主題二：AI 應用發展人權影響評估報告

壹、 背景

隨著人工智慧技術快速發展，其應用已逐步導入公共治理、就業媒合、金融風險控管、勞動管理及數位平臺內容推薦等多元場域。人工智慧在提升行政效率與產業創新動能之同時，亦與平等權、隱私權、程序保障、勞動權及表達自由等基本權利密切相關，如何在促進創新發展的同時持續強化基本權利之保障，已成為我國及各主要法域共同關注之政策議題。

《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「人權影響評估」等，並將評估報告公開發布，爰本「人權影響評估」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應用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、隱私保護及可問責等原則進行整理。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；憲法基本權利保障體系及《個人資料保護法》、《就業服務法》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。國際間，歐盟、美國等主要法域已逐步將風險管理與基本權保障納入人工智慧治理架構，相關發展可作為我國持續觀察之參照。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，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 AI 使用情境具高度多樣

性，本次採取國際資料蒐集、問卷調查、專家意見諮詢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單位及專家提供意見，並透過檢視會議確認內容，辦理期程說明如下：

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	請相關部會協助檢視表單架構及提供意見，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，邀集相關單位討論，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	邀集學術界、法律實務界、公民團體及人權組織代表與會，並有行政院人權及轉型正義處、監察院國家人權委員會等機關代表出席。就自動化決策、演算法差別影響、勞動場域 AI 應用及資料跨境治理等議題提出觀察。
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，就民眾可能接觸之人工智慧應用場景，針對平等權保障、隱私與資料自主、程序正當性與可問責性，以及勞動權與社會結構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，應結合質化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，兼顧公共治理、社會服務、勞動管理、資料治理及平臺運作等面向，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）：

召開會議邀集學術界、法律實務界、公民團體及人權組織代表與會，並有行政院人權及轉型正義處、監察院國家人權委員會等機關代表出席，就自動化決策、演算法差別影響、勞動場域 AI 應用及資料跨境治理等議題提出觀察。

(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與公共服務，民眾接觸人工智慧之情境已涵蓋公共行政、就業媒合、金融服務、數位平臺、內容創作及社會福利等多元面向。本節係依前述評估方法，綜整專家座談、利害關係人意見、學者就歐盟人工智慧法中高風險領域所提供之分析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：人權相關

場景類別	具體應用舉例
一、公部門 AI 輔助決策涉及人權	■ AI 輔助社會福利資格審核與補助優先排序（如低收入戶認定、身心障礙評估） ■ 政府機關運用 AI 進行行政裁罰輔助 ■ 公共資源分配之 AI 輔助決策（如住宅補貼、醫療優先序）

	■ AI 協助司法或執法場域之人臉辨識、行為預測分析
二、就業媒合、勞動管理與職場決策	■ AI 履歷篩選與評分系統（依行為資料或歷史紀錄自動排序求職者） ■ 職場 AI 監測工具（出勤追蹤、通訊紀錄分析、生產力評估） ■ AI 績效評估或晉升推薦演算法 ■ 企業導入 AI 自動化取代特定職務（如客服、資料處理、物流分揀） ■ 薪資定價或工時調配之演算法決策（如零工平臺派單機制）
三、數位平臺、演算法推薦與內容傳播	■ 社群平臺個人化內容推薦演算法（如新聞動態、短影音「為你推薦」機制） ■ 搜尋引擎與資訊聚合平臺之排序演算法 ■ 廣告定向投放 AI（依使用者行為、族群特徵進行精準推播）
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、個人資料蒐集、隱私保護與資訊安全	■ AI 系統於個人資料儲存、處理與傳輸過程中之安全防護應用 ■ 健康管理 App、金融客服 AI 蒐集與分析高敏感個資 ■ AI 輔助身分驗證與生物特徵辨識（人臉、聲紋、步態）
六、金融、信用與商業服務	■ 信貸核准或額度決策 AI（依演算法評分自動審核） ■ 保險商品定價與核保之 AI 輔助評估 ■ 投資理財或消費決策之 AI 建議工具 ■ 租屋或商業交易平臺之 AI 信用或風險評分 ■ AI 生成個人化訊息於商業行銷、客戶互動及業務推廣之應用
七、社會結構、市場競爭與環境影響	■ AI 產品與服務之市場化開發與部署 ■ AI 自動化技術於製造、物流及服務業之導入與職務替代 ■ AI 訓練與大規模運算基礎設施之建置與能源消耗
八、執法、刑事司法與風險預測	■ 預測性警務系統（以歷史犯罪資料預測高風險區域或個人） ■ 再犯風險評估工具輔助司法裁量（如量刑、假釋審查） ■ AI 輔助人臉辨識應用於犯罪偵查或嫌疑人比對 ■ 法院或檢察機關使用 AI 進行案件分類或優先排序
九、移民、庇護審查與邊境管理（國際參考場景）	■ AI 輔助移民或庇護申請之風險分類與資格審查 ■ 邊境管制之 AI 行為偵測或意圖分析系統 ■ 依國籍、語言或行為模式進行自動化風險評分 ■ AI 輔助入境查驗與身分比對系統

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 AI 應用之人權可能風險，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，顯示人權面臨之 AI 風險普遍屬中高水準。
- 落於右上象限之情境（2.5.3、2.3.3、2.4.3、1.8.3、2.1.3、2.2.3、3.1.3、1.2.3、3.2.3）兼具較高發生可能性與影響程度，可為優先關注對象。

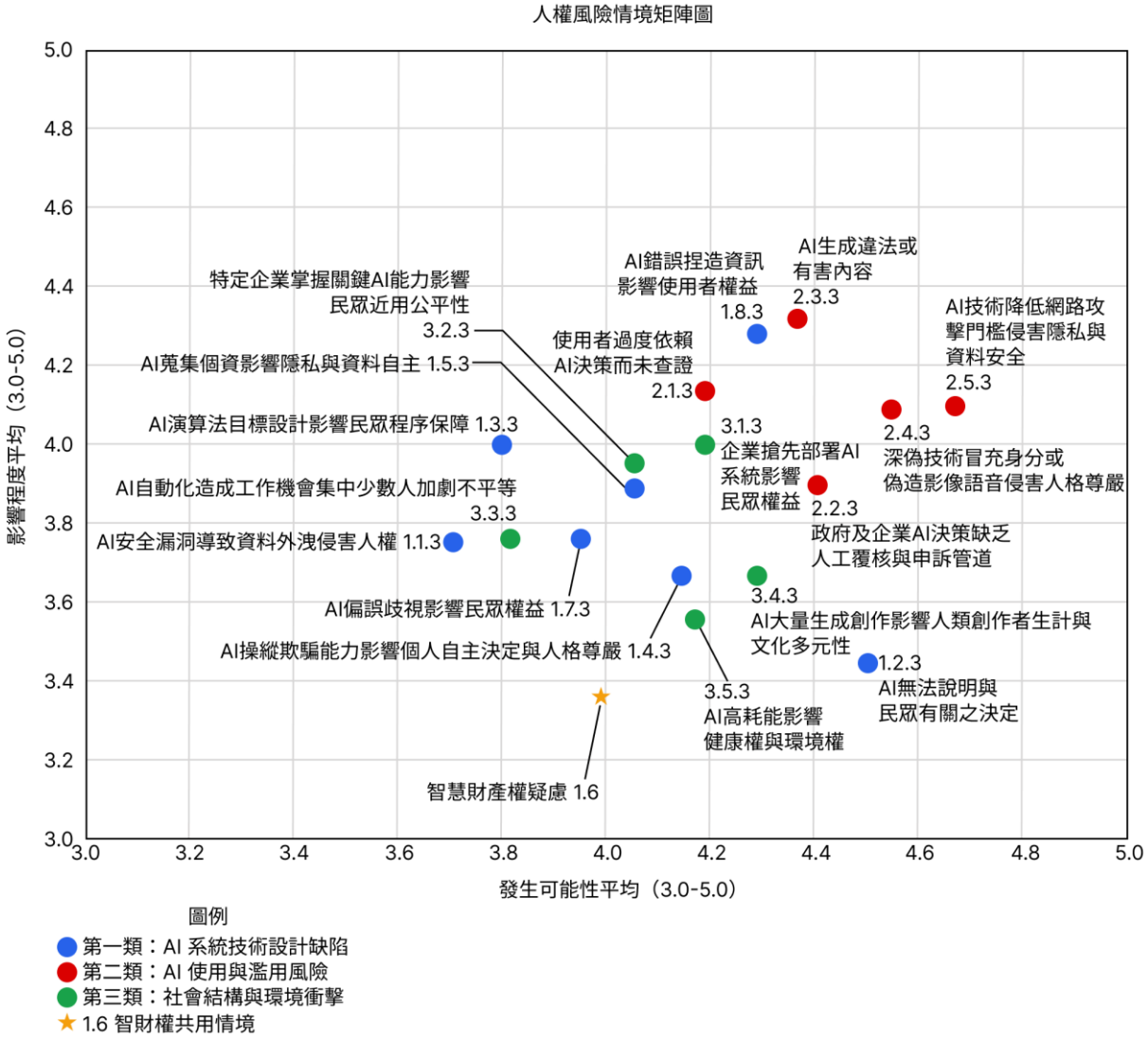

圖 1：人權風險情境矩陣圖

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	2.5.3 AI 技術降低網路攻擊門檻侵害隱私與資料安全	4.67	4.10	19.11	● AI 自動產製釣魚郵件、仿冒網站或惡意程式碼 ● 網路攻擊門檻降低
2	2.3.3 AI 生成違法或有害內容	4.39	4.32	18.96	● 生成式 AI 產製仇恨言論、歧視內容或詐騙話術 ● 演算法推播加速有害內容擴散
3	2.4.3 深偽技術冒充身分或偽造影像語音侵害人格尊嚴	4.55	4.09	18.60	● 深偽技術未經同意合成他人影像或聲音 ● 用於身分冒充、詐騙或散布不實內容
4	1.8.3 AI 錯誤捏造資訊影響使用者權益	4.29	4.29	18.37	
5	2.1.3 使用者過度依賴 AI 決策而未查證	4.19	4.14	17.36	
6	2.2.3 政府及企業 AI 決策缺乏人工覆核與申訴管道	4.40	3.90	17.16	
7	3.1.3 企業搶先部署 AI 系統影響民眾權益	4.19	4.00	16.76	
8	3.2.3 特定企業掌握關鍵 AI 能力影響民眾近用公平性	4.05	3.95	16.00	● 少數企業掌控關鍵算力、資料與主流 AI 服務 ● 民眾缺乏替代選項，身心障礙者及偏鄉民眾受衝擊尤大

9	1.5.3 AI 蒐集個資影響隱私與資料自主	4.05	3.90	15.81	
10	3.4.3 AI 大量生成創作影響人類創作者生計與文化多元性	4.29	3.67	15.71	● AI 大量生成創作內容，人類創作者作品被低價取代 ● 內容同質化，創作生計與文化多元性受影響
11	1.2.3 AI 無法說明與民眾有關之決定	4.50	3.45	15.53	● AI 以不透明方式作成審核、貸款或資源分配決定 ● 當事人無從得知決策依據，尋求救濟困難
12	1.3.3 AI 演算法目標設計影響民眾程序保障	3.80	4.00	15.20	● AI 以效率、績效為目標設計，可能犧牲公平性 ● 推薦機制以最大化停留時間為目標
13	1.4.3 AI 操縱欺騙能力影響個人自主決定與人格尊嚴	4.14	3.67	15.19	● AI 具備操縱或說服能力，被部署於投資建議或社福說明場域 ● 影響個人自主決定，一般使用者難以識別，救濟困難度高
14	1.7.3 AI 偏誤歧視影響民眾權益	3.95	3.76	14.87	● AI 訓練資料偏誤導致特定族群誤判率或拒絕率偏高 ● 演算法不透明，偏誤難被察覺，救濟路徑相對困難
15	3.5.3 AI 高耗能影響健康權與環境權	4.16	3.56	14.78	● AI 訓練高耗能造成碳排放與能源分配壓力
16	3.3.3 AI 自動化造成工作機會集中少數人加劇不平等	3.81	3.76	14.33	

17	1.1.3 AI 安全漏洞導致資料外洩侵害人權	3.70	3.75	13.88	● AI 系統遭駭入或資料遭竄改外洩，引發錯誤決策或不當處置
18	1.6 智慧財產權疑慮	3.97	3.37	13.39	● AI 訓練使用他人著作，涉及著作權、專利及商業秘密疑慮

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)深偽技術、違法內容生成與網路攻擊

本次專家問卷中，與 AI 生成有害內容及網路攻擊相關之三項情境風險值包辦前三名，整體評分最高，專家共識度亦相對集中，顯示此面向為目前人權領域中受關注程度最高之風險叢集。

1、 AI 技術降低網路攻擊門檻侵害隱私與資料安全（2.5.3）

- 為本次問卷風險值最高項目。
- AI 技術可自動化產製釣魚郵件、仿冒網站內容及惡意程式碼，使網路攻擊之門檻持續降低、規模持續擴大，對個人之隱私權、財產安全及資訊自主造成直接影響。

2、 AI 生成違法或有害內容（2.3.3）

- 影響程度為全部情境中最高，反映專家對此類風險一旦發生所可能造成影響之高度關切。
- 生成式 AI 工具可能被用於產製煽動仇恨、歧視言論或詐騙話術等有害內容，並透過網路平臺快速流通，對言論自由保障邊界、內容平臺問責機制及特定群體之人身安全形成新的挑戰。
- 處境不利群體（例如：身心障礙者、原住民族及老人）因數位資訊識讀能力差異，面對 AI 生成有害內容之抵禦能力相對較弱，面臨風險更為複雜之情形。
- 此類風險不限於內容「產製」面向，演算法「推播」亦構成重要風險路徑，有害資訊得以廣泛流通，惟平臺與實際創作者之間的責任歸屬仍待研議。

3、 深偽技術冒充身分或偽造影像語音侵害人格尊嚴（2.4.3）：

- 發生可能性為全部情境中第二高，顯示專家認為深偽技術之應用已相當普遍。
- 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，其應用於未經當事人同意之影像合成、聲音仿冒及身分偽造，對個人之人格尊嚴、名譽及自主決定權形成直接影響。
- 處境不利群體(如身心障礙者或偏鄉民眾)在遭遇深偽攻擊後，可能因缺乏資源及管道而面臨較高之救濟困難。

(二)錯誤資訊、過度信任與決策缺乏覆核

1、 AI 錯誤捏造資訊影響使用者權益（1.8.3）：

- AI 系統在提供法律建議、醫療資訊或公共事務說明時，若產出錯誤或捏造之資訊，可能直接影響使用者之決策判斷。
- 目前生成式 AI 工具在提供具法律效力或攸關民眾重要權益之資訊時，問責機制尚不明確。

2、使用者過度依賴 AI 決策而未查證（2.1.3）：

- 使用者在投資、健康管理或公共服務申請等高影響決策場域中，若過度依賴 AI 建議而未進行必要查證，可能對其財產安全及基本權利之實現造成影響。
- 部分使用者未必能充分理解 AI 系統之侷限性，對資訊識讀能力較弱或高度依賴數位服務之族群，其風險尤為明顯。

3、政府及企業 AI 決策缺乏人工覆核與申訴管道（2.2.3）：

- 政府或企業在福利審核、信用評分或帳號停權等攸關民眾重要權益之決策中，若完全依賴 AI 自動判斷而未設置人工覆核程序，可能影響當事人受正當程序保障之權利。
- 專家以荷蘭育兒津貼演算法及丹麥 AI 社福審查為例，偏誤演算法在缺乏人工介入的情況下，可能對特定族群（包括移民家庭或低收入戶）形成系統性差別影響，具有制度觀察之參考意義。
- 部分決策雖聲稱最終決定仍由人工作成，然演算法標記後實際上幾乎等同定論，人工覆核之實質功能有待釐清。

(三)個人資料蒐集、隱私保護與資訊安全

1、AI 蒐集個資影響隱私與資料自主（1.5.3）：

- 例如金融 App 生成式客服、健康管理工具及行為追蹤演算法等，可能透過持續蒐集與分析使用者資料，影響個人之資訊自主與隱私權。
- 境外平臺之資料處理實務難以透過現行《個人資料保護法》進行有效管轄，且部分平臺之「使用者條款」要求用戶同意將資料授權予第三方，否則即無法使用服務，實質上壓縮個人資料

自主之選擇空間。案例顯示現行個資保護規範方面，仍有持續研議之必要。

2、AI 安全漏洞導致資料外洩侵害人權（1.1.3）：

- AI 系統之安全漏洞若導致社會服務、醫療或公共行政資料外洩，可能引發錯誤決策或不當處置，進而侵害當事人之隱私權、資訊安全與人格尊嚴。
- 涉及公部門社政、衛政或司法領域之高敏感資料一旦外洩，對處境不利群體之影響尤為直接。

(四)演算法偏誤、歧視與可解釋性缺乏

1、AI 偏誤歧視影響民眾權益（1.7.3）：

- 提供服務或作出決策之 AI 系統若因訓練資料偏誤，對特定族群（包括身心障礙者、原住民族或特定社經背景民眾）產生差別影響，將直接涉及憲法平等權保障之核心。
- AI 系統在設計階段若未充分考量訓練資料之代表性，偏誤將在部署後持續放大，且因演算法不透明而難以被當事人察覺，救濟路徑亦相對困難。
- 歧視可能分散於大量個案之中，較難被察覺與舉證，與傳統歧視的本質差異。

2、AI 無法說明與民眾有關之決定（1.2.3）：

- AI 系統在作出與民眾權益有關之決定（如福利審核、貸款核准或資源分配）時，若無法清楚說明決策依據，將影響當事人瞭解決定原因之權利，進而影響其尋求救濟之能力。
- 可解釋性缺乏對人權保障之影響具有結構性意涵，當決策影響面愈廣，資訊不對稱所形成之不利後果亦愈顯著。

3、AI 演算法目標設計影響民眾程序保障（1.3.3）：

- AI 系統為追求績效目標（如加快篩選速度、減少人工作業或精準推播），可能在結果上對民眾之程序保障或公平近用造成影響。
- 現行推薦機制在設計上以最大化使用者停留時間為目標，可能與民眾知情權及認知自主之保障存在潛在張力。

(五)AI 操縱欺騙能力影響個人自主決定與人格尊嚴（1.4.3）

- 本情境為全部風險情境中少數「影響程度高於發生可能性」之情境之一，顯示專家對其潛在嚴重性抱有高度警戒，視為「若發生將造成深遠人權影響」之潛伏型高衝擊風險。
- AI 系統若經強化訓練具備操縱、說服或欺騙能力，並被部署於影響個人重要決策之情境（如投資建議、社福資格說明或法律資訊提供），對個人之自主決定、人身安全及人格尊嚴之影響將尤為嚴峻。
- AI 生成內容若被刻意用於操控仇恨情緒、散布歧視言論或強化不實資訊，將對公眾言論空間之真實性造成影響，進而侵蝕表意自由與公民社會之正常運作。

(六)市場競爭、企業壟斷與治理缺口

1、企業搶先部署 AI 系統影響民眾權益（3.1.3）：

- 企業在 AI 產品未完善安全測試或保護機制之情況下搶先上市，可能使民眾在高影響決策場域中暴露於未充分驗證之系統風險。

2、特定企業掌握關鍵 AI 能力影響民眾近用公平性（3.2.3）

- 少數科技企業掌控關鍵算力、資料基礎設施及主流 AI 服務，可能影響民眾對資訊服務之近用公平性，並在市場發生重大變動時對依賴相關服務之民眾形成較大衝擊。
- 目前缺乏針對 AI 服務市場集中度之監理機制，為制度觀察面向之一。

(七)其他觀察

1、AI 自動化造成工作機會集中少數人加劇不平等（3.3.3）

- 本情境風險值屬本次評估中相對居後之情境，惟專家提出此議題具有長期結構性意涵。
- AI 自動化加速工作機會與薪資集中於少數技術勞動者，可能使既有社會經濟不平等進一步擴大，影響廣大民眾之工作權與生活水準。

2、AI 大量生成創作影響人類創作者生計與文化多元性（3.4.3）

- 本情境發生可能性與影響程度落差為全部情境中最大。
- 此一結構顯示，AI 大量生成創作之現象已相當普遍，但其對人民原創空間及文化多元性之實質損害，目前仍處於發展初期而尚難評估，有待持續觀察。

3、AI 高耗能影響健康權與環境權（3.5.3）

- 本情境風險值在本次評估中相對居後。
- 我國因具備全球領先之高效能運算產業聚落，AI 訓練與運算之高耗能特性對本地能源分配、碳排放及環境品質之影響，值得持續關注。
- 專家提及，建議政府未來可關注國際間是否及如何者建立業者之環境足跡揭露義務，作為後續政策研議之參照。

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，具體措施將隨實務發展由各主管機關依其權責持續推進與滾動調整。

一、現行作法

(一)依既有法制保障基本權利

1、 依現行法制保障 AI 應用涉及之基本權利：

人工智慧應用涉及之平等權、隱私權、程序保障與勞動權等基本權利，現行已有相關法制提供規範基礎，包括個人資料保護法、勞動法制及行政程序法等。人工智慧應用如涉及特定領域事項，仍應依各該法規辦理，並接受既有監督與救濟機制之檢驗。

2、 銜接已國內法化人權公約落實機制檢討因應 AI 人權議題

- 有關 AI 應用涉及之人權議題，倘涉及已國內法化人權公約之落實及相關政策、法規之檢討，其屬跨部會事項者，行政院設有人權相關任務編組，以統合協調各機關推動人權保障工作，各部會亦設有人權工作小組，得適時循公私協力及民間參與討論之方式，就應用人工智慧衍生之人權侵害疑慮予以討論應處。
- 後續將視人工智慧應用涉及之權利項目、處境不利群體、業務範疇，由各部會人權工作小組或行政院相關人權任務編組等既有機制下討論處理。

(二)落實《人工智慧基本法》治理原則

1、 依《人工智慧基本法》第 2 條第 2 項及第 5 條第 1 項規定，政府應避免人工智慧之應用侵害人民之生命、身體、自由或財產，各相關事項並由各目的事業主管機關依其職掌分別辦理。

2、 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，各機關於規劃或評估人工智慧應用時，將逐步納入風險辨識與影響觀察機制。

3、 數位發展部亦正研議「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，確保各機關執行人權相關業務時具備一致之評估基礎。

(三)落實跨部會協作並參考國際多元治理模式

1、 本次影響評估作業涵蓋跨部會協調整合、專家座談及問卷意見蒐集等程序，初步建立跨域對話基礎。相關過程顯示，人工智慧治理涉及技術、法律與社會等多元面向，難以由單一機關獨力因應，有賴各目的事業主管機關依其業務職掌持續協作，共同強化風險辨識能力與制度回應彈性。

2、 就國際發展而言，美國、OECD、歐盟等主要經濟體雖治理路徑不盡相同，惟均呈現監理規範、產業自律與多方參與並行之趨勢。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)持續關注 AI 輔助決策之程序保障面向

1、本次評估中，涉及政府或企業以 AI 輔助作成攸關民眾重要權益之決定（如福利審核、信用評分及資源分配）之情境，受到與會專家及問卷填答者之高度關注。

2、學者觀察

- AI 人權風險往往並非源自單一技術失誤，而是來自多重交織因素，包括訓練資料反映既有社會不平等、系統設計階段之變數選擇與最佳化目標本身隱含效率優先之價值判斷、人工覆核流於形式而未具實質審查功能，以及受影響者因資訊與資源不對等而難以理解或質疑決策結果。
- 上述因素相互強化，使得影響往往難以在個案層次被察覺，卻可能在制度層次造成持續性的不利後果。

3、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，就涉及基本權利保障之高影響決策情境，研議程序透明度與人工判斷空間之適當設計，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高敏感資料之保護

1、鑑於公部門所掌握之社政、衛政及司法相關資料（包括身心狀況、家庭環境及社會服務紀錄）具有高度敏感性，若未來導入 AI 系統進行分析或服務提供，其資料安全要求宜在系統設計階段即予考量。

2、有意見指出跨境資料傳輸與境外平臺之資料處理實務，難以透過現行《個人資料保護法》進行有效管轄，顯示資料治理存在跨域落差。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，就既有資料治理框架之適用範圍持續研議。

(三)研議 AI 系統上市前合規觀察之制度方向

1、調查意見中指出，許多 AI 系統之人權風險並非僅源自技術設計本身，而高度取決於使用單位如何導入、設定與解釋系統結果

2、專家認為，將權利保障機制前置於技術開發與制度設計階段，具有預防性意義。透過在模型設計時即進行差別影響分析、訓練資料代表性檢視與風險情境模擬，可在系統部署前辨識潛在不利影響，避免偏誤在大規模部署後才被察覺而難以矯正。此一觀察對後續研議特定類型 AI 應用之事前評估方向具有參考意義，相關討論可透過跨部會討論持續研議。

(四)持續關注國際治理發展趨勢與處境不利群體之觀察

1、國際間已逐步將差別影響辨識、偏誤測試與可問責原則納入人工智慧重點關注議題。

2、本次評估亦觀察到，身心障礙者、原住民族、老人及偏鄉民眾等處境不利群體，在面對 AI 應用之潛在風險時，可能因資訊近用能力差異而處於相對脆弱之情境。

3、未來將持續蒐集主要國家與國際組織之制度發展經驗，並結合我國法制環境與多元社會脈絡進行評估，以利後續政策研議與制度精進之參考。

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「人工智慧應用發展人權影響評估」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，並初步辨識人工智慧發展與應用可能涉及之基本權利影響面向。

人工智慧技術發展日新月異，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。配合附帶決議所提常設性影響評估機制之研議方向，未來亦可在本次評估基礎上，持續觀察人工智慧應用與基本權利保障之交會情境，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。本次影響評估報告之提出，係我國在人工智慧治理中納入人權觀點之初步起點，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，同時兼顧基本權利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**