

## 國家人權委員會就《人工智慧基本法》草案之建議

- 一、鑒於人工智慧技術應用類型與範圍不斷增長，已逐漸成為人們日常生活的一部分，並促進社會進步，帶來生活上的便利，已為各界納入各種產品生產或服務提供之領域。然而，人工智慧技術之使用伴隨著隱私、言論自由、歧視等基本權利侵害風險，各國也因此訂定各種對人工智慧監測、人權風險影響評估等建議或執行方法。聯合國人權事務高級專員辦事處(OHCHR)也針對生成式AI可能導致的人權風險進行分類，並積極討論如何將《聯合國企業與人權指導原則》(United Nations Guiding Principles on Business and Human Rights, UNGP)及人權影響評估、盡職調查實踐於人工智慧系統生命週期活動。
- 二、伴隨著歐盟《人工智慧法》(AI Act)的問世，國科會於2024年7月15日預告《人工智慧基本法》草案(下稱本草案)，開啟對人工智慧可能造成人權侵害之議題進行回應。鑑於本草案係基本法，屬原則性之指引，其內容已包含永續發展、人類自主、隱私保護、資安與安全、透明可解釋、公平不歧視及問責等七大基本原則；且說明未來之四大推動重點，即創新合作及人才培育、風險管理及應用負責、權益保障及資料利用、法規調適及業務檢視，涵蓋美國、新加坡相關人權問責範圍，且與歐洲理事會(the Council of Europe)2024年5月提出《AI與人權、民主及法治框架公約》有關人工智慧系統生命週期活動相關原則、風險影響評估等立場無衝突之處。
- 三、針對本草案內容，本會建議如下：
  - (一) 本草案第三條第六款規定「公平與不歧視：人工智慧研發與應用過程中，應盡可能避免演算法產生偏差及歧視等風險，不應對特定群體造成歧視之結果」，惟「應盡可能」一詞之使

用屬不確定法律概念，恐無法有效保障人民免於歧視之權利。建議將「應盡可能」改為更具強制力的法律用語，確保規範具備法律強制性，並有效防止演算法產生偏差及歧視風險。

- (二) 本草案第十條規定「各目的事業主管機關得循前項風險分級框架，訂定其主管業務之風險分級規範」，第十二條亦提及「政府應依人工智慧風險分級，透過標準、驗證、檢測、標記、揭露、溯源或問責等機制提升其可信任度，建立人工智慧責任、救濟、補償等規範，明確責任歸屬與歸責條件」。換言之，基本法要求政府及各目的事業主管機關就人工智慧之應用，設立相關責任機制。然而，草案並未見對相關主管機關所制定之風險分級、評估驗證及救濟補償等規範設置監管機制，以確認是否落實，建議再酌。

四、基於本草案為基本法性質，為確保人權之落實，建議未來制定執行法規時，應參酌國際人權趨勢：

- (一) 歐盟《人工智慧法》規定，將建立公開的歐盟資料庫，以提高對於構成較高風險的人工智慧系統之透明度，除了規範高風險人工智慧系統的供應商之外，也要求實際使用系統的部署者應遵守透明度義務。為進一步提升我國人工智慧系統之透明與可解釋性，建議亦應強化對公部門及私部門之透明度要求。
- (二) 國際人權組織 Article 19 2024 年 4 月 3 日「歐盟《人工智慧法》未能為人權樹立典範」(EU: AI Act fails to set gold standard for human rights)指出，歐盟《人工智慧法》在保障人權方面存在諸多不足之處，特別是在情緒辨識及生物特徵分類系統之應用上。該法雖禁止此類技術於職場及教育機構中使用，然執法及移民機關卻被允許使用，恐有侵犯個人隱私及歧視弱勢群體之風險。為此，建議未來於政策制定及技術應用過

程中，實施人權影響評估，確保技術之應用符合人權保障原則，避免對特定群體造成不當侵害。

(三) 人工智慧的快速發展涉及大量能源消耗，惟相關環境影響資訊卻鮮為人知，建議亦應評估包含人工智慧系統之能源消耗資訊，藉以了解其對環境之潛在影響之可能性。

(四) 為確保人工智慧之使用不致對人民之基本權利造成侵害，或降低潛在侵害之風險，建請數位發展部等相關部會於訂定相關規範、標準文件及工具之制定過程，納入人權考量之各項因素，可參考歐盟《人工智慧法》(AI Act)之風險分級、《聯合國企業與人權指導原則》(United Nations Guiding Principles on Business and Human Rights, UNGP) 對於人權盡職調查之框架，建立符合國際人權規範之識別、評估與驗證程序，必要時讓本會一同參與。

(五) 本草案第十三條規定「政府為因應人工智慧發展，應避免技能落差，確保勞動相關權益」。對於企業如何在政府政策或規範下，踐行對勞動權益之保障，建議相關部會得與企業合作，透過人權教育推廣等方式，協助企業內化人工智慧應用與勞動權益保障之理念，並建議相關部會及企業朝提升職業技能輔導、協助合理職務調整等方式，確保既有勞動者之各項權益獲得有效保障。