僅供委員問政所需參考
不代表本院意見或立場

編號：2520

議題研析

一、題目：歐盟及美國人工智慧之風險監理法制模式對我國立法之啟發

二、議題所涉法規

人工智慧基本法草案

三、背景說明

人工智慧(Artificial Intelligence, AI）近年來快速發展，AI技術應用如雨後春筍遍及醫療、交通、金融、教育、娛樂、生產及銷售等各個領域，企業透過AI提供的服務雖然為民眾生活帶來便利，也帶動相關產業蓬勃發展，然而也隱含諸多風險，引發偽造、散播不實資訊、侵犯隱私，甚至威脅國家社會安全等問題，例如利用「擬人化機器人」影響選舉，不法份子利用「深度偽造」(Deepfake)作為詐騙工具等[1]。世界AI技術先進國家，例如歐盟及美國等，紛紛針對AI系統之風險發布相關監理法律或措施以為因應。考量AI發展影響社會層面廣泛，而我國目前尚無AI專法及產業評測標準，國家科學及技術委員會表示將研擬「人工智慧基本法草案」[2]，本院委員亦極關切並擬具草案[3]，引發民眾對於我國AI相關立法之關注與討論。

四、探討研析

[1] 社論，借鏡國外 加速AI立法管理，經濟日報，113年5月27日，第A2版。
[2] 張瓊，國科會擬定AI基本法草案 聚焦應用合法性等7面向，中央社，112年7月4日；范正祥，人工智慧基本法草案 國科會：年底前提出，中央社，113年3月19日。
[3] 立法院第11屆第1會期第11次會議議案關係文書，院總第20號，委員提案第11003774號，113年4月24日印發，立法院第11屆第1會期第9次會議議案關係文書，院總第20號，委員提案第11002784號，113年4月10日印發。

（一）歐盟與美國之 AI 風險管理機制簡介

AI 對個人和社會的潛在風險近年已逐漸顯現，促使各國政府針對 AI 帶來的風險積極建構監理機制，其中尤以有無必要制定不分行業別一體適用的「全面性監管專法」備受討論。歐盟採取制定專法模式，美國則採取不制定專法之監理策略[4]。以下爰就歐盟與美國之 AI 風險管理機制進行簡介。

1、歐盟

歐洲議會於 2024[5]年 3 月通過(Artificial Intelligence Act, AIA)，為全球首部 AI 監管專法。AIA 採取「風險基礎管制」(Risk-based Approach)模式，將 AI 系統應用上伴隨而生的潛在風險區分為 4 個級別，具體設定 AI 系統所應受到的規範密度[6]，對於違反之企業將被處以 3,500 萬歐元或全球年營收 7%的罰款(以金額較高者為準)[7]：

(1)不可接受的風險(Unacceptable Risk)係指對於人類的基本權利有所威脅或侵害者，包括操控認知技術、剝削特定弱勢族群、社會評分行為、執法單位在公共場所進行即時遠端生物辨識系統、預測犯罪行為、透過網路或閉路監視器擷取臉部影像以建立或擴增臉部辨識資料庫、在工作場所或教育機構進行情緒識別等[8]。依據 AIA 第 5 條規定將全面禁止使用該系統[9]。

(2)高風險(High Risk)之判斷係由 AIA 以附件清單列表方式規範[10]，

---

[4] 郭戎晉，〈國際趨勢下之人工智慧監管可能模式與臺灣推動課題〉，《全國律師》，第27 卷，第 6 期，112 年 6 月，頁 18。

[5] 本文有關年份之使用，原則以民國紀年表述，惟涉及全球性及國際性事件部分，改採西元紀年表述。

[6] 郭戎晉，同註 4，頁 26。

[7] 林妤柔，全球首部 AI 法案！6 月生效，罰款最高 3,500 萬歐元，科技新報，113 年 5 月 22 日，網址：https://technews.tw/2024/05/22/eu-worlds-first-major-law-for-ai/，最後瀏覽日期：113 年 6 月 21 日。

[8] 邱嘉琪，淺談歐盟人工智慧法案，中華經濟研究院，113 年 5 月 30 日，頁 1，網址：https://www2.itis.org.tw/NetReport/NetReport_Detail.aspx?rpno=48902536&type=netreport，最後瀏覽日期：113 年 6 月 21 日。

[9] 王煦棋，〈金融業 AI 監管的十字路口：從風控角度談臺灣 AI 法制〉，《當代法律》，第 28 期，113 年 4 月，頁 30-31。

[10] 王煦棋，同註 9，頁 31。

包括作為其他商品之安全零組件或獨立使用之 AI 系統，屬於歐盟產品安全法所規範之產品，例如機器、玩具、航空器、車輛、醫療儀器及電梯等，以及特定應用 AI，例如非禁止之生物特徵識別系統、重要基礎設施、教育與職業培訓、就業(如僱用及員工管理)、私人與公共服務(如銀行業信用評等及保險業評估系統)、執法、移民及邊境管制、司法行政與民主程序等[11]。針對高風險 AI 系統應遵守之義務包括，建立風險評估與緩衝系統、建立涵蓋產品生命週期之風險管理系統、記錄系統活動以確保該結果得以有效追溯、製作技術文件提供必要資訊，供主管機關評估其合法性、提供使用者清楚充分之資訊、透明度與資訊條款、適當人為監督機制以減少風險、確保 AI 系統之正確性、穩定性及網路安全等[12]。

(3)有限風險(Limited Risk)[13]包括與人互動的 AI 系統(如聊天機器人)、情緒辨識系統與生物特徵分類系統、偽造或變造影像、聲音、視訊內容的深偽系統，適用透明性義務及標註(識別)要求等，須告知使用者正在與 AI 系統互動或內容經由 AI 生成等[14]。

(4)低或最小風險(Low and Minimal Risk)係除上述之外的 AI 系統，不受 AIA 規範[15]。

2、美國

美國國家標準暨技術研究院(National Institute of Standards and Technology, NIST)於 2023 年 1 月正式發布「AI 風險管理框架」(Artificial Intelligence Risk Management Framework, AI RMF1.0)[16]，以幫助 AI 應用之決策與預期目標及價值保持一致，並降

[11] 邱嘉琪，同註 8。
[12] 邱嘉琪，同註 8。
[13] 亦有稱為特定透明風險(Specific Transparency Risk)者，詳見王煦棋，同註 8，頁 31；許莉美，歐盟「人工智慧法(AI Act)」於本(2023)年 12 月 9 日達成政治協議，經濟部國際貿易署，112 年 12 月 20 日，網址：https://www.trade.gov.tw/Pages/Detail.aspx?nodeID=45&pid=775685，最後瀏覽日期：113 年 6 月 21 日。
[14] 邱嘉琪，同註 8。
[15] 王煦棋，同註 9，頁 32。
[16] Crosswalks to the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)，NIST，113 年 4 月 26 日，網址：https://www.nist.gov/itl/ai-risk-management-framework/crosswalks-nist-artificial-intelligence-risk-management-framework，最後瀏覽日期：113 年 6 月 25 日。

低 AI 應用風險[17]。AI RMF 第一部分表明如何識別、減輕及最小化涉及 AI 技術的風險與潛在危害，並提出 7 項「可信賴 AI 系統」(trustworthy AI systems)關鍵特徵，包括有效性與可靠性、安全、穩固性與彈性、負責任與透明度、可解釋性與可闡述性、隱私強化、公平並管理有害偏見，以降低 AI 衍生風險並評估 AI 系統可資信賴與否。第二部分在風險有效識別的基礎上，建構 AI 風險應有的治理架構與管理標準，並提出以治理、路徑、量測及管理為核心之風險管理架構設計，同時闡述公私部門應如何落實 AI 風險管理體系[18]。NIST 並持續就 AI RMF 研提配套資料[19]。此外，2023 年 10 月美國拜登總統發布之「關於安全、可靠和值得信賴的 AI 行政命令」(Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence)亦要求持續制定 AI 安全和保障之新標準[20]。

（二）參考國外制度研議以 AI 風險管理為核心之監理立法模式

歐盟 AIA 採用風險分級方式，針對不同風險的 AI 系統賦予不同程度義務，美國則以鼓勵為主，並將 AI 風險聚焦在國家、經濟與公衛等領域之安全，規範具有嚴重風險之兩用基礎模型(dual-use foundational models)[21]的開發與交易，以及相關之大型資料中心[22]，

[17] 落實全球 AI 治理 美國發布可信賴 AI 行政命令，中央社，112 年 12 月 20 日，網址：https://www.cna.com.tw/postwrite/chi/360136，最後瀏覽日期：113 年 6 月 22 日。

[18] 郭戎晉，同註 4，頁 28。

[19] AI RISK MANAGEMENT FRAMEWORK，NIST，網址：https://www.nist.gov/itl/ai-risk-management-framework，最後瀏覽日期：113 年 6 月 22 日。

[20] FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence，THE WHITE HOUSE，2023 年 10 月 30 日，網址：https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/，最後瀏覽日期：113 年 6 月 21 日。

[21] 亦有稱為開放基礎模型(open foundation models)，只基於廣泛資料且通常以自我監督式學習方式加以訓練的 AI 模型，包含至少數百億參數，適用廣泛情境，且能展現出或經輕易修改而展現出給定任務的高水準表現，從而對國家、經濟、公衛安全，或三者之組合，構成嚴重風險。詳見黃禾田，美、歐對人工智慧（AI）管理模式異同之初探，中華經濟研究院 WTO 及 RTA 中心，112 年 11 月 15 日，網址：https://web.wtocenter.org.tw/Page/21/391300，最後瀏覽日期：113 年 6 月 21 日；Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights，National Telecommunications and Information Administration，網址：https://www.ntia.gov/federal-register-notice/2024/dual-use-foundation-artificial-intelligence-models-widely-available#ftn24，最後瀏覽日期：113 年 6 月 21 日。

[22] 黃禾田，同前註。

並以行政命令方式推動「產業管理標準」，以建立可持續運作的管理機制，助益公私部門識別 AI 系統風險[23]。AIA 之法律框架雖確保 AI 運用之公共安全性、提高使用者之信賴（需求面）及研發者或供應商之法安定性（供應面）[24]，但亦有學者認為 AIA 之各個風險級別的區分標準恐有未臻明確之虞，且由於 AI 系統應用伴隨而生的風險大抵處於持續且快速變動狀態，AIA 對此似亦欠缺合適的風險調整機制[25]。美國對 AI 之監管並非直接給予一定義務，係在既有規範下授權各機關為一定行為，或發布指引由業者遵循[26]。美國之 AI 行政命令可為未來 3 至 5 年的 AI 技術保留監管空間[27]。歐盟與美國對於 AI 之監管模式雖各有不同，但均以風險管理為核心。

AI 系統具有複雜性、不透明性、不可預測性、自主性(Autonomy)，涉及複雜且專業的領域，無論是政府機關進行監理或是企業進行法規遵循都會產生相當大的成本[28]，如何在兼顧 AI 產業創新發展，與國家、社會安全及人權保障之前提下，制定相關法律或監理措施具有相當難度。爰此，建議或可考量從 AI 設計、研發、部署及應用各階段，參酌外國立法例評估 AI 系統應用或輸出結果對個人或特定族群可能造成之風險，規劃預先採取之風險管理措施[29]，並授權相關機關訂定風險辨識與管理之標準，以為 AI 技術日後發展保留監管彈性。

撰稿人：安怡芸

[23] 郭戎晉，同註 4，頁 28。
[24] 陳錫平(主講人)，公共性與 AI 論壇(十二)歐盟人工智慧規則草案之初探——以市場、風險、價值及信賴為核心的管制架構，台灣人工智慧行動網，111 年 3 月 4 日，網址：https://ai.iias.sinica.edu.tw/eu-ai-regulation-draft-minutes/，最後瀏覽日期：113 年 6 月 21 日。
[25] 郭戎晉，同註 4，頁 28-29。
[26] 王煦棋，同註 9，頁 33。
[27] 黃禾田，同註 22。
[28] 賴文智，〈從歐盟首部AI監管法，看臺灣AI立法方向與挑戰？〉，《會計研究月刊》，第457期，112年12月，頁106。
[29] 落實全球 AI 治理 美國發布可信賴 AI 行政命令，同註 17。