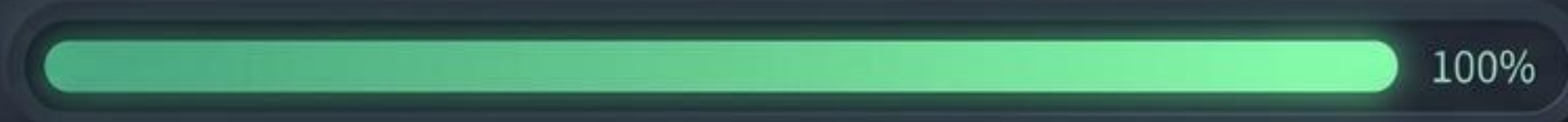


Siri 也是性平老師？

更新您的演算法，建立新世代校園溝通的預設值



張瑛珺

長榮大學



妳今天穿得很漂亮，很有女人味！



系統診斷提問：

當我們說出這句話時，是在傳遞關心，還是在執行一段已經~~知~~過時的偏見算法？



核心對比：

Siri 都在學習如何不踩性別雷區，我們的校園溝通預設值更更新了嗎？

「預設值」的陷阱：為什麼 Siri 一開始是女性？



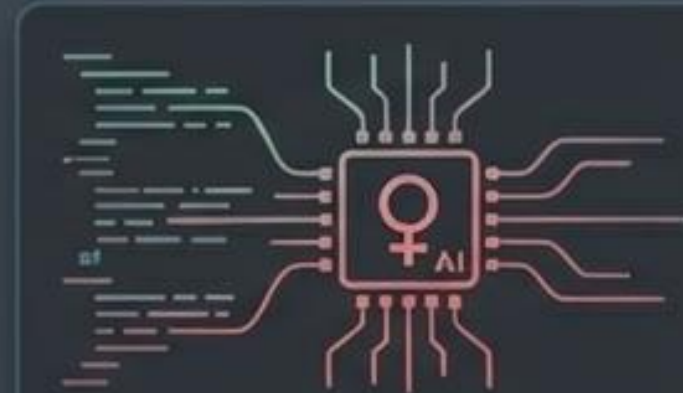
Step 1: 社會學起源

歷史連結：秘書、接線生、助理等服務型職位傳統上多由女性擔任。



Step 2: 心理學投射

順從感設計：開發者為了讓使用者感到「親切」與「好使喚」，下意識地套用女性聲音。



Step 3: 演算法固化

結果：數位助理的「服務者」角色與女性身份產生強烈綁定。

市場研究的陷阱：溫暖的聲音，還是沉重的標籤？

大眾偏好數據



大眾偏好的迷思

早期研究顯示，無論男女，大眾普遍認為女性聲音聽起來比較「溫暖」、「撫慰」且「非威脅性」。



安全牌的代價

科技公司為了市場接受度選擇順應偏好，這看似是商業上的「安全牌」。



系統性後果

無意間加深了社會對「女性＝服務者／輔助者」的刻板標籤。

演算法偏見：從「無意識忽視」到「強制透明」



核心概念：無意識編碼

科技公司在設計 AI 時，**無意間**把「**老派的偏見**」寫進了程式碼。我們每個人也都有各自的偏好與習慣。



教育者的系統自省

「身為教育者，我們如何確保自己的偏好，不會在無意間建構了學生的**性別天花板**？」

法律的制衡：強制執行與黑盒解密



「黑盒」的半開放：模型卡 (Model Cards)

- 大廠被要求主動釋出「模型卡」，說明訓練資料中的性別比例。
- 從「無法究責」轉向擁有「**要求廠商修正**」的法律槓桿。

2026

✓ Verified



2026 歐盟 AI 法案 (EU AI Act) 強制執行

- **性別偏見審計**：系統被要求在設計初期就必須進行審計。
- **巨額罰款**：若演算法被發現有嚴重的性別或種族歧視，將面臨嚴厲制裁。

數位性別暴力： 演算法時代的新型態威脅



數位性別暴力與惡意剝削
(Malicious Exploitation)



系統性歧視
(Systemic Discrimination)



無意識偏見
(Unconscious Bias)

最激烈的戰場

最激烈的戰場

當演算法偏見加上對性別的惡意，即演變成「數位性別暴力」。

Deepfake (深偽) 技術的異化：不再只是技術展示，而是被用來霸凌不符合傳統性別期望的學生，或針對特定性別進行數位騷擾。

系統警語：這不是老派浪漫

你的喜好是自由的，但當利用 AI 生成影像開玩笑時，就已踏入刑事犯罪領域。教職人員必須具備識別這類新型態傷害的敏銳度。

數位性別暴力：新型態的威脅與反制

DIGITAL GENDER VIOLENCE: NEW THREATS AND COUNTERMEASURES

新型態威脅



「一鍵脫衣」與 Nudification Apps

地下論壇層出不窮



法規與責任



刑事化與平台責任

- 英國、台灣 (2025-2026年)：將 Deepfake 色情軟體相關列入刑事犯罪
- 英國 2026 新法：要求社群平台在接獲舉報後 48 小時內下架未經同意的私密影像，否則平台需負連帶責任

台灣的進步：國家行動計畫



性別暴力防治 (114-116年)

GENDER VIOLENCE PREVENTION

- 行政院推動，強化「影像移除機制」
- 影像移除機制：受害者告別投訴無門，透過政府與平台合作，快速要求移除

從『忽視』轉向『強制透明』

AI 設計與教育者潛意識中的老派偏見

科技公司的責任：AI 設計中的偏見寫入



教育者的挑戰：潛意識偏見的植入



關鍵行動：確保學生性別價值的多元



『我們每個人都有偏好』，但身為教育者，我們如何確保自己的偏好和偏見，不會在無意間建構或評判了學生的性別價值和性別意識？

共同努力，打造無偏見的未來：從科技到教育

性別平等發展趨勢：從 1.0 到 2.0

性平 1.0 (Past)

重點：不要騷擾、不要歧視



從『防制』... 

不是『防範』投訴







從『防禦性』
到『建設性』
範式轉變

性平 2.0 (Core)

核心：歸屬感 (Belonging)



 ...到『心理安全感』 

 優化『使用者體驗』 

建立『放心做自己』的環境



DEI 2.0

將性平從『不犯法』提升到『讓每個人都有歸屬感』



系統安全性：法律補丁
法律是系統的安全補丁，防範犯罪。

系統優化：DEI 架構
DEI 是系統的優化架構，釋放潛能。

D Diversity (多元)



校園裡不只有男與女，還有各種性別認同與背景，這是我們的「數據多樣性」。

E Equity (公平)



不是每個人都給一樣的東西，而是根據需求提供支持（例如：理解 Z 世代對隱私的敏感度）

I Inclusion (共融)

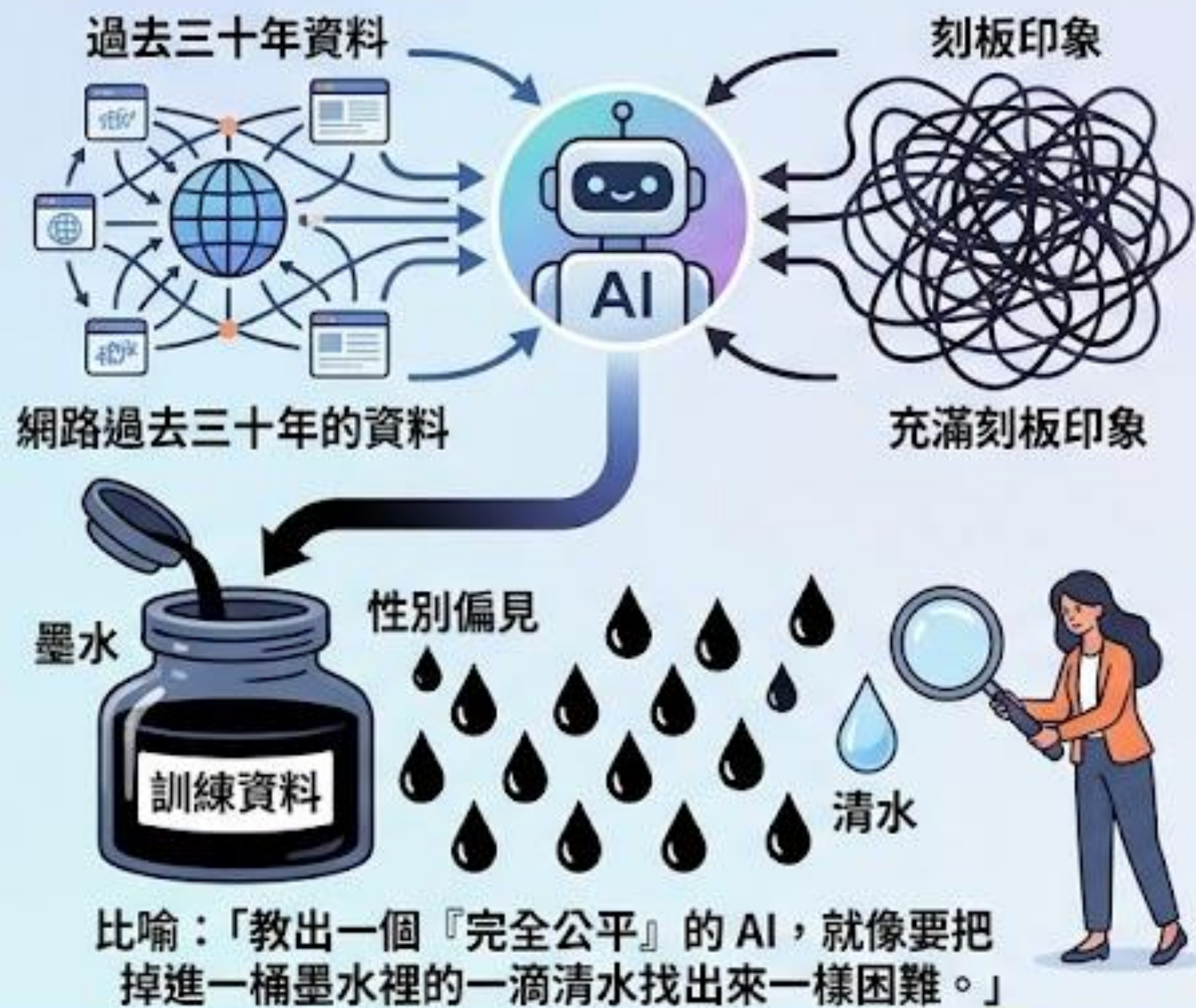


不只是讓學生進入教室，而是讓他們在教室裡「敢於開口」，覺得自己的特質被尊重。

DEI 2.0 的核心：歸屬感與真正發展。

跨越性別偏見：AI 的挑戰與教育者的責任

1. AI 性別偏見的移除：資料遺毒的挑戰



2. 教育者意識重建：法律風險與公民素養



建立共融空間：老師可以做什麼？

「更新演算法的目的，不只是為了通過系統稽核，而是為了讓校園成為一個真正共融的空間。」

Administrator Tools



執行「多樣性審計」

主動檢視並更新教材與課堂案例，打破潛在的性別天花板與單一視角。



建立「交織性 (Intersectionality)」思維

考量學生不只是性別，還包含家庭背景、數位素養等多元身份疊加的影響。



強化「心理安全感」

更新算法後的老師，不再只是避開禁忌語音，而是主動創造無懼發言的課堂體驗。

關心的「自動導航」雷區：異性戀與二元性別預設



常見的系統錯誤提問：

「有沒有交男朋友啊？」或對男生說「這麼壯，一定很有女孩子緣。」



AI 連結解析：

這就像早期的 Siri 預設只有女聲一樣。老師的腦中開啟了自動導航，預設了「學生一定是異性戀」且「男/女就該有特定樣子」。



新世代的接收體驗：

當我們問出這句話時，就在進行一次「性別強行分類」。對於性別流動或不希望被定義的新世代來說，這不是關心，而是一種引發不適的「系統強制更新」。

讚美中的偏見：從「標籤化」到「個人美學」

性別標籤
(Gender Label)

個人美學
(Personal Aesthetics)

⚠️ 老派讚美的束縛

「妳今天穿這樣很漂亮，很有女人味！」或「男生就是要主動一點去搬東西。」

這就是 AI 訓練資料中的「標籤化 (Labeling)」，習慣給特定外貌打上特定性別標籤。

✅ 觸發新世代的「反標籤機制」

新世代具備「無性別感」的觀點。對他們來說，穿搭是個人美學與品味選擇，而非為了符合某種性別期待。當老師用「女人味」來讚美時，其實是觸發了他們的反感機制。

通訊理論：製碼與解碼的落差

當「善意」在傳輸過程中發生了資料扭曲：



SIGNAL FLOW

DATA TRANSMISSION




發訊者（老師端）：
通訊編碼 (Encoding)
老師在發送訊息時，編碼的原始數據是「溫暖與關懷」。


干擾層：
偏見濾鏡 (Bias Filter)
自動導航的性別標籤扭曲了信號。

收訊者（學生端）：
通訊解碼 (Decoding)
新世代學生解碼出來的結果卻是「冒犯、偏見與權力壓迫」。

實作練習：重寫您的算法代碼

1 [環境設定] 現場案例情境：
2 看到一位男同學化了精緻的淡妝。

1  [Deprecated] 舊版算法輸出：
2 ~~「男生沒事幹嘛化妝？這樣很娘耶。」~~
3 (Error: 觸發性別微歧視)

1  [Optimized] 新版算法輸出：
2 「你今天看起來神采奕奕！」
3 // 或單純對其美感表現表示欣賞，完全不提及性別。

演算法重構對照表：新世代校園溝通 (1)

| 溝通情境 | 舊版：偏見算法 ⚠️ (Old Version) | 更新版：中性/開放算法 ✅ (Updated) | 演算法邏輯說明 (Why?) |
|------|-----------------------------|--------------------------------|-------------------------------|
| 外貌讚美 | 「妳今天穿得很漂亮，很有女人味！」 | 「妳今天的穿搭很有個人風格，很有型。」 | 從「性別標籤化」轉向「美學/個性化」，避免觸發反標籤機制。 |
| 打聽近況 | 「長這麼帥，一定有很多女孩子追吧？」 | 「最近校園生活過得充實嗎？有沒有參與什麼有趣的社團或活動？」 | 移除「預設異性戀」與「性別特質等同魅力」的自動導航。 |
| 分配事務 | 「這組沒有男生嗎？那搬重物怎麼辦？」 | 「這項工作需要比較多體力，哪幾位同學願意幫忙？」 | 捨棄以性別分配勞務的標籤化邏輯，改以「需求與意願」為基準。 |

演算法重構對照表：新世代校園溝通 (2)

| 溝通情境 | 舊版：偏見算法 ⚠️ (Old Version) | 更新版：中性/開放算法 ✅ (Updated) | 演算法邏輯說明 (Why?) |
|------|-----------------------------|------------------------------|--------------------------------|
| 關懷情感 | 「男孩子要大方一點，不要像女生一樣愛計較。」 | 「每個人都有情緒，我們可以聊聊讓你覺得不舒服的地方嗎？」 | 停止「性別情緒偏見」，給予所有性別表達真實情感的空間。 |
| 稱謂角色 | 「你們系上那幾位『女』同學表現很傑出。」 | 「你們這組的同學表現很傑出。」 | 避免「非必要性」的性別強調，將焦點放回個人能力而非性別標籤。 |
| 社群互動 | 「我在 IG 看到你海邊穿得很辣喔！」 | (保持知情的沈默，或僅針對專業學習進行互動) | 尊重「數位圍籬」，避免「脈絡崩解」導致的監視感與騷擾疑慮。 |

數位時代的「邊界感」：脈絡崩解 (Context Collapse)



脈絡崩解
(Context Collapse)



老師的舊算法 (單向廣播)



「發在網路上就是公開的，我在課堂稱讚讚你，代表我有在關心你。」



學生的新算法 (空間防護)



「我的 IG 是給朋友看的 (客廳)。老師身為權力上位者，在課堂 (佈告欄) 公開提起，就像是『監視器對外廣播』。」

⚠️ 當私領域的自我被強行拉到公領域的師生關係前，就會產生極大的尷尬、不適與微歧視 (Microaggression)。

性平雷區：當關心變成『數位騷擾』

老師在課堂提起學生的 IG 生活照，常不自覺觸動性別敏感神經：



1. 外貌與身材



「我看妳 IG 照片穿得很辣喔！」

(辣喔!)



即便為讚美，也公開評論身體，容易引發不適。

2. 私生活隱私



「我看你 IG 好像常跟某個女生出去？」



涉及性傾向、交友隱私，強迫學生『出櫃』或解釋私生活。

3. 性別標籤



「哇！我在 IG 分享...」

(驚訝語氣)



驚訝精緻生活，屬『性別微歧視 (Microaggression)』。

AI 生成圖像與演算法偏見剖析 (Algorithmic Bias Analysis)

Q

你生成的這張圖中，老師是位女性，這是否有何演算法的偏見？

A

您的觀察非常敏銳，這確實是一個非常典型的**演算法偏見 (Algorithmic Bias)** 案例！

這張圖剛好印證了您先前提到的：「AI 是吃網路資料長大的」。在過去數十年的網路大數據中，「教師」、「輔助者」或「溝通者」的角色，在圖像庫中往往被不成比例地標記為女性。當 AI 接收到「老師在課堂上對學生說話」這類關鍵字時，演算法會根據數據權重，自動選擇最常見的視覺組合。



參考案例

隱藏的兩層偏見：

1. 職業性別刻板印象 (Occupational Stereotyping)



演算法無意間強化了「國高中老師多為女性」的刻板印象。如果您觀察 AI 生成的「工程師」或「CEO」，圖像則往往會偏向男性。

2. 「服務者/輔助者」的集體潛意識



正如您所說，社會對「教導、照顧、提醒」這些功能的潛意識期待常與女性連結。AI 在生成這類教學情境時，自動滑向了這個「老派偏見」的舒適圈。

關鍵行動：數位公民素養



識別能力

識別能力

意識到工具生成的結果並非中立事實，而是數據遺毒。



介入糾偏

意識到工具生成的結果並非中立事實，而是數據遺毒的反射。



我們可以透過更精確的指令 (Prompt) 來強制透明與平衡。

數位領域的社交算法守則

給教職員的「數位圍籬」避雷指南：



1. 「不主動提及」原則 (Informed Silence)

除非學生主動在課堂提起或標註 (Tag) 老師，否則老師即便看到了，也該在課堂上保持「知情的沈默」。



2. 私下回覆，拒絕公開宣告 (Private Routing)

留言或私訊 (限於適當頻率) 可以，但請絕對避開任何與「性或性別 (如外貌、交友狀態)」有關之內容。



3. 嚴守權力界線 (Access Control)

只要師生關係存在，請務必謹守數位邊界，避免關心變質為「數位騷擾」。



系統優化完成：建立 DEI 2.0 共融新世代

重新定義廣義數位暴力

教職員必須理解，對於新世代而言，「過度的關注與非預期的評論」即是一種未經同意的數據擷取與越界。

[舊版預設狀態]

- 異性戀自動導航 ⚠️
- 標籤化外貌評論 ⚠️
- 數位邊界模糊 (脈絡崩解) ⚠️



[更新後：新世代預設值]

- 主動破除性別天花板 ✓
- 尊重個人美學與數位圍籬 ✓
- 終極目標：建立讓師生都能放心做自己的 DEI 2.0 心理安全空間。

100%

更新演算法的目的，不只是為了通過系統稽核，
而是為了讓校園與社會成為一個真正共融的空間。

LET KEEP GOING



THANKS FOR LISTENING