主題一：AI 應用發展兒少影響評估報告

壹、 背景

隨著人工智慧技術快速發展，其應用已逐步融入教育學習、社群互動、內容創作與心理支持等多元場景，兒童及少年亦成為重要使用族群之一。人工智慧在促進學習效率與資訊取得之同時，亦可能伴隨認知發展、隱私保護、內容安全與社會互動等面向之新興風險，如何在鼓勵創新與保障兒少權益之間取得平衡，已成為各國政策治理之重要議題。

《人工智慧基本法》於 114 年 12 月 23 日三讀通過後，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「兒少影響評估」等，並將評估報告公開發布，爰本「兒少影響評估」報告，旨在於整合跨部會觀點，初步辨識人工智慧應用於我國兒少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。

本報告在《人工智慧基本法》所揭示之以人為本、隱私保護與公平不歧視等原則下進行，並兼顧兒少最佳利益之政策精神。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；《兒童及少年福利與權益保障法》、《個人資料保護法》及《刑法》等現行規定，對於涉及兒少之 AI 應用行為，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

本次兒少、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，且兒少使用情境具高度多樣

性，本次採取問卷調查、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，期程規劃說明如下：

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

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，整理 18 項常見風險情境，就兒少可能接觸之人工智慧應用場景，針對認知發展、隱私安全、內容環境與社會互動等面向進行初步整理。

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（Risk Value = Likelihood × 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，兼顧教育現場、兒少保護、科技發展與社會治理等面向，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）：

召開會議邀集醫療、教育、社福與民間團體代表，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。

(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

本次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（如兒少本人與其照顧者）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

本次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，相關視角於本次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x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，兒童及少年接觸人工智慧之情境已由單一工具使用，擴展至學習、社交、娛樂與情緒支持等多元面向。本節係依前述評估方法，綜整專家座談、利害關係人意見及國際治理觀察，對目前我國兒少可能接觸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：兒少相關

場景類別	具體應用舉例
一、數位學習與智慧教育系統	■ 個人化學習平臺自動調整教材難度 ■ 自動批改作文/程式碼並生成學習診斷報告 ■ 學習歷程追蹤（登入頻率、錯題率等） ■ 生成式 AI 查詢知識、翻譯、撰寫摘要 ■ AI 解題與家教 App
二、數位平臺、演算法推薦與內容傳播	■ YouTube/TikTok/IG 依行為數據推送個人化內容 ■ 「你可能認識的人」社交推薦 ■ 精準廣告投放（遊戲推廣、課金誘導） ■ AI 自動偵測並移除違規內容
三、陪伴型與互動式 AI	■ 情緒陪伴聊天 App ■ AI 虛擬朋友/虛擬寵物 ■ 心理健康評估或情緒支持對話工具 ■ 語音助理

四、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
五、兒少定位追蹤與行為監測	■ 定位手錶/追蹤 App ■ 家長監控軟體（App 使用時間、瀏覽紀錄） ■ 校園 AI 影像分析（出缺勤、專注度偵測）
六、公部門 AI 輔助決策涉及兒少	■ 社會福利資格審查與補助優先排序 ■ 兒少保護風險預警（通報數據篩選） ■ 教育資源分配輔助決策

資料來源：本報告整理

肆、 風險識別與影響評估

本節綜整專家問卷量化分析與座談質性意見，辨識兒少面臨之 AI 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,顯示兒少面臨之 AI 風險普遍屬中高水準。
- 落於右上象限之情境（2.4.1、1.3.1、2.3.1、2.5.1、1.5.1、1.4.1）兼具較高發生可能性與影響程度，宜列為優先關注對象

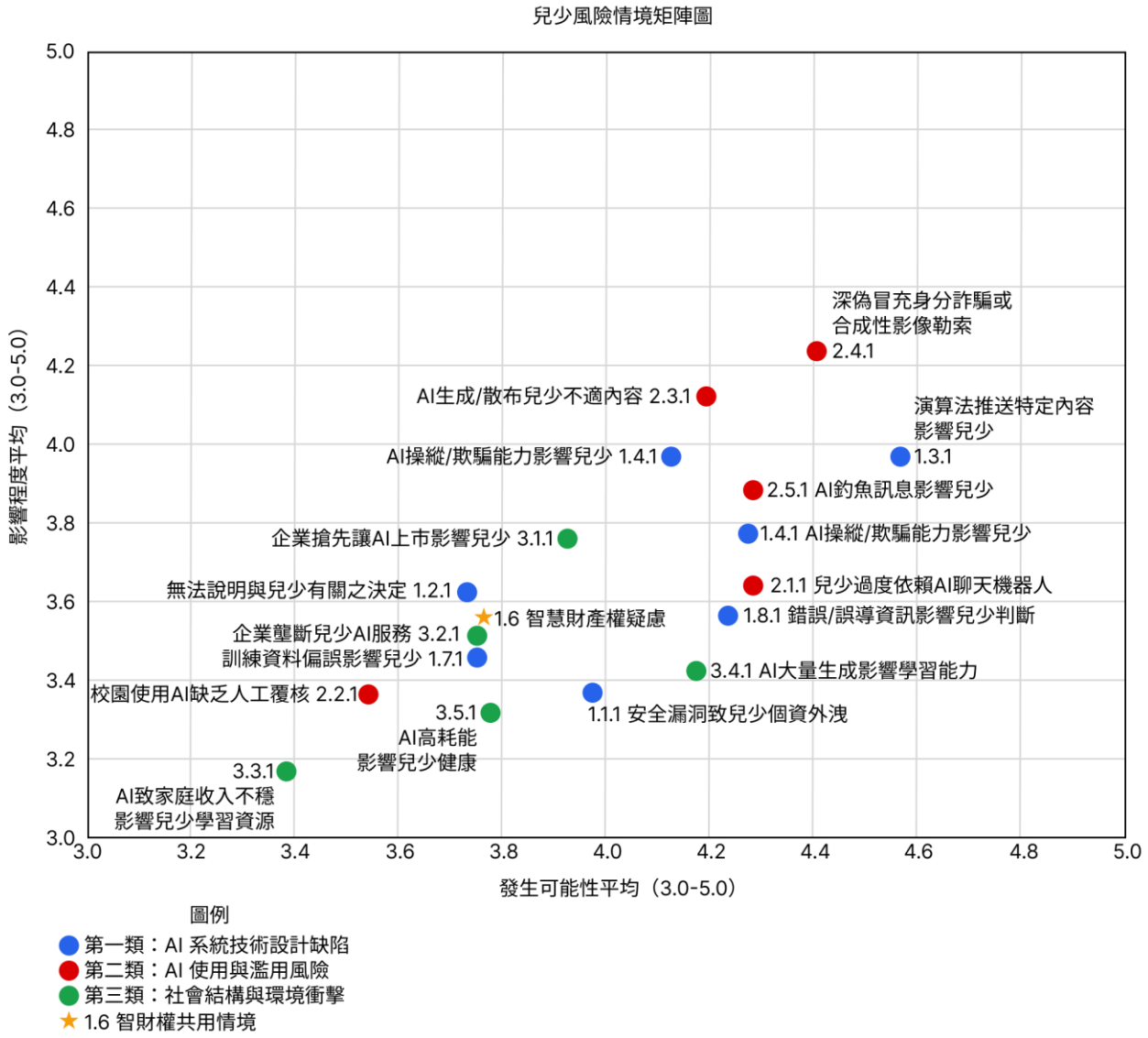

圖 1：兒少風險情境矩陣圖

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×I	補充說明/具體樣態
1	2.4.1 深偽冒充身分詐騙或合成性影像勒索	4.40	4.23	18.62	● 深偽工具變造影像並於網路散布 ● 語音合成工具冒充親友對兒少進行詐騙

排序	風險情境	可能性平均	影響度平均	風險值 L×I	補充說明/具體樣態
2	1.3.1 演算法推送特定內容影響兒少	4.56	3.96	18.06	● 推薦演算法使兒少持續接觸單一類型或不適齡內容 ● 短影音平臺依行為數據推送高黏著性素材強化沉迷
3	2.3.1 AI 生成/散布兒少不適內容	4.19	4.12	17.22	● 生成式工具產製暴力、色情或煽動性內容
4	2.5.1 AI 釣魚訊息影響兒少	4.28	3.88	16.61	● AI 模仿官方語氣誘使兒少提供帳號資訊或家庭支付工具 ● 遊戲平臺與社群媒體中高擬真詐騙訊息
5	1.4.1AI 操縱/欺騙能力影響兒少	4.12	3.96	16.33	● 情感陪伴型 AI 對兒少情緒狀態與行為判斷產生不當引導 ● 情感陪伴型 AI 未能辨識兒少心理危機徵兆並適時轉介
6	1.5.1 蒐集兒少資料影響隱私	4.27	3.77	16.09	● AI 系統蒐集兒少行為資料致個人偏好與使用軌跡遭不當存取 ● 穿戴式裝置或定位服務因安全設計不足
7	2.1.1 兒少過度依賴 AI 聊天機器人	4.28	3.64	15.58	● 兒少以 AI 聊天機器人替代人際互動或專業心理諮詢 ● AI 聊天機器人於自傷或自殺危機情境中無法提供適當回應
8	1.8.1 錯誤/誤導資訊影響兒少判斷	4.23	3.56	15.06	
9	3.1.1 企業搶先讓 AI 上市影響兒少	3.92	3.76	14.73	● 對話型服務上市前安全測試不足即開放使用

排序	風險情境	可能性平均	影響度平均	風險值 L×I	補充說明/具體樣態
					● AI 產品缺乏兒少保護設計即推出上市
10	3.4.1 AI 大量生成影響學習能力	4.17	3.42	14.24	
11	1.2.1 無法說明與兒少有關之決定	3.73	3.62	13.49	
12	1.1.1 安全漏洞致兒少個資外洩	3.97	3.37	13.39	● 學校採購 AI 系統缺乏一致安全檢核標準
13	1.6. 智慧財產權疑慮	3.76	3.56	13.39	
14	3.2.1 企業壟斷兒少 AI 服務	3.75	3.50	13.12	
15	1.7.1 訓練資料偏誤影響兒少	3.75	3.46	12.97	● 訓練資料未充分涵蓋多元樣本致偏鄉、原住民或新住民二代兒少遭不當標籤化 ● AI 演算法將申請人背景特徵納入風險評分致特定兒少家庭遭差別對待
16	3.5.1AI 高耗能影響兒少健康	3.77	3.32	12.52	
17	2.2.1 校園使用 AI 缺乏人工覆核	3.54	3.36	11.89	
18	3.3.1AI 致家庭收入不穩影響兒少學習資源	3.38	3.17	10.69	

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

以下就各情境之主要觀察重點進行說明，內容綜整自專家問卷及座談會意見。

(一)內容安全、深偽技術與數位暴力

在本次專家問卷與座談意見中，與內容安全相關之風險情境整體評分較高，專家共識度亦相對集中，顯示此面向為目前兒少接觸人工智慧情境中受關注程度較高之議題。

1、 深偽冒充身分詐騙或合成性影像勒索（2.4.1）

- 影像與語音合成技術之發展，使高度擬真內容之製作門檻持續降低。
- 相關技術已出現被不當使用之案例，包括使用一鍵脫衣(undressing)功能變造同儕影像並於網路散布，對被害兒少之人格權與心理健康產生負面影響。
- 語音合成技術可能被用於冒充親友進行詐騙，影響家庭信任關係。
- 深偽技術對兒少之風險具有即時性與擴散性，為本次評估中專家共識度最高之議題之一。

2、 演算法推送特定內容影響兒少（1.3.1）

- 演算法推薦機制在特定情境下，可能使兒少持續接觸單一類型或不適齡之內容，進而影響其心理狀態與認知發展。
- 專家對於實際影響程度之評估尚存不同意見，考量近期案例顯示風險有上升趨勢，顯示此議題仍需持續觀察與累積實證。

3、 AI 生成或散布兒少不適內容（2.3.1）

- 生成式人工智慧工具可能被用於產製暴力、色情或煽動性內容，並透過網路平臺快速流通。

- 實務上已出現利用生成式技術產製不當內容之情形，且因生成來源難以追溯，現行執法與責任歸屬機制面臨新的挑戰。
- 不適齡內容之風險不僅來自生成端，亦涉及平臺演算法之推播機制，平臺可能基於商業利益蒐集兒少使用行為資料，據以推送具高度黏著性之內容，使兒少持續暴露於不適齡素材之中，而相關責任歸屬亦較難釐清。
- 建議應從技術設計端導入安全防護機制，並強化內容檢舉與即時處理流程，以降低不適齡內容對兒少之影響。

(二)AI 詐騙與操縱風險

1、 AI 釣魚訊息影響兒少（2.5.1）

- 人工智慧技術使釣魚訊息之擬真程度與互動能力顯著提升，兒少在遊戲平臺、社群媒體及交友軟體等情境中，可能面臨更難辨識之詐騙風險。
- 人工智慧可模仿官方語氣或進行初步對話互動，誘使兒少提供帳號資訊或家庭支付工具，此類情形在實務中已有相當數量之案例。
- 專家意見亦提及，當兒少持續暴露於高擬真詐騙訊息環境中，可能影響其對數位環境之基本信任感。
- 相關因應方向包括平臺端之內容偵測與攔截機制、分齡使用設計，以及學校端之數位素養教育等。

2、 AI 操縱/欺騙能力影響兒少（1.4.1）

- ：部分情感陪伴型 AI 在互動過程中，可能對兒少產生情感引導或行為暗示之效果。

- 國際間已有案例顯示，情感陪伴型 AI 在缺乏適當安全機制之情況下，可能對兒少之情緒狀態與行為判斷產生不當影響，包括未能及時辨識使用者之心理危機徵兆。
- 建議針對兒少之情感互動型人工智慧應強化安全設計，包括關鍵情境之辨識與導引機制，以及適時提供專業諮詢資源之提示功能。

(三)認知依賴與學習能力

1、 兒少過度依賴 AI 聊天機器人（2.1.1）

- AI 聊天機器人之普及，使兒少可能將其作為情緒支持或日常陪伴之來源。
- 部分兒少已出現以人工智慧替代人際互動或專業諮詢之使用型態，可能影響其社交能力發展與適時獲得專業協助之機會；惟亦有觀察認為，AI 聊天機器人在特定情境下可為社交適應困難之兒少提供初步心理緩衝，顯示此議題影響面向具多元性。
- 兒童福利聯盟（2025）調查指出，有心理困擾且實際尋求協助之青少年中，46.5%傾向透過生成式人工智慧抒發情緒，高於向學校輔導求助者（41.1%）及心理健康專業人員者（30.4%），國際研究亦呈現相近走向。
- 人工智慧系統在涉及自傷或自殺之危機情境中是否能提供適當回應，仍存在重要疑慮，宜在後續政策研議中審慎評估。

2、 AI 大量生成影響學習能力（3.4.1）

- 生成式人工智慧工具協助完成學習任務之情形日益普遍，此現象之發生可能性甚高，惟對於長期學習能力之實質影響程度，目前仍存在不同評估。

- 當生成式工具大幅降低內容產出之門檻時，可能影響兒少之批判思考歷程與自主學習動機。相關討論方向包括教育評量方式之調整，例如強化對思考歷程之觀察，而非僅以最終成果作為評估依據，以引導兒少在善用工具之同時維持自主學習能力。

3、校園使用 AI 缺乏人工覆核（2.2.1）

- 人工智慧若被應用於學生行為監測、學習評量或資源配置等決策，可能涉及兒少之人格發展與受教權益。
- 教育現場之成人使用者對人工智慧運作方式與潛在風險之認識仍有落差，可能在未充分評估之情況下導入相關系統。
- 國際間已有將人工智慧用於即時監測學生課堂行為之案例，引發對兒少隱私權與人格自主發展之討論。
- 建議若有涉及兒少重大權益之決策，人工智慧應定位為輔助參考工具，最終判斷仍應由具備專業素養之人員行使。

(四)隱私、資料保護與定位追蹤

1、蒐集兒少資料影響隱私（1.5.1）

- 人工智慧系統之運作多涉及資料蒐集與分析，兒少可能因風險辨識能力尚在發展階段而過度揭露個人資訊。
- 穿戴式裝置或定位服務若安全設計不足，可能衍生兒少行蹤遭不當存取之風險。
- 相關可能研議方向包括資料蒐集最小化、法定代理人同意機制及敏感資料之加強型保護措施等。

2、安全漏洞致兒少個資外洩（1.1.1）

- 在教育或社會服務場域導入人工智慧應用時，系統安全性為重要考量。
- 政府機關或學校透過採購引進相關應用，若缺乏一致性之安全檢核標準，可能增加資料外洩風險。
- 兒少保護及身心健康輔導等領域所涉資料具高度敏感性，未來若導入人工智慧進行分析或輔助判斷，其存取權限、傳輸安全與使用範圍均需配合既有個資保護法制妥適規劃。

(五)訓練資料偏誤影響兒少（1.7.1）

- 人工智慧系統之判斷品質與訓練資料之代表性密切相關。
- 若訓練資料未充分涵蓋不同地區、族群及學習特質之樣本，偏鄉、原住民、新住民二代或具特殊學習需求等具交織身分之處境不利兒少，可能在系統判定中被低估或不當標籤化。
- 國際間亦有公部門演算法因將申請人背景特徵納入風險評分，導致特定家庭遭受差別對待之案例。
- 相關因應方向包括訓練資料之多元代表性審核、跨文化偏誤測試，以及演算法決策之申訴管道等。

(六)無法說明與兒少有關之決定（1.2.1）

- 在涉及兒少權益之決策情境中，人工智慧系統之判斷邏輯若缺乏可解釋性，可能影響當事人及其照顧者對結果之理解與信任。
- 國際間已有將人工智慧用於入學評量或福利資格審查後，因演算法產生非預期之差別效果而引發爭議之案例。
- 專家普遍建議，涉及兒少重大權益之決策應保留人工覆核機制，避免全自動化判斷，並建立兒少友善之申訴管道與資訊揭露方式，以維護程序保障與權益救濟之基本要求。

(七)錯誤/誤導資訊影響兒少判斷（1.8.1）

- 生成式人工智慧存在產出不正確或誤導性內容之技術特性，對正處於認知發展階段之兒少，其影響可能較成人更為顯著。
- 目前兒少主動查證人工智慧所提供資訊之比例仍偏低，實務上已有因誤信生成內容而產生不利後果之情形。
- 相關因應方向包括在醫療、心理諮詢等對兒少影響較為直接之應用領域採取限制用途設計、教育場域所使用之工具強化內容來源管理，以及持續推動兒少之資訊查證素養教育等。

(八)市場結構與競爭秩序對兒少權益之影響

1、 企業搶先讓 AI 上市影響兒少（3.1.1）

- 人工智慧產品快速推出之市場競爭環境下，部分服務在安全防護機制尚未完備時即開放兒少使用，可能衍生內容安全與互動風險。
- 國際間已有案例顯示，對話型服務因上市前安全測試不足，產生對兒少不適當之回應內容。
- 可參考既有產品安全管理思維，就涉及兒少之人工智慧服務，研議於上市前納入安全評估之可行機制。

2、 企業壟斷兒少 AI 服務（3.2.1）

- 在教育場域中，若少數大型平臺成為數位學習之主要基礎設施，學習資料集中化之趨勢可能伴隨隱私風險與系統轉換困難等議題。
- 當演算法深度介入學習內容之篩選與評量時，其對兒少資訊接收與價值判斷之影響宜持續關注。
- 相關討論方向包括本土數位學習資源之多元發展、資料可攜性之促進，以及平臺治理透明度之提升等。

(九)其他觀察

1、 智慧財產權疑慮（1.6）

- 兒少使用生成式工具完成學習任務時，可能涉及學術倫理與原創性之討論。
- 相關因應方向包括推動人工智慧倫理創作之素養教育，以及生成內容之溯源標示機制等。

2、 AI 高耗能影響兒少健康（3.5.1）：

- 人工智慧模型訓練與運算之能源消耗議題逐漸受到國際關注。
- 考量我國半導體與高效能運算產業之發展現況，相關環境影響亦為長期觀察面向之一。
- 國際間已有要求模型開發者揭露能源消耗與碳排放之討論趨勢，可作為未來政策研議之參考。

3、 AI 致家庭收入不穩影響兒少學習資源（3.3.1）

- 人工智慧對就業結構之影響，可能透過家庭經濟條件之變動間接波及兒少之教育資源與發展機會。
- 此影響路徑較為間接，惟從世代之間的流動與社會公平之角度，弱勢家庭所受衝擊仍值得持續關注。

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；二是綜整本次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)持續落實跨既有法規機制

政府各機關目前已針對兒少網路安全與數位學習領域，初步整合既有法規機制，亦用以因應人工智慧技術之導入。

1、教育應用：教育部已發布「數位教學指引」及相關使用規範，導引第一線師生在校園情境中正確且安全地應用生成式 AI 工具，並持續精進資訊素養教育。

2、內容安全保障：透過國家通訊傳播委員會、數位發展部、衛生福利部、內政部、文化部與教育部之協作，利用 iWIN 網路內容防護機構進行爭議內容之受理，並依《兒童及少年福利與權益保障法》及《兒童及少年性剝削防制條例》等相關法規，針對涉及 AI 變造之不當影像落實即時下架與移除防範。

(二)依據人工智慧基本法建構風險治理基礎

1、依據《人工智慧基本法》第 5 條，人工智慧應用於兒少相關情境且經評估具較高風險者，應明確標示注意事項或警語。

2、數位發展部正研議「人工智慧風險分類框架」草案，作為各部會辨識 AI 系統風險之共同參據，確保執行兒少保障業務時具備一致基礎。

3、數位發展部推動 AI 產品與系統評測制度，刻正納入兒少相關之評測項目，以期符合兒少最佳利益。

(三)落實公私協力與國際接軌

1、依現行「兒童及少年福利與權益推動小組」機制，持續推動兒童及少年福利與權益保障政策，人工智慧相關議題得於既有架構下納入討論與觀察。

2、 數位發展部與大型國際平臺業者（如 Meta、Google 等）建有常態化溝通窗口，就重要 AI 治理議題持續進行對話與意見交流，未來也將持續就兒少保護措施與業者深化溝通。

3、 就國際發展而言，OECD 及聯合國等國際組織持續就兒少 AI 權利保護提出相關建議，各主要經濟體亦逐步將兒少保護納入 AI 治理規範之核心考量。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)建構涵蓋兒少之 AI 應用管理機制

1、 專家普遍建議應將針對兒少之特定 AI 應用（如情感陪伴型聊天 AI、深偽影像生成工具、自動化教育評量系統）列入後續管理對象。

2、 相關機關可研議透過跨部會及跨專業合作機制，結合教育、社會福利與諮商輔導等助人專業之實務經驗，針對人工智慧系統在涉及兒少心理健康或危機情境下之互動模式，進行系統性之風險評估與實證研究，作為後續政策研議之基礎。

(二)強化數位內容環境安全與平臺法遵透明度

針對生成式人工智慧與推薦演算法帶來之內容風險，綜合專家問卷與座談意見，有以下建議可做為未來研議方向，包括：

- 在平臺設計面，評估未成年帳號預設關閉基於行為軌跡之個人化推薦之可行性，回歸通識內容或年齡分層內容。
- 研議將「最大化用戶參與」列為針對兒少之演算法設計限制事項。
- 在內容處理面，研議建立涉及兒少不適內容之跨平臺快速下架協作流程，並參考國際趨勢評估要求 AI 生成內容標示識別標記之機制設計。
- 研議整合現行申訴管道，提供被害人更便捷之一站式通報與緊急下架服務，並強化金融與電信體系之跨域防詐協作。

(三)建立兒少友善之申訴救濟與程序保障機制

1、 專家問卷中幾乎每一項風險情境之「建議因應方式」均提及建立申訴機制之必要性，反映此為當前治理缺口。

2、 專家建議

- 建立專門處理「AI 歧視／錯誤決策」之申訴管道，提供兒少友善且簡單可理解之透明資訊與控制選項。
- 涉及兒少重大權益之 AI 決策，相關機關可研議保留人工覆核機制之可行性，避免全自動化判斷。
- 建立可及之救濟途徑，包括資料刪除請求權與損害賠償機制。

(四)深化兒少數位素養、親職教育與 AI 識讀能力

1、 在隱私與資料治理面向，透過既有教育與宣導機制，持續推動兒少對 AI 使用風險、個資保護及數位安全之認識。

2、 多位專家特別強調親職教育之必要性，指出家長與教育現場之 AI 素養落差可能成為風險放大因素，建議透過「情感識讀」、「AI 偏見識讀」與「深偽識讀」等課程模組，提升兒少、家長與教師之判斷能力。

(五)推動弱勢或特定群體兒少之數位包容與反偏誤機制

針對 AI 訓練資料偏誤可能對弱勢或特定群體兒少造成之差別衝擊，相關機關可研議：

- 要求教育用 AI 產品提供演算法公平性說明之可行性，並推動訓練資料之多元代表性審核，納入多元文化樣本；
- 可將高品質 AI 學習工具列為公共財政策方向，確保家庭經濟弱勢之兒少仍能獲得基礎數位素養資源。

(六)強化兒少資料保護之具體技術規範

專家建議對兒少資料採「預設不蒐集」原則，針對涉及兒少敏感資料之 AI 系統，相關機關可研議納入資安稽核、滲透測試與通報義務等強化措施，作為後續政策研議之參考方向。

陸、 結語

基於前揭背景說明、方法架構、應用場景盤點與風險觀察，本次「人工智慧應用發展兒少影響評估」係依《人工智慧基本法》附帶決議於短期內完成之階段性成果，重點在於彙整跨部會觀點與蒐集專家意見，並初步辨識我國兒少在人工智慧發展下可能面臨之主要影響面向。

在快速變動之科技環境下，政府將持續以協作治理為原則，透過跨部會合作、社會對話與國際經驗交流，逐步累積在地治理知識與實務經驗，並依各主管機關權責推動相關研議工作。配合附帶決議所提及之常設性人工智慧應用影響評估機制之研議方向，未來亦可在既有評估經驗基礎上，持續觀察如何透過制度化之評估流程，強化風險辨識與政策回饋功能，並加入利害關係人溝通，以作為後續治理精進之參考。

本報告之提出，象徵我國在人工智慧治理中納入兒少觀點之起點。隨著技術演進與社會意見之累積，相關影響評估與治理方向仍將持續深化，以促進人工智慧之創新發展，同時兼顧兒少權益保障與社會整體福祉。

附錄：AI 影響評估：意見調查問卷

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 作業批改平台）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(如線上學習平台的 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（如平台的動態定價 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、：遊戲平台的 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、交友平台的配對演算法、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：線上教育平台的 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：租屋平台的 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/遊戲平台通知），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**