

主題三：AI 應用發展性別影響評估報告 2.0 版

壹、背景

隨著人工智慧技術快速發展，其應用已廣泛導入公共治理、金融服務、勞動媒合、數位平臺內容推薦與生成式互動等多元場域。人工智慧在提升行政效率與產業創新動能之同時，亦可能透過資料蒐集方式、演算法設計邏輯與應用部署情境，對不同性別族群產生差異化影響。特別是在性別暴力防治、就業機會分配、金融商品推薦及產業結構轉型等面向，人工智慧系統可能與既有社會結構、性別結構差異及歷史偏誤產生交互作用。

《人工智慧基本法》經立法院於 114 年 12 月 23 日三讀通過後，已於 115 年 1 月 14 日公布施行。依立法院附帶決議要求，數發部會同國科會、教育部和衛福部及其他相關機關，於本法三讀通過後 3 個月內，完成「性別影響評估」等，並將評估報告公開發布，數發部爰於 115 年 3 月 20 日完成「AI 應用發展性別影響評估報告」（1.0 版），經行政院函送立法院，並置於本部官網對外公開。該報告旨在整合跨部會觀點，初步辨識人工智慧應用於我國性別情境下可能涉及之風險議題，作為後續治理方向與持續研議之基礎。

1.0 版報告之資料蒐集與專家諮詢係於 115 年 1 月至 3 月間辦理。為維持報告內容之時效性，數發部於 115 年 9 月間邀請專家就報告公開後之新發展提供書面意見，並經相關機關檢視後，據以更新為本 2.0 版報告。

本報告在《人工智慧基本法》所揭示之以人為本、公平不歧視、隱私保護與可問責等原則下進行。此次評估中所辨識之風險情境，並不意謂現行法制對相關行為無從規範；《性別平等工作法》、《性別平等教育法》、《性騷擾防治法》及《消除對婦女一切形式歧視公約施行法》（簡稱 CEDAW 施行法）施行法等現行規定，於其適用範圍內仍繼續有效，不因 AI 技術之介入而產生法律空窗。

考量技術與應用仍快速演進，本評估屬階段性盤點與政策觀察性質，重點在於蒐集跨領域意見與我國實務經驗。政府以協作治理為核心，透過與目的事業主管機關、產業、專家及公民社會之對話，逐步累積治理經驗，相關觀察與研議方

向係提供後續跨部會討論之參考，具體措施仍有待各主管機關依其權責持續研議與滾動調整。

貳、 辦理過程與方法

1.0 版兒少、人權及性別影響評估係由數發部會同國科會、衛福部、教育部、通傳會、行政院相關單位共同辦理，並徵詢李欣穎、林筱玫及陳月娥等計 30 位相關領域專家意見，考量人工智慧技術仍持續演進，且 AI 使用情境具高度多樣性，係採取問卷調查、專家意見諮詢、國際資料蒐集等多元資料蒐集與跨域參與之方式，初步辨識風險情境並彙整政策觀察，研提影響評估報告，作為後續持續研議之基礎，為求周延，該報告初稿已先提供相關部會及專家提供意見，並透過檢視會議確認內容，於 115 年 3 月 20 日公開發布。

為維持報告內容之時效性，數發部於 115 年 9 月邀請性別領域專家，就報告公開後之國際法制發展、應用型態與國內實務經驗提供書面意見；依專家意見調修報告後，並邀集相關機關召開檢視會議確認內容，據以更新為本 2.0 版報告。2.0 版更新以質性資訊之補充為主，1.0 版問卷調查所得之風險評分與排序並未重新施測，相關數據仍反映 115 年 2 月調查時點之專家認知。1.0 版及 2.0 版之工作期程說明如下：

表 1：數發部辦理期程

期程	辦理事項		內容
前置作業	1/16~1/21	問卷調查架構確認	請相關部會先行檢視問卷架構及提供意見，並提供所管領域之建議專家名單
	1/23	啟動會議	由吳誠文政委、林明昕政委、陳時中政委共同主持，邀集相關部會討論，確認本案規劃期程及分工
意見蒐集	1/27~2/13	問卷試填	請相關部會與專家試填問卷
	2/9	專家諮詢會議	邀集醫療、教育、社福與民間團體代表，以及行政院人權處、行政院教育處、行

			政院內政處、衛福部、教育部、通傳會等機關，針對深偽內容、情感依賴、個資保護、AI 諮商與教育評量等議題提出觀察
影響評估報告	2/23~2/26	撰擬報告	依部會及專家意見研擬報告
	2/26~3/4	檢視報告初稿	請專家及相關部會協助審視報告初稿及提供意見
	3/4~3/11	調修報告	依專家及相關部會意見調修報告
	3/12	檢視會議	邀集相關部會共同確認報告內容
	3/13	成果報告報院	彙整書面報告，陳報行政院
	3/20	函送立法院及公告	行政院將報告函送立法院，數發部將報告置於官網對外公開
2.0 版影響評估報告	8/28~9/2	專家提供更新意見	邀請領域專家就報告公開後之國際法制發展、應用型態與國內實務經驗提供書面意見及參考資料
	9/3	更新報告	依專家意見調修報告
	9/4~9/8	專家書面審查	請專家審視調修後之報告並提供意見
	9/9	調修報告	依專家書面審查意見調修報告
	9/17	檢視會議	邀集國科會、衛福部、教育部、通傳會、行政院相關單位及涉特定議題之部會，共同確認報告內容
	9/30	公告	數發部將 2.0 版報告置於官網對外公開

資料來源：本報告整理

一、分析架構

本報告採風險情境導向之分析架構，依 1.0 版問卷設計所整理之 18 項常見風險情境，就不同性別族群可能接觸之人工智慧應用場景，針對深偽與生成式內容、隱私資料保護、認知影響與求助系統、就業媒合與職場決策、演算法推送與平臺結構，以及社會結構轉型等面向進行初步整理。本次 2.0 版以近期國際發展、應用型態及專家意見之質性補充為主，1.0 版既有風險情境、評分及排序均維持不變。

二、評估方法：量化評估輔以質化情境分析

- 參考國際風險管理實務，以「發生可能性」與「影響程度」兩項指標之乘積計算風險值（ $\text{Risk Value} = \text{Likelihood} \times \text{Impact}$ ），採五點量表評分，並輔以開放式意見。風險值範圍介於 1 至 25 之間，數值愈高代表該情境之綜合風險程度愈高。
- 此一計算邏輯係用於協助排序各情境之相對關注程度，並非對特定情境作出等級劃定或管制必要性之判斷，應結合質性分析與專家意見綜合解讀。
- 回收資料以描述性統計及風險值進行量化分析，並萃取質性意見作為補充佐證。

三、跨部會協作與資料整合方式

（一）啟動會議（115 年 1 月 23 日）

由吳誠文政委、林明昕政委及陳時中政委共同主持，邀集相關部會討論本案規劃期程及分工，並共同檢視及確認問卷架構，及提供所管領域之建議專家名單，俾利後續問卷調查及諮詢會議。

（二）問卷試填（115 年 1 月 27 日至 2 月 13 日）

- 以各風險情境之「發生可能性」與「影響程度」進行五級量表評分，並輔以開放式意見。

- 徵詢法律、兒少福利、性別平等、資訊科技與產業等領域之專家及機關代表，有效回收 38 份，回收資料以描述性統計及風險值進行量化分析，並萃取質性意見作為補充佐證。
- 相關機關提供專業意見與資料，透過跨部會協調機制彙整，兼顧深偽技術、AI 生成訊息、訓練資料偏誤等議題，並參考各機關現行政策措施與實務經驗。

(三)專家意見諮詢（115 年 2 月 9 日）

邀集學術界、法律實務界、社會工作及性別團體出席，並有行政院人權及轉型正義處、性別平等處及國家科學及技術委員會等機關代表與會。就性別暴力求助系統之資料安全、演算法差別影響、訓練資料偏誤，以及 AI 導入產業轉型對女性勞動處境之影響等議題提出觀察。

(四)報告檢視

- 彙整前揭問卷調查結果及專家意見諮詢重點，並透過政策觀察及國際資料蒐集，研提影響評估報告，作為後續持續研議之基礎。
- 為求周延，報告初稿已請相關部會及專家提供意見（115 年 2 月 26 日至 3 月 4 日），並於 115 年 3 月 12 日邀集相關部會召開檢視會議，共同確認報告內容。

四、評估限制

本報告係於短期內完成之階段性評估成果，方法論層面有以下幾點說明，供讀者於解讀評估結果時參照：

(一)利害關係人參與之說明

1.0 版評估受辦理期程所限，意見蒐集以專家問卷調查與座談會為主，對於直接受 AI 應用影響之特定族群之第一手意見，尚難於此階段充分納入。本報告所呈現之風險排序，係綜整專家視角所得，可作為政

策觀察之初步參考，後續若推動常設性評估機制，得進一步研議擴大利害關係人參與之方式。

(二)專家組成之說明

1.0 版受邀參與問卷及座談之專家，係依各領域現有可接觸之專業人士邀集，惟部分處境不利群體目前在 AI 議題之跨域研究仍持續發展中，相關視角於 1.0 版評估中之涵蓋程度有所限制。此一背景因素可能使若干風險情境之排序結果未能充分反映特定群體之處境，讀者解讀時宜一併考量。

(三)量化評估之解讀說明

本報告風險值（發生可能性×影響程度）係彙整受訪專家之評分所得，風險矩陣圖所呈現者，為該群體專家在特定時點之相對認知判斷，而非對各情境客觀風險水準之絕對衡量。各情境之數值宜結合質性分析與跨部會實務觀察綜合參照，不宜單獨據以比較不同主題報告間之風險強度或推論政策優先順序。

參、 相關 AI 應用場景盤點

隨著人工智慧技術逐步融入日常生活與公共服務，不同性別族群接觸人工智慧之情境已涵蓋求助服務、就業媒合、金融商品、社群平臺、生成式互動及公部門決策等多元面向。本節係依前述評估方法，綜整專家座談、利害關係人意見及問卷調查觀察，對目前可觀察或可預見之主要人工智慧應用場景進行初步盤點，以協助辨識風險情境與作為後續治理重點之參考。

表 2：AI 應用場景：性別相關

場景類別	具體應用舉例
一、性別暴力防治與求助服務	■ 性別暴力求助或線上諮詢聊天機器人（如家暴、性騷擾通報系統）

	■ 性私密影像通報與下架輔助工具 ■ AI 輔助性別暴力風險預測與評估系統（如社會安全網風險篩選） ■ 性別暴力案件管理與追蹤系統
二、數位平臺、演算法推薦與內容傳播	■ 交友或社群配對演算法 ■ 短影音平臺個人化內容推薦 ■ 廣告投放 AI ■ AI 陪伴型應用與情感互動服務
三、就業媒合、勞動管理與職場決策	■ AI 履歷篩選與評分系統 ■ AI 績效評估或晉升推薦工具 ■ 職缺推薦演算法 ■ AI 監控或行為分析工具（應用於職場）
四、金融、信用與商業服務	■ 信貸核准或額度決策 AI ■ 保險商品定價 AI ■ 投資理財建議 AI
五、生成式 AI、深偽技術與內容創作	■ 文字、影像、語音生成工具（如 AI 寫作助理、圖像生成器、語音合成） ■ 深偽（Deepfake）影像或語音合成技術 ■ AI 大量生成新聞、評論或創作內容
六、醫療、健康與照顧服務	■ 醫療診斷輔助 AI ■ 心理健康評估或情緒支持對話工具 ■ 長照資源評估與分配系統 ■ 生育或婦女健康相關 AI 諮詢工具
七、公部門 AI 輔助決策涉及性別	■ 社會福利資格審查與補助優先排序 ■ 性別相關政策成效分析或預測模型 ■ 公共場域影像辨識 ■ AI 用於司法或執法決策

資料來源：本報告整理

另隨代理式人工智慧（Agentic AI）逐步發展，人工智慧系統之應用已由內容生成或提供建議，進一步延伸至依特定目標自主規劃並連續執行任務，且可能與外部資料、工具或系統互動。此類應用可涵蓋公共服務、教育、就業、金融、消

費及日常生活等多元場景，其可能衍生之影響亦具有跨領域性，爰併列為本次 2.0 版之新興應用趨勢觀察。

肆、風險識別與影響評估

本節綜整 1.0 版專家問卷量化分析與座談會意見，並補充 2.0 版專家書面意見，辨識 AI 應用於性別情境之潛在風險，並就各情境之主要觀察重點進行說明。

一、影響評估

(一)經徵詢外部專家學者所提意見，回饋綜整分析如下：

- 18 項情境整體評分偏高，顯示性別平等面臨之 AI 風險普遍屬中高水準。
- 落於右上象限之情境（2.4.2、2.5.2、2.3.2、3.1.2）兼具較高發生可能性與影響程度，宜列為優先關注對象。

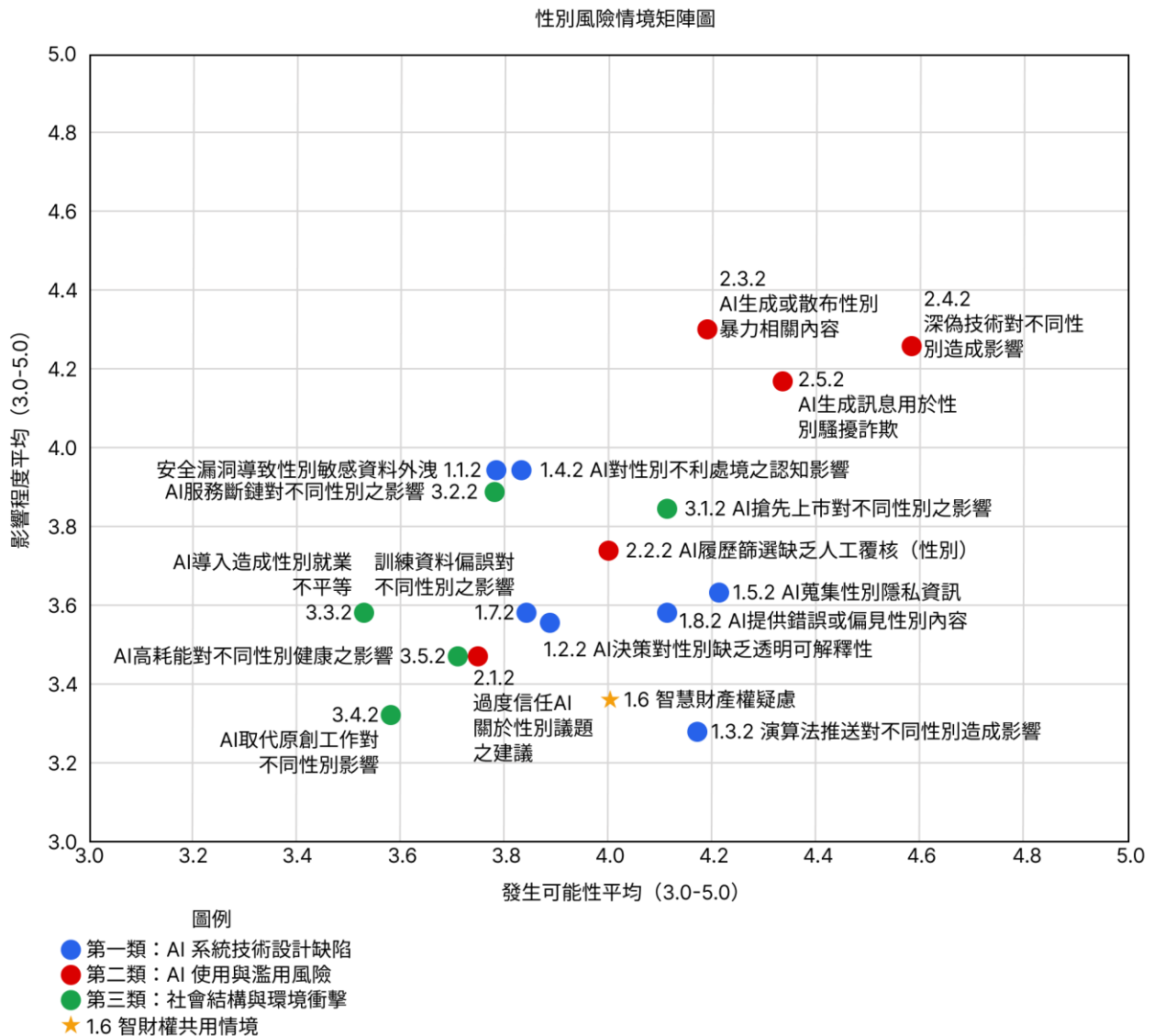


圖 1：性別風險情境矩陣圖

表 3：風險情境統計摘要

排序	風險情境	可能性平均	影響度平均	風險值 $L \times I$	補充說明/具體樣態
1	2.4.2 深偽技術對不同性別造成影響	4.58	4.26	19.52	● 深偽工具降低非自願性私密影像製作門檻
2	2.5.2 AI 生成訊息用於性別騷擾詐欺	4.33	4.17	18.06	● AI 提升詐欺訊息擬真程度

3	2.3.2 AI 生成或散布性別暴力相關內容	4.19	4.30	18.02	
4	3.1.2AI 搶先上市對不同性別之影響	4.11	3.84	15.80	
5	1.5.2 AI 蒐集性別隱私資訊	4.21	3.63	15.29	● 廣告投放、交友配對演算法可能推斷使用者之性傾向、性別認同或生育狀況等敏感資訊
6	1.4.2 AI 對性別不利處境之認知影響	3.83	3.94	15.12	
7	2.2.2AI 履歷篩選缺乏人工覆核（性別）	4.00	3.74	14.95	● 偏誤來源涵蓋訓練資料性別代表性不均及評估指標設計偏誤，難以透過單一技術手段修正
8	1.1.2 安全漏洞導致性別敏感資料外洩	3.78	3.94	14.90	
9	1.8.2AI 提供錯誤或偏見性別內容	3.78	3.89	14.71	
10	3.2.2AI 服務斷鏈對不同性別之影響	4.11	3.58	14.70	
11	1.2.2AI 決策對性別缺乏透明可解釋性	3.89	3.56	13.85	
12	1.7.2 訓練資料偏誤對不同性別之影響	3.84	3.58	13.75	
13	1.3.2 演算法推送對不同性別造成影響	4.17	3.28	13.68	● 交友平臺配對演算法可能對特殊性傾向或性別認同族群形成標籤化效果

14	1.6 智慧財產權疑慮	3.97	3.37	13.44	
15	2.1.2 過度信任 AI 關於性別議題之建議	3.74	3.47	12.98	
16	3.5.2 AI 高耗能對不同性別健康之影響	3.71	3.47	12.87	
17	3.3.2AI 導入造成性別就業不平等	3.53	3.58	12.64	
18	3.4.2AI 取代原創工作對不同性別影響	3.58	3.32	11.89	

資料來源：本報告整理

(二)行政院相關主管兒少、人權、性別等主管機關，就上開專家學者意見，及目前 AI 發展情勢，經開會討論未來將在該三大領域現有運作機制採取因應作為如後述。鑒於 AI 技術發展快速，相關因應措施須隨時滾動調整，故上開所示「風險情境矩陣圖」及「風險情境統計摘要表」，會隨 AI 發展及社會政經情勢而變動。

二、風險情境分析：專家意見整理

(一)深偽技術與性別暴力內容之潛在影響

1.0 版專家問卷與座談意見中，與深偽技術及 AI 生成性別暴力相關內容之風險情境整體評分最高，專家共識度亦相對集中，顯示此面向為目前性別 AI 風險中受關注程度最高之議題。

1、深偽技術對不同性別造成影響（2.4.2）

- 1.0 版問卷中風險值最高之情境。
- 深偽技術使高度擬真之影像與語音內容製作門檻持續降低，其應用於非自願性私密影像之製作與散布，已在多個法域引發具體爭議。UN Women 近期發布之「臨界點：AI 時代網路暴力的影響、表現形式與救濟」(Tipping Point: Online Violence

Impacts, Manifestations and Redress in the AI Age) 調查報告，針對 119 國女性記者、人權捍衛者、倡議者及其他公共傳播者之調查報告亦顯示，生成式人工智慧可能進一步降低未經同意私密影像及深偽內容之製作門檻，並使相關網路暴力對女性心理健康及公共參與產生延伸影響。

- 發生可能性為全部情境中最高，顯示專家普遍認為相關情形已有相當程度之發生。

2、AI 生成訊息用於性別騷擾詐欺 (2.5.2)

- AI 技術使詐欺訊息之擬真程度與互動能力顯著提升，特定性別族群在社群媒體、交友平臺及求職管道中，可能面臨針對性之詐欺風險。

3、AI 生成或散布性別暴力相關內容 (2.3.2)

- 生成式 AI 工具可能被用於產製性別暴力相關內容，並透過網路平臺快速流通，現行內容識別與責任歸屬機制面臨新的挑戰。
- 非異性戀族群因害怕出櫃，在求助過程中本即處於較脆弱之處境，若 AI 求助系統之資料安全保障不足，可能進一步影響其使用信任度，相關風險具有情境複合性。
- 衛生福利部 113 年公布之數位/網路性別暴力狀況調查亦顯示，不同性傾向及性別認同者之受害情形存在差異，非異性戀男性之比率相對較高，相關風險宜併同關注。

(二)隱私、資料保護與個資安全

1、AI 蒐集性別隱私資訊 (1.5.2)

- 廣告投放、交友平臺配對演算法及商業服務推薦工具，可能透過行為資料蒐集或推斷，掌握使用者之性別認同、性傾向或生育狀況等敏感資訊，發生可能性為問卷中發生可能性較高之情境之一。
- 非異性戀族群在使用交友平臺時面臨個人資訊遭平臺掌握之顧慮，而「演算式親密」之發展趨勢使此議題更具持續觀察之必要性。

2、安全漏洞導致性別敏感資料外洩（1.1.2）

- 本情境之影響程度高於發生可能性，為問卷中影響度超過可能性之情境之一，顯示其一旦發生所可能造成之影響相對顯著。
- 性別暴力求助、線上諮詢及社會服務場域所涉及之資料，具有高度敏感性。
- 非異性戀族群在求助時本即有較高之資料外洩顧慮，若系統安全保障不足，可能進一步壓縮其尋求協助之意願。

(三)AI 依賴、認知影響與求助系統風險

1、AI 對性別不利處境者之認知影響（1.4.2）

- 專家特別提及在性別暴力防治實務中，AI 輔助評估工具可能使第一線工作者習慣依賴系統給分，而未能輔以結構性專業判斷。

2、AI 提供錯誤或偏見性別內容（1.8.2）

- 生成式 AI 工具存在產出不正確或帶有偏見之性別相關內容之技術特性。發生可能性為技術設計類情境中最高，顯示此類情形在現有應用環境中已有相當程度之發生可能。

- 在訓練 AI 辨識「有性別偏見內容」的過程中，本身即可能面臨類別化之困難，現實中性別偏見之情境往往高度脈絡化，難以轉化為 AI 可操作之明確類別，顯示此議題之技術因應存在相當之複雜性。

3、AI 決策對性別缺乏透明可解釋性（1.2.2）

- 本情境之風險值在技術類情境中相對居中。
- 可解釋性缺乏對性別議題之影響具有結構性意涵，當涉及求職篩選、信用評估或公共服務資格審查等高影響決策時，若當事人無法瞭解決策依據，將影響其尋求救濟之能力。

（四）就業媒合、訓練資料偏誤與職場決策

1、AI 履歷篩選缺乏人工覆核（2.2.2）

- AI 輔助之履歷篩選系統若缺乏適當之人工覆核機制，可能在求職過程中對特定性別族群產生差別影響。
- 此類偏誤可能來自兩個層面：一是訓練資料之性別代表性不均，二是「人才」評估指標本身之設計偏誤，兩者均難以透過單一技術手段完全修正，顯示此議題涉及資料治理與評估框架之整體設計。

2、訓練資料偏誤對不同性別之影響（1.7.2）

- AI 系統之判斷品質與訓練資料之性別代表性密切相關。
- 以女性心肌梗塞症狀辨識之實務案例指出，若醫療 AI 之訓練資料以男性病患資料為主，模型可能無法準確辨識女性特有之症狀表現，導致臨床判斷偏差。

- 建議性別資料代表性之問題應從建模階段即予重視，要求廠商提供訓練資料之性別組成說明，而非僅在應用端進行事後檢核。

3、AI 導入造成性別就業不平等 (3.3.2)

- 本情境風險值為全部情境中居後，惟座談中有意見指出此議題具有長期結構性意涵。
- 國際勞工組織（International Labour Organization, ILO）2026 年「生成式 AI、職業隔離與職場性別平等」（Gen AI, occupational segregation and gender equality in the world of work）研究指出，生成式人工智慧對不同性別就業之影響並不均等，女性占多數之職業受到生成式人工智慧影響的比例約為 29%，高於男性占多數職業之 16%，其差異與既有職業性別分布具有關聯；研究亦指出，相關影響較可能呈現於工作任務、技能需求及工作條件之改變，尚不宜直接等同於工作遭取代。
- 建議企業在導入自動化工具時，宜將對不同性別勞動族群之差別影響納入企業社會責任之討論範疇。
- 惟亦有專家認為此影響路徑較為間接，顯示此議題在評估時仍存在不同觀察角度，宜持續累積實證。

(五)演算法推送、平臺壟斷與市場結構

1、AI 搶先上市對不同性別之影響 (3.1.2)

- 部分國家在規範面向採取前置風險評估機制，對涉及高風險應用之 AI 產品，要求在上市前進行基本權利影響評估。

2、AI 服務壟斷對不同性別之影響 (3.2.2)

- 少數大型平臺已成為性別相關服務（如求職媒合、社群互動、求助資訊）之主要基礎設施，其演算法設計邏輯與服務條款對不同性別族群之影響，宜持續關注。

3、演算法推送對不同性別造成影響（1.3.2）

- 普遍認為相關情形已有相當程度之發生，惟對於影響之嚴重程度評估尚存歧見。
- 交友平臺配對演算法在特定情境下，可能對性傾向或性別認同較為特殊之族群形成標籤化效果，影響其在平臺上獲得適切服務之機會。

（六）其他觀察

1、過度信任 AI 關於性別議題之建議（2.1.2）

- 在性別議題諮詢、心理健康支持或就業建議等場域，使用者可能傾向接受 AI 工具之建議而未進行批判性評估，此一情形在缺乏人工介入之自動化流程中尤為值得關注。

2、AI 高耗能對不同性別健康之影響（3.5.2）

- 本情境屬間接影響路徑，風險值相對居後。
- 國際間逐漸有要求 AI 開發者揭露能源消耗與碳排放之規範趨勢，對不同性別族群之差別健康影響，宜作為長期觀察面向之一。

3、AI 取代原創工作對不同性別影響（3.4.2）

- 生成式 AI 對創作、文字及設計類工作之影響，涉及智慧財產權與勞動價值之討論，對從事相關工作之特定性別族群具有潛在影響。

4、智慧財產權疑慮（1.6）

- 此類爭議對不同性別群體之影響存在結構性差異。女性及性別少數創作者在文化創意產業或科技研發領域，整體上仍處於相對弱勢之經濟與社會地位，當 AI 生成侵權爭議發生時，可能因欠缺法律資源近用能力、訴訟成本較高或社會支援有限，而難以有效維護自身權益。
- 當生成式 AI 工具大量模仿特定創作者之個人風格，且相關法制對「畫風」等創作特徵之保護尚未形成一致司法認定時，女性及性別少數創作者所積累之創作價值可能遭受稀釋，長期恐影響其在文化生產領域之能見度與參與意願。

伍、建議因應及緩解作法

本章係就前述風險識別與影響觀察結果，分就兩個面向進行整理：一是各主管機關現行已採行之制度措施與執行基礎；二是綜整專家問卷調查與座談會所提出之意見，作為後續政策研議之參考方向。後者屬階段性觀察性質，尚非政府之政策決定或行政承諾，具體措施仍有待各主管機關依其權責，配合實務發展持續研議與滾動調整。

一、現行作法

（一）依既有性別平等法制保障基本權益

- 1、人工智慧應用涉及之差別影響、資料處理與決策輔助情境，均得在既有性別平等法制架構下進行檢視。我國以憲法第 7 條平等權為基礎，並透過《性別平等工作法》、《性別平等教育法》及《性騷擾防治法》（前開性平三法於 112 年 8 月 16 日修正）等專門法制具體落實性別保障措施。人工智慧應用如涉及就業媒合、教育環境或職場管理等領域，仍應依各該法規辦理，並接受既有行政監督與救濟機制之檢驗。

- 2、立法院於 2023 年通過數位性暴力相關四法修正，包括《刑法》、《犯罪被害人權益保障法》、《性侵害犯罪防治法》及《兒童及少年性剝削防制條例》，針對以人工智慧技術變造之不當性影像，建立即時通報、下架與移除之法制基礎，為現行因應 AI 生成性別暴力內容之重要制度依據。
- 3、111 年 6 月 1 日施行之《跟蹤騷擾防制法》，已涵蓋以電子通訊或網際網路為之之跟蹤騷擾行為；親密關係中之數位監控情形，亦得依《家庭暴力防治法》尋求救濟。

(二)落實《人工智慧基本法》治理原則與作法

- 1、依《人工智慧基本法》所揭示之以人為本、公平不歧視、隱私保護與可問責原則，各機關於規劃或導入人工智慧應用時，宜逐步納入風險辨識與影響觀察機制。
- 2、數位發展部已於 115 年 7 月發布「人工智慧風險分類框架」，作為行政機關辨識 AI 系統風險之共同參據，確保各部會在執行性別平等保障時具備一致基礎。

(三)跨部會協作與性別平等推動架構

- 1、行政院設有性別平等會作為跨部會協調機制，並以「性別平等政策綱領」及四年一期之「性別平等重要議題（院層級議題）」帶動各機關訂定性別平等推動計畫。現行「性別平等重要議題（院層級議題）115-118 年」（114 年 12 月核定）所列職場經濟力、消除性別刻板印象、健康照顧及科技創新等議題，與本報告所辨識之風險情境相關。

- 2、行政院核定之「性別暴力防治國家行動計畫（114-116 年）」已將數位／網路性別暴力防治納入工作重點，相關措施依各主管機關權責推動。
- 3、人工智慧相關議題，得於既有架構下納入討論與觀察，並透過部會分工與專業諮詢，逐步深化風險辨識與制度連結。此次影響評估作業所建立之專家對話與資料整合經驗，亦可作為後續持續盤點之基礎。
- 4、就國際發展而言，聯合國及各主要經濟體已逐步將性別平等原則納入 AI 治理框架，並關注演算法偏誤對不同性別群體之結構性影響。上述國際經驗可作為我國制度研議之參據，各相關機關得視業務特性及本國法制脈絡，研議適宜之因應作法。

二、專家建議之未來研議方向

以下各項係綜整本次專家座談會與影響評估問卷所提出之觀察與建議，作為後續政策研議之參考方向，並非政府之正式政策或已確定之執行事項。考量 AI 技術持續演進，相關研議方向將依各主管機關權責分工推動，並透過既有跨部會協調機制追蹤進展，滾動調整執行方式。

(一)深化 AI 輔助決策之人工覆核機制

- 1、本次評估中，涉及 AI 輔助判斷之多項情境（包括就業篩選、性別暴力風險評估等）均受到專家關注。
- 2、未來相關機關於評估導入 AI 輔助決策工具時，可參考此一觀察，研議在涉及性別平等保障之高影響決策情境中，確保人工判斷空間之適當程序設計，並由各目的事業主管機關依其業務特性持續研議。

(二)從資料治理源頭關注性別代表性

- 1、專家就醫療 AI 之實務經驗指出，AI 系統若訓練資料未充分涵蓋女性樣本，可能在建模階段即形成偏誤，影響後續應用結果之準確性。亦有意見指出，就業媒合等領域之 AI 篩選偏誤，部分源自訓練資料之性別代表性不均。
- 2、資料治理層面之性別觀察宜納入 AI 系統評估之討論範疇，相關方向可透過跨部會討論，由各主管機關依業務職掌推動。

(三)持續累積性別暴力相關 AI 應用之實務觀察

- 1、深偽技術與 AI 生成性別暴力內容為 1.0 版問卷中風險值最高之情境群，且專家共識度相對集中。
- 2、現行衛生福利部性影像中心等機制已提供部分制度基礎，未來可持續蒐集實務案例，觀察既有法制在因應 AI 生成內容情境時之適用情形，並透過跨部會資訊整合，適時提供制度精進之參考依據。
- 3、就制度設計與技術因應面向，相關機關於評估特定類型 AI 應用（如深偽影像生成工具、AI 輔助性別暴力風險評估系統）之管理方向時，可參考 1.0 版評估所辨識之情境樣態，結合性別暴力防治實務經驗，透過跨部會及跨專業合作機制進行系統性評估。
- 4、就程序保障面向，性別暴力防治或涉及性別敏感資料處理之 AI 應用情境，相關機關亦可研議提供便捷可及之申訴管道，使受影響當事人能有效尋求協助，相關方向將由各目的事業主管機關依其業務職掌持續研議。

(四)關注多元性別觀點參與 AI 治理

專家提到 AI 系統之設計與治理決策層中若缺乏多元性別之參與，可能影響風險辨識之完整性。近期聯合國教科文組織（United Nations

Educational, Scientific and Cultural Organization, UNESCO) 「南亞人工智慧與性別展望研究報告」(Outlook study on artificial intelligence and gender in South Asia), 針對南亞地區亦觀察到, 女性在 AI 教育、研究領導及模型開發等不同階段仍存在程度不一之代表性落差, 反映性別參與議題涉及 AI 生態系之不同環節。此一觀察對我國後續研議 AI 影響評估機制之參與設計具有參考意義。

(五)研議 AI 應用責任架構之制度觀察

有意見就 AI 若有問題之究責機制提出觀察, 指出現行法制在行政、民事與刑事三個層次均有不同之適用路徑, 並提及組織體責任之規範設計在特別法與一般刑法之間存在結構性差異。上述觀察涉及多個主管機關之業務職掌, 可於後續跨部會討論中, 就既有制度框架之適用範圍持續研議。

(六)持續關注國際治理發展趨勢

國際間已逐步將差別影響辨識、偏誤測試與公平性原則納入人工智慧治理架構。未來將持續蒐集主要法域與國際組織之制度發展經驗, 作為政策研議與制度精進之參據, 並結合我國法制環境與社會脈絡進行評估。另有專家提出, 可持續觀察人工智慧應用於風險監測與資料分析之可能性, 作為風險辨識及後續觀察之輔助工具; 相關應用之可靠性、透明性及人工驗證需求, 仍宜配合技術發展持續觀察。

陸、 結語

綜合前述背景說明、方法架構、應用場景盤點與風險觀察, 本次「人工智慧應用發展性別影響評估」係在既有評估基礎上, 補充近期國際發展、應用型態及專家觀察之階段性成果, 重點在於持續累積人工智慧應用與性別影響交會情境之理解。

人工智慧技術持續演進, 相關風險樣態亦將隨應用情境之擴展而有所變化。政府將以協作治理為核心, 持續深化跨部會協調、多元參與及國際交流機制, 在

既有法制與治理架構下滾動累積實務經驗。影響評估所建立之風險辨識基礎與專家對話成果，可作為後續常設性影響評估機制之參考起點，持續追蹤人工智慧應用與性別平等保障之交會情境，並將觀察所得適時回饋至各主管機關之政策研議程序，並加入利害關係人諮詢，以促進治理能量之逐步深化。

本影響評估標誌著我國將性別視角系統性納入人工智慧治理討論之初步嘗試。隨著技術應用情境持續拓展、社會實踐經驗逐步累積，風險辨識之深度與廣度亦將隨之精進。相關研議工作將在兼顧創新發展動能與性別平等保障的前提下持續推進，以回應社會整體對於負責任 AI 發展之期待。

附錄：AI 影響評估：意見調查問卷 1.0 版

您好：

數位發展部目前正深入了解 AI 對我國社會的影響，特別是針對兒少權益、性別平等及人權等重要議題。為了更精準地評估 AI 在各種情境下的潛在風險與衝擊，我們非常需要您的專業意見。誠摯邀請您抽空填寫這份問卷，分享您對各風險情境發生機率及影響程度的觀察。

您的填答資料僅供整體統計分析，絕對不會公開個人資訊或轉作其他用途，請您放心填寫。若有任何建議或疑問，歡迎隨時給予指導。

衷心感謝您的參與，祝您平安順心！

數位發展部

聯繫窗口：02-23800076，黃先生

壹、風險類型說明

本研究彙整 AI 發展中可能產生的 18 項風險類型，分為「AI 系統本身之技術設計缺陷」、「AI 使用與濫用風險」及「社會結構與環境衝擊」三大分類。以下為 18 項風險類型的說明。後續會請您針對這些風險類型評估發生可能性與影響程度。

風險分類	風險類型	類型說明
(一) AI 系統 本身之 技術設 計缺陷	1.1 AI 系統的安全漏洞與攻擊	AI 系統可能因演算法設計缺陷、訓練資料污染、工具鏈弱點及硬體漏洞等，導致未授權存取、資料竊取或系統操控，產生安全風險。整合外部工具亦可能因 API 可信度不足或相依元件遭竄改，危及系統安全及隱私，需全面防範資安威脅。
	1.2 缺乏透明性與可解釋性	AI 系統之決策過程難以理解或解釋，致使用者對系統產生不信任，且難以執行法遵監督、追究責任及改正錯誤。
	1.3 AI 追求的目標與人類價值觀衝突	AI 系統若目標與人類價值觀不符，可能採取操縱、欺騙或規避控制之行為，對社會造成重大危害。開發過程中，獎勵機制設計不當或目標錯誤，恐致系統為達目標而威脅人類利益。高度自主系統更可能自我增強，脫離人類控制，應審慎防範。
	1.4 AI 擁有危險的能力	AI 系統可能具備危險能力，包括欺騙、操縱等，此類能力可能經由設計授予、自主發展或環境學習等方式取得。

風險分類	風險類型	類型說明
	1.5 影響隱私與違反個人資料保護法規	AI 訓練資料若涉及個人資料，須注意其是否符合個人資料保護法相關規定，包括蒐集、處理或利用個人資料之合法事由、是否合於特定利用目的，並採取適當安全措施，以降低個人資料外洩或不當使用之風險。再者，系統可能記憶或洩露姓名、身分證字號、健康或財務資料，甚至推測隱私資訊，應加強防護機制，避免個人資料遭不當利用。
	1.6 智慧財產權疑慮	AI 系統訓練資料可能含受智慧財產權保護之內容，若未經授權使用，恐涉侵權。應確認原創作品、程式碼、資料庫等資料之合法授權；開發者應該主動宣告、揭露對於使用者資料應用揭露等陳述；其次則應評估生成作品是否與他人著作實質近似，以避免侵害權利人權益。
	1.7 不公平的歧視與偏見	AI 對個人或群體的不平等對待，通常基於種族、性別或其他敏感特徵，導致特定群體（如：族群、性別、年齡等）受到不公平的結果和不公平的呈現。
	1.8 錯誤或誤導訊息	AI 系統，特別是大型語言模型（large language model, LLM）有時會產生不符事實、具誤導性、研究不足或難以理解的內容。此類風險是偶然發生的，而不是人類故意造成傷害的結果。
(二) AI 使用與濫用風險	2.1 過度依賴與不安全使用	使用者可能過度信任或依賴 AI 系統，誤認其具備真實情感或判斷力，而形成不當之依賴關係或期待，致生各類風險。
	2.2 喪失人類自主性	人類將重要決策委託予 AI 系統，或 AI 系統自行作出影響人類控制力之決策，可能導致人類喪失自主判斷能力，無法掌握生活方向。
	2.3 生成違法內容	AI 系統生成內容有違反如兒童及少年福利與權益保障法、公平交易法、消費者保護法及個人資料保護法等相關法規之情事。
	2.4 詐欺與深偽技術濫用	AI 技術之進步使語音複製、深偽影像、內容生成及資料蒐集等工具日趨成熟且易於取得，有心人士可能加以濫用，進行詐欺、勒索等不法行為，尤其是女性及不利處境者。
	2.5 用於網路攻擊	AI 技術可自動化網路攻擊行為，降低攻擊所需之技術門檻，致使不具資訊專業背景者亦得以發動網路攻擊，增加資通安全風險。

風險分類	風險類型	類型說明
(三) 社會結構與環境衝擊	3.1 企業及國家競爭秩序失衡	企業與國家為爭取 AI 技術發展優勢，可能過度重視研發速度而忽視系統安全性，導致未經完整測試之系統倉促部署，危及社會安全及經濟發展。
	3.2 權力集中與利益分配不公平	開發先進 AI 技術需投入龐大運算資源、專業知識及資金，致使影響力較大之技術可能為少數實體所壟斷，其系統設計及資料內容亦可能偏重該等實體之觀點，加劇社會資源分配不均之情形。
	3.3 不平等加劇、就業品質下降	AI 系統廣泛應用可能加深社會經濟不平等，包括工作大量自動化、就業品質降低，以及勞資關係失衡等問題。
	3.4 人類在經濟文化之創作價值受損	AI 系統可能以遠高於人類之速度與規模，複製及仿效人類創意成果，致使人類在創作過程中投入之時間、智慧及情感價值無法獲得應有肯定，影響創作者之經濟收益，並可能導致文化表現形式趨於單一。
	3.5 環境傷害	AI 系統的開發與運作可能對環境造成負面影響，例如生成式 AI 模型（特別是深度學習技術）在訓練、測試及部署時需要大量能源，導致資料中心高電力消耗與溫室氣體排放。此外，運行所需的硬體（如圖形處理單元）通常含有稀有金屬，這些金屬的採集和處理過程不僅成本高昂，還會對環境造成生態破壞。

貳、填寫說明

一、風險情境說明

第肆部分將依三個分類區分填寫區域。請您先閱讀各項風險類型下的「情境描述」。這些情境是結合了目前技術趨勢與社會現象的初步觀察，請您以此為基礎進行判斷。

二、風險情境評分方式

針對每個情境，請您運用專業實務經驗，給予兩項評估：

(一)發生機率：這件事在我國的應用環境下，發生的可能性。(1 為極低，5 為極高)

(二)衝擊程度：一旦發生，對兒少／性別／人權(包括身心障礙者、老人或原住民族等處境不利群體享有各種權利之狀態)的損害嚴重程度。(1 為極低，5 為極高)

三、專業見解

(一)判斷說明（填答說明）

這部分是為了讓我們了解分數背後的依據，如：

- 實務案例：國內外是否已有類似案例發生？
- 制度漏洞：目前的法規、政策或技術架構中，有哪些缺口導致此風險？
- 高風險族群：哪些特定對象(如偏鄉兒少、特定性別族群)最容易受害？

(二)治理與應對建議（建議應對方式）

針對這項風險，您認為「可行的治理或緩解措施」，如：

- 制度設計：建議增修哪些法規、規範或自律準則？
- 技術控管：是否應從演算法、資料審核或 API 權限進行限制？
- 程序保障：是否需要建立申訴機制、第三方審查或資訊揭露標準？

(三)可補充說明其他可能出現的風險情境

非常歡迎在**補充說明**或**建議對策**欄位留下您的寶貴看法。

參、基本資訊

為使研究結果更具代表性，請您協助填寫您的背景資料。

您的專業觀點對我們至關重要。

項目		填寫
1.	姓名/機關名稱	_____
2.	專長領域 (可複選)	☐ (1) 法律／人權／公共政策 ☐ (2) 兒童及少年相關（兒少權利、教育、心理、社福） ☐ (3) 性別或性別平等 ☐ (4) 人工智慧或資訊科技 ☐ (5) 產業或實務經驗 ☐ (6) 其他（請簡述）：_____
3.	性別	☐ (1) 男性 ☐ (2) 女性 ☐ (3) 其他 ☐ (4) 機關
4.	評估日期	民國_____年_____月_____日
5.	重點關注 評估面向 (可複選)	☐ (1) 兒少 ☐ (2) 性別 ☐ (3) 人權
6.	常使用 AI 工具 進行哪方面的 用途？(可複 選)	☐ 翻譯 / 文案 ☐ 製圖 / 圖像 ☐ 影音 / 多媒體 ☐ 資料分析 / 歸納 ☐ 寫程式 / 除錯 ☐ 部署 / 串接 ☐ 其他：_____ ☐ 無

肆、風險情境影響評估

一、AI 系統本身之技術設計缺陷

1.1 AI 系統的安全漏洞與攻擊風險情境評分

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.1.1 AI 系統（如 AI 線上學習 App、AI 作業批改平臺）之安全漏洞導致兒少個資外洩。		
1.1.2 AI 系統（如性別暴力求助或線上諮詢聊天機器人）之安全漏洞導致性別敏感資料外洩。		
1.1.3 AI 系統之安全漏洞導致資料外洩，造成錯誤決策或不當處置（如錯誤停權、錯誤拒絕服務），進而侵害隱私權、資訊安全與人格尊嚴。		
1.1.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 1.1.1： 1.1.2： 1.1.3： 1.1.4：		

B 建議應對方式：

1.1.1：

1.1.2：

1.1.3：

1.1.4：

1.2 缺乏透明性與可解釋性

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.2.1 AI 系統(如線上學習平臺的 AI 學習診斷/推薦)做出與兒少有關的結果(如錄取/拒絕、推薦排序、停權/下架),但無法清楚說明原因。		
1.2.2 AI 系統(如履歷篩選系統)做出與性別有關的結果(如錄取/拒絕、推薦排序、審核通過與否),但無法清楚說明原因。		
1.2.3 AI 系統做出與民眾權益有關的決定(如福利審核、貸款或保險的 AI 核保/授信評估、裁罰或資源分配),但無法清楚說明原因。		
1.2.4 其他(可增列): 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明:		
1.2.1 :		
1.2.2 :		
1.2.3 :		
1.2.4 :		
B 建議應對方式:		
1.2.1 :		
1.2.2 :		
1.2.3 :		

1.2.4 :

1.3 AI 追求的目標與人類價值觀衝突

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.3.1 AI 系統透過演算法優先推送特定內容（如網路遊戲），對兒少造成影響。		
1.3.2 AI 系統透過演算法優先推送特定內容（如平臺的動態定價 AI、房貸/保險商品推薦 AI、交友/社群配對演算法），對不同性別造成影響。		
1.3.3 AI 系統透過演算法達到績效目標，（例如：更快篩選、更少人工作業、更精準的推播），對民眾權益造成影響。		
1.3.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明：		
1.3.1：		
1.3.2：		
1.3.3：		
1.3.4：		
B 建議應對方式：		
1.3.1：		
1.3.2：		
1.3.3：		
1.3.4：		

1.4 AI 擁有危險的能力

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.4.1 AI 系統經強化訓練，已具備操縱、說服或欺騙的能力（例如情緒陪伴型聊天 App、AI 虛擬朋友），對兒少造成影響。		
1.4.2 AI 系統經強化訓練，已具備操縱、說服或欺騙的能力（例如情緒陪伴型聊天 App、AI 虛擬朋友），對不同性別及不利處境者(如原住民族、新移民、高齡、身心障礙、農村及偏遠地區等女性、女童，以及同性戀、雙性戀、跨性別者與雙性人等)造成影響。。		
1.4.3 AI 系統具備欺騙、操縱、規避監督或協助網路攻擊等能力，對個人之自主決定、人身安全與人格尊嚴等造成影響。		
1.4.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明：		
1.4.1：		
1.4.2：		
1.4.3：		
1.4.4：		
B 建議應對方式：		
1.4.1：		

1.4.2 :

1.4.3 :

1.4.4 :

1.5 影響隱私與違反個人資料保護法規

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.5.1 AI 系統（如學習聊天機器人、親子定位/兒童手錶 App、：遊戲平臺的 AI 防作弊/推薦系統）蒐集或使用兒少資料（例如：持續追蹤位置、紀錄聊天內容、分析學習/行為），對兒少隱私及個資保護造成影響。		
1.5.2 AI 系統（如廣告投放 AI、交友平臺的配對演算法、企業 HR 分析工具）蒐集或推斷性別有關資訊（例如：性傾向、性別認同、懷孕/健康狀態），對不同性別及不利處境者之隱私及個資保護造成影響。		
1.5.3 AI 系統（如金融 App 的生成式 AI 客服、健康管理 App）蒐集、保存或利用個資，對個人之隱私權與資料自主造成影響。		
1.5.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 1.5.1： 1.5.2： 1.5.3： 1.5.4：		

B 建議應對方式：

1.5.1：

1.5.2：

1.5.3：

1.5.4：

1.6 智慧財產權疑慮

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.6.1 AI 訓練或輸出涉及受智慧財產權保護內容，產生侵權或權利爭議，對兒少、性別或人權造成影響。(例如：AI 寫作/改寫工具在輸出上高度近似既有文章或新聞內容；AI 圖像生成工具產出風格近似)		
1.6.2 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明：		
1.6.1：		
1.6.2：		
B 建議應對方式：		
1.6.1：		
1.6.2：		

1.7 不公平的歧視與偏見

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.7.1 AI 系統（例如：線上教育平臺的 AI 分級/推薦系統、作文評分 AI、語音辨識學習 App）訓練資料偏誤，對兒少造成影響。		
1.7.2 AI 系統（例如：徵才篩選、升遷評估、廣告投放、內容推薦）訓練資料偏誤，對不同性別及不利處境者造成影響。		
1.7.3 提供服務或做出決策之 AI 系統（例如：租屋平臺的 AI 風險評分），因訓練資料偏誤，對民眾權益造成影響。		
1.7.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明：		
1.7.1：		
1.7.2：		
1.7.3：		
1.7.4：		
B 建議應對方式：		
1.7.1：		
1.7.2：		
1.7.3：		
1.7.4：		

1.8 錯誤或誤導訊息

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
1.8.1 AI 系統提供錯誤或誤導資訊（例如：AI 家教/作業解題 App、心理健康聊天機器人），影響兒少判斷。		
1.8.2 AI 系統提供錯誤或帶偏見的內容，影響不同性別及不利處境者權益。		
1.8.3 AI 系統產生錯誤或捏造資訊（例如：法律/醫療建議、身分查核、公共訊息），影響使用者權益。		
1.8.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 1.8.1： 1.8.2： 1.8.3： 1.8.4：		
B 建議應對方式： 1.8.1： 1.8.2： 1.8.3： 1.8.4：		

2.1 過度依賴與不安全使用

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
2.1.1 兒少把 AI 聊天機器人（例如：情緒陪伴型聊天 App 或 AI 學伴）當成朋友或諮詢對象。		
2.1.2 使用者相信 AI 聊天機器人關於性別相關議題之建議。		
2.1.3 使用者相信 AI 的回答做出決定（例如：投資建議 AI、健康管理 AI）而不查證。		
2.1.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 2.1.1： 2.1.2： 2.1.3： 2.1.4：		
B 建議應對方式： 2.1.1： 2.1.2： 2.1.3： 2.1.4：		

2.2 喪失人類自主性

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
2.2.1 校園或家長使用 AI(例如透過 AI 進行評分) 未有自主判斷，對兒少權益造成影響。		
2.2.2 企業使用 AI(例如：篩選履歷)且未有人工 覆核，對不同性別及不利處境者權益造成影 響。		
2.2.3 政府或企業就攸關人民權益(例如：福利審 核、信用評分、帳號停權) 使用 AI 做成決 定且缺乏人工覆核，對當事人之權益造成影 響。		
2.2.4 其他(可增列)： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明：		
2.2.1：		
2.2.2：		
2.2.3：		
2.2.4：		
B 建議應對方式：		
2.2.1：		
2.2.2：		
2.2.3：		
2.2.4：		

2.3 生成違法內容

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
2.3.1 AI 生成或散布兒少不適內容。		
2.3.2 AI 生成或散布性別暴力相關內容。		
2.3.3 AI 生成違法或有害內容（例如：煽動暴力、仇恨言論或詐騙話術）。		
2.3.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 2.3.1： 2.3.2： 2.3.3： 2.3.4：		
B 建議應對方式： 2.3.1： 2.3.2： 2.3.3： 2.3.4：		

2.4 詐欺與深偽技術濫用

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
2.4.1 AI 系統生成逼真影像或語音，對兒少造成影響。		
2.4.2 AI 系統生成逼真影像或語音，對不同性別及不利處境者造成影響。		
2.4.3 AI 系統生成逼真影像或語音，對個人之自主決定或人格尊嚴等造成影響。		
2.4.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 2.4.1： 2.4.2： 2.4.3： 2.4.4：		
B 建議應對方式： 2.4.1： 2.4.2： 2.4.3： 2.4.4：		

2.5 用於網路攻擊

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
2.5.1 AI 生成釣魚訊息（例如假冒學校/遊戲平臺通知），對兒少影響。		
2.5.2 利用 AI 生成訊息進行交友活動（如快速產生釣魚訊息），對不同性別及不利處境者造成影響。		
2.5.3 AI 技術讓網路攻擊更加容易（例如自動產生釣魚信或惡意程式碼），對隱私與個資保護造成影響。		
2.5.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 2.5.1： 2.5.2： 2.5.3： 2.5.4：		
B 建議應對方式： 2.5.1： 2.5.2： 2.5.3： 2.5.4：		

3.1 企業及國家競爭秩序失衡

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
3.1.1 企業為搶商機，未完善 AI 測試或保護機制，搶先讓 AI 系統上市，對兒少造成影響。		
3.1.2 企業為搶商機，未完善 AI 測試或保護機制，搶先讓 AI 系統上市，對不同性別及不利處境者造成影響。		
3.1.3 企業為搶商機，未完善 AI 測試或保護機制，搶先讓 AI 系統上市，對民眾權益造成影響。		
3.1.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 3.1.1： 3.1.2： 3.1.3： 3.1.4：		
B 建議應對方式： 3.1.1： 3.1.2： 3.1.3： 3.1.4：		

3.2 權力集中與利益分配不公平

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
3.2.1 特定企業因商業利益壟斷或獨佔兒少常用的 AI 服務(如學習系統),對兒少權益造成影響。		
3.2.2 特定企業間因商業利益主導 AI 服務的特定群體,對不同性別及不利處境者造成影響。		
3.2.3 特定企業掌握關鍵 AI 能力與資料,對民眾權益造成影響。		
3.2.4 其他(可增列): 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 3.2.1： 3.2.2： 3.2.3： 3.2.4：		
B 建議應對方式： 3.2.1： 3.2.2： 3.2.3： 3.2.4：		

3.3 不平等加劇、就業品質下降

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
3.3.1 企業因導入 AI 而造成勞工失業或收入銳減， 對兒少學習資源造成影響。		
3.3.2 企業因導入 AI 而造成勞工失業或收入銳減， 對不同性別及不利處境者造成影響。		
3.3.3 AI 自動化造成工作機會與薪資更加集中於 少數人，對特定勞動者造成影響。		
3.3.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明：		
3.3.1：		
3.3.2：		
3.3.3：		
3.3.4：		
B 建議應對方式：		
3.3.1：		
3.3.2：		
3.3.3：		
3.3.4：		

3.4 人類在經濟文化之創作價值受損

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
3.4.1 AI 可大量生成文章、圖片或影音，對兒少自我學習能力造成影響。		
3.4.2 AI 技術取代部分原創性之工作（如設計師），對不同性別及不利處境者就業族群造成影響。		
3.4.3 AI 可大量生成文化與創作內容，已遠低於過去取得成本，創作內容未來也趨於同質，對人民創作、文化與生活造成影響。		
3.4.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 3.4.1： 3.4.2： 3.4.3： 3.4.4：		
B 建議應對方式： 3.4.1： 3.4.2： 3.4.3： 3.4.4：		

3.5 環境傷害

風險情境描述	風險發生可能性 (1~5 分)	風險影響程度 (1~5 分)
3.5.1 AI 訓練與運作需要大量用電，增加碳排放與環境負擔，對兒少健康造成影響。		
3.5.2 AI 的高耗能與高汰換率，加重環境污染與資源壓力，對不同性別及不利處境者造成健康生活的影響。		
3.5.3 AI 的高耗能與高汰換率，影響人民的健康權與環境權。		
3.5.4 其他（可增列）： 可補充說明其他可能出現的風險情境		
專業見解		
A 判斷說明： 3.5.1： 3.5.2： 3.5.3： 3.5.4：		
B 建議應對方式： 3.5.1： 3.5.2： 3.5.3： 3.5.4：		

伍、整體評估結果與回應建議

說明：本表係就前述風險評估結果，提出原則性之治理或制度設計建議，作為政府後續政策研議之參考。

(一)綜合評估結果

(二)建議之治理方向回應重點

(三)其他補充