

立法院第 11 屆第 3 會期  
教育及文化委員會、交通委員會

制定「人工智慧基本法草案」  
聯席公聽會  
書面報告

中央研究院

中華民國 114 年 4 月 16 日

主席、各位委員女士、先生：

感謝各位委員對本院各項業務的協助與指教，承蒙貴委員會邀請列席制定「人工智慧基本法草案」聯席公聽會，謹就生成式人工智慧風險研究，提出簡要報告：

因應全球席捲而來之人工智慧技術發展及生成式人工智慧可能帶來之風險問題，本院於 112 年 11 月 1 日成立「生成式 AI 風險研究小組」，邀集院內外跨領域的專家學者共同組成，主要針對「生成式 AI 風險及其管理機制」、「研究開發大型語言模型之著作權保護機制」，以及「本院研究人員以中研院名義從事相關學術活動之規範」等問題，進行跨領域研究，並提出相關建言；透過小組成員的共同努力，本院於 113 年 7 月 8 日將初步研究報告函送大院。小組目前持續針對特定的 AI 風險主題進行深入研究，並適時對外公布研究成果與政策建言。

其中就生成式人工智慧在以下 5 個層面可能產生的風險加以分析，並提出相應之對策，分述如下：

- (一) 著作權等智慧財產權保護問題：強調訓練資料和產出內容可能涉及的著作權、商標及營業秘密侵害風險，並建議透過立法或主管機關指引、契約條款授權機制等方式加以規範。
- (二) 個人資料保護與資料治理問題：關注生成式人工智慧可能導致的個人敏感資料洩漏、生成或推斷風險，以及訓

練資料來源透明度的問題，並提出建立防護監管機制、確保資料合法性與透明度、強化去識別化技術等對策。

(三)技術風險治理與使用者保護課題：探討幻覺、危險或暴力建議、人機互動及系統性風險，以及對未成年或易受傷害族群使用者的保護，並提出測量系統可靠性、進行攻擊演練、建立申訴機制、確保訓練資料安全等對策。

(四)民主影響評估與公共領域風險治理：警惕生成式人工智慧能對資料整全性和多元公共意見形成造成的威脅，並建議建立內容出處驗證機制、加強資訊共享與安全合作等。

(五)社會衝擊與環境影響的風險治理：分析資安攻擊、不當社會倫理觀念及高耗能對環境的潛在風險，並提出建立使用者回饋流程、進行影響評估、確保訓練資料安全及記錄能源消耗等對策。

總體而言，除分析生成式人工智慧在學術應用及社會各層面可能帶來的風險外，並從法律、技術、倫理等多重角度提供具體的應對策略與治理方向。

有關人工智慧基本法草案，尊重主責機關及各界意見，其中如有涉及法律或資訊等專業疑義，本院樂於協助提供意見。

以上報告，敬請

主席及各位委員、女士、先生指教，謝謝！