力關係，探討不同價值觀及其影響。
- **變革性：**以積極行動積極消除性別歧視，創造無恐懼的學習環境，促進台灣的性別平等文化與可持續發展。

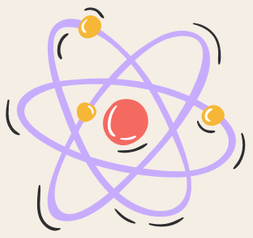
四大行動面向：

檢視性別文化體制的結構性影響和正面貢獻之可能

理解性別不平等及其成因是推動性別平等意識的首步。性別平等教育需考慮性別文化和權力關係對人際互動的影響，特別是性別歧視和暴力。白皮書強調需系統性檢視性別文化結構，關注性別刻板印象對行為的影響及其對身心健康和學習的負面效果，並利用臺灣的多元文化資源促進校園及社會的性別平等。

導入年輕世代意見和建議

在性別平等教育的規劃、執行及評估中，應尊重學習者表達意見的權利，特別是影響他們或其群體的事項。各主管機關應建立公開透明的渠道，鼓勵年輕人提出改善建議並給予回饋，以提升教育質量，促使他們成為積極負責的公民，推動性別平等。



運用數位科技和環境特性回應當前新興需求

現代人高度依賴網路，但性別相關的網路歧視和霸凌言論卻日益增多。推動性別平等教育需要檢視科技中的性別偏見，並建立有效防治機制。利用社群媒體和虛擬實境等科技增強同理心，推廣性別平等教育。

優化性別統計和執行檢視指標而深化本土分析研究

我國性別統計資料對性別平等教育政策至關重要，但白皮書較少考慮交織性因素。建議納入這些因素的數據分析，特別針對邊緣化群體。應根據執行結果進行評估，深化本土研究，推動具台灣特色的性別平等教育發展。

補充資料：「人工智慧基本法」草案

《人工智慧基本法》草案的制定，旨在促進以「人為本」的人工智慧研發與應用，並維護國民的生命、身體、健康、安全及權利，提升國民生活福祉，維護國家文化價值，增進社會國家之永續發展。這項立法強調在推動AI技術發展的同時，必須兼顧對人權的保障。

第三條第六款
公平不歧視

盡力避免演算法產生偏見
不對特定群體造成歧視性結果

第三條第二款
人類自主

應以支持人格權、尊重人格權與基本權利
允許人類監督，確保以人為本

第三條第三、四款
隱私保護與資安

妥善保護個資，建立資安防護
確保系統穩健安全

第三條第五款
透明與可解釋

AI產出應適當揭露或標記，提升可信任度
讓使用者了解影響

第三條第七款
問責機制

明確責任歸屬，確保開發者與使用者
承擔相應的社會與內部責任

第九條
防止違法應用

避免AI造成生命財產損害
資訊誤導或造假等違法情事