《人工智慧基本法》制定了，然後呢？

林勤富*

第三波人工智慧（Artificial Intelligence, AI）興起從深度學習革命、大型語言模型、生成式 AI 迅速迭代出新，到近年代理型 AI 的破壞式創新應用，除龐大效益外，也衍生諸多新興風險，或對既有制度與秩序帶來挑戰。

面對 AI 迅速發展所帶來的經濟、社會與政治變遷，各國採取了不同治理模式試圖回應，反映其背後的制度基因、歷史脈絡與管制偏好。Anu Bradford 在《數位帝國》一書中就將美中歐區分為三種治理模式：[1] 美國的市場驅動模式（market-driven model）以去管制及強化競爭為核心，建構以軟法為主的治理體系，為了維持美國領先地位，美國總統川普撤銷原拜登政府的 AI 治理基本原則，並於 2025 年 12 月發布《確保國家人工智慧政策框架》禁止各州制定與該行政命令（去管制與鼓勵創新）相悖之立法，[2] 以貫徹聯邦主導地位並廣泛促進 AI 發展；不過，美國又於 2026 年 3 月發布《國家人工智慧立法框架》重新強調兒童保護、言論自由等權利導向的治理底線。[3] 中國的國家主導模式（state-driven model）則由國家引導科技發展與經濟成長，並透過數位監控建立起龐大的數位集權（digital authoritarianism）體制，針對科技雖不通盤管制，但在個別應用上強調社會穩定與國家控制、積極介入。[4] 與前述二者相較，歐盟的權利驅動模式（rights-driven model）則制定以硬法保障權利為核心的《人工智慧法》，[5] 禁止具

* 國立清華大學科技法律研究所教授

1 Anu Bradford. (2023). *Digital Empires: The Global Battle to Regulate Technology,* 7-10.

2 The White House, *Ensuring a National Policy Framework for AI,* The White House (11 Dec. 2025), https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/

3 The White House, *President Donald J. Trump Unveils National AI Legislative Framework,* The White House (20 Mar. 2026), https://www.whitehouse.gov/articles/2026/03/president-donald-j-trump-unveils-national-ai-legislative-framework/

4 如中國國家互聯網信息辦公室公布「人工智能擬人化互動服務管理暫行辦法」（2026 年 7 月 15 日起施行），禁止生成與傳播危害國家安全、損害國家榮譽和利益，或者散布謠言擾亂經濟與社會秩序等內容，並針對特定服務提供者定有須進一步進行安全評估的規定。

5 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828, 2024 O.J. (L 1689) 1.

不可接受風險的 AI 應用，並規範高風險 AI 應用；不過，歐盟執委會考量過於嚴格繁複的規範恐箝制產業創新，又於 2025 年 11 月公布《數位綜合規則草案》，嘗試簡化規定並降低法遵負擔，更延後《人工智慧法》諸多條文的生效期程。[6]

從前述規範的推陳出新與模式變遷，可以看出目前 AI 治理並非可一錘定音的議題。原因在於科技發展速度與高度不確定性，不同國家的管制哲學與制度相依，乃至於地緣政治變動與美中競爭、科技安全化（securitization）等趨勢，不但使治理機制必須滾動檢討，也使國家間規範取徑有交錯、調和甚至對立的現象。根據史丹佛大學人本 AI 研究中心之全球調查，2023 年時美國對於中國仍有壓倒性優勢，[7] 但在 2025 年起兩者間 AI 模型效能差距已顯著縮小，[8] 美中爭逐全球 AI 霸權的態勢不言可喻，競爭層面更從技術發展、供應鏈韌性，延伸到規範與標準制定、國家安全等面向。

若將視角轉回亞洲，韓國主要參考歐盟模式，但在同時側重產業創新的考量下於 2024 年 12 月 26 日率亞洲之先通過《人工智慧發展與建立信任基礎基本法》，[9] 日本則於 2025 年 5 月 28 日通過《人工智慧技術研究開發及應用促進法》，[10] 越南國會亦於 2025 年 12 月 10 日通過《人工智慧法》，[11] 而我國則於同年 12 月 23 日通過《人工智慧基本法》（下稱本法），正式開啟我國 AI 治理新局。

首先，有必要釐清我國 AI 專法的治理邏輯：管什麼、由誰來管、如何管？「掌握 AI 權力者」與「受到 AI 影響者」間的權力關係變遷及基本權利保護毋寧

---

[6] Proposal for a Regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus), COM (2025) 837 final (19 Nov. 2025).

[7] *See* Loredana Fattorini, Nestor Maslej, Ray Perrault, Vanessa Parli, John Etchemendy, Yoav Shoham, & Katrina Ligett, *The Global AI Vibrancy Tool,* HAI, at 21 (2025), https://hai.stanford.edu/assets/files/global_ai_vibrancy_tool_paper_november2024.pdf; Stanford University Human-Centered Artificial Intelligence (HAI), The 2025 AI Index Report 86, 97 (2025) [hereinafter 2025 AI Index Report]

[8] 2025 AI Index Report, supra note 8, at 97-8; Stanford University Human-Centered Artificial Intelligence (HAI), The 2026 AI Index Report 71, 77 (2026).

[9] Press Release, *The AI Basic Act Comes into Force to Lay the Foundation for Korea to Become an AI G3,* Ministry of Science and ICT (22 Jan. 2026), https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&bbsSeqNo=42&nttSeqNo=1214

[10] *Outline of the Act on Promotion of Research and Development, and Utilization of AI-related Technology (AI Act),* Japanese L. Translation Database Sys., https://www.japaneselawtranslation.go.jp/en/laws/view/4972 (last visited 25 Apr. 2026).

[11] Thuy Dung, *Government News: First-ever Law on Artificial Intelligence approved,* Vietnam Go‘t Portal (11 Dec. 2025), https://en.baochinhphu.vn/first-ever-law-on-artificial-intelligence-approved-111251211093619398.htm

是應重點關注的 AI 治理核心，後者可謂包山包海，前者則包含了掌握 AI 系統開發與訓練的私人科技巨擘、也包含部署、使用 AI 的公私部門行為者。因此，一部妥適的專法應該要盡可能採納全生命週期（lifecycle）的觀點，[12] 涵蓋公私部門對於具體 AI 系統的訓練資料蒐集與處理、模型開發與驗證、環境測試、標準制定、系統部署、應用與救濟等，甚至應涵蓋前階段教育、社會、經濟、產業、基礎建設等面向的政策準備。

從這點看來，本法不同條文分別觸及各類 AI 治理議題，如推動研發應用應遵循原則（第 4 條）、評估驗證（第 5 條）、知識技能與倫理教育（第 7 條）、基礎設施（第 8 條）、資料開放、共享及再利用機制（第 13 條）、研發應用過程之個資保護（第 14 條）、風險管理的國際標準（第 16 條）、責任歸屬（第 17 條）及政府風險內部控制（第 19 條）等，原則上涵蓋了不同生命週期階段的重點面向，誠值贊同。不過，除了以第 4 條基本原則作為定錨外，多處強調創新優越性、[13] 研發與產業發展的重要性，並未清楚闡明各階段、各面向的治理關聯性，似無全生命週期觀點的「牽一髮而動全身」規範思考，有待進一步建構與釐清。例如，基礎設施、資料治理、AI 應用的權利侵害問題，與風險管理與責任歸屬密不可分，完善治理亦有賴透明性與可解釋性之要求、救濟取徑，乃至於 AI 教育與倫理框架之落實，但若僅分別賦予不同機關不同任務、分列不同面向的問題，可能忽略了全生命週期觀點的核心規範意涵。

本法的中央主管機關為國家科學及技術委員會（下稱國科會），具體事項若涉及各目的事業主管機關職掌者則由該機關辦理。在國科會之上，行政院亦應成立國家人工智慧戰略特別委員會，審議國家人工智慧發展綱領，由國科會負責幕僚作業（第 6 條）。本法另特別賦予數位發展部（下稱數發部）「推動與國際介接之人工智慧風險分類框架」之任務，以協助其他機關訂定風險管理規範，並針對高風險 AI 系統對兒少利益等影響，提供或建議評估驗證工具或方法（第 5

[12] *See* U.S. National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (2023); *see also* Organization for Economic Co-operation and Development, *OECD Framework for the Classification of AI systems*, OECD Digital Economy Papers No. 323 (22 Feb. 2022), https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html

[13] 本法不同條文多次提及「創新」，可見立法者對於創新的重視，本法第 11 條更規定相關法規之解釋與適用如生扞格，在依循第 4 條基本原則的前提下，以促進新技術與服務之提供為優先原則。不過，「促進新技術與服務之提供」本身極不明確，法院、行政機關、科技產業界，誰有能力以及依據何種標準進行判斷？本項規定亦與既有的法規範適用順序思維有異，既非基於法位階理論、亦非基於特別法優於普通法等法適用理論，具體操作上可能迭生爭議。

條及第 16 條）。此外，本法第 18 條採近年盛行的基本法立法模式，[14] 要求政府各機關在本法施行後兩年內完成法規與行政措施之檢討與改進。

從治理任務的具體明確性來看，國科會作為（跨部會的）中央主管機關，肩負最大部分治理任務，但最缺乏具體指引，僅有第 4 條羅列重要原則（例如避免歧視與不平等），雖然提供制度續造的政策指引，或可一定程度作為節制行政權力的框架，但該條規定仍屬高度抽象（相較於各國立法例已有較為充實之規範內容，不論是軟法或硬法），具體制度形塑仍幾乎全盤仰仗行政機關，不易在彈性與不確定性間穩定以及在不同價值與利益間權衡，甚至衍生治理界線模糊的問題。而前述各機關之法規及行政措施檢討合致性之要求，雖設有明確時程，然而究竟各機關應依據何種標準進行檢討改進，卻同樣缺乏引導。[15] 倘若缺乏較具體的指引，或各機關缺乏遵循所需的能力與資源，則可能陷入各自為政甚至相互衝突的 AI 治理思維，這對於資源本身較少的臺灣來說，毋寧有重複、乃至於錯誤投入有限資源的風險，亦不利建構真正的部門（sectoral）治理。

相對而言，數發部在本法下的治理任務明確，較能採取具體行動。不過，值得注意的是，相對於歐盟《人工智慧法》明確以風險「分級」為基礎的治理模式，並透過附件與後續準則之制定，盡可能呈現不同風險層級 AI 系統應用情狀的具體圖像，[16] 我國基本法雖多處提及風險與高風險之概念，[17] 隱含有風險分級之設定，但並未釐清其具體內涵，而僅要求數發部推動相關風險「分類」框架（第 16 條第 1 項），然而「分級」與「分類」究竟有何不同？前者既然級數程度有別，自有高、低（可忽略）風險的區別，或是如歐盟一般劃下不可接受之風險紅線，從而也會伴隨著不同層次的「程序要求」及「法律效果」，因此更加符合法規範邏輯。若是「分類」，則可能主要以風險發生的來源、類型、效果、場域為主，是否能推導出程序要求與法律效果分層治理的結論，而分門別類後的 AI 應該有什麼程序要求與法律效果，則難以從條文中推知。[18] 因此，雖然我國基本法

[14] 例如《社會福利基本法》第 30 條、《海洋基本法》第 16 條等。

[15] 其他例子又如，本法第 14 條強調之個人資料保護與第 13 條的「資料開放、共享及再利用機制」應如何在第 4 條的隱私保護、資安與資料治理原則下受權衡與實行，亦難以從立法中探知指引方向並框定執行者權力。

[16] 林勤富（2025）。〈歐盟《人工智慧法》之制度設計、規範內涵與治理侷限〉，《中研院法學期刊》36 期，頁 38-50。

[17] 如第 4 條與第 16 條提及「風險」，第 5 條與第 17 條則提及「高風險」。

[18] 此等差異並非文字遊戲，從數發部目前的人工智慧風險分類框架（草案）來看，其所進行之「分類」除了廣泛識別風險來源與類型外（涵蓋 AI 技術設計缺陷、部署操作與人機互動問題、社會與環境衝擊等面向，第 16 條第 1 項），亦延伸到評估不同風險類型造成對「人民基本權利、生命安全、財產保障或社會秩序之嚴重危害」之「高風險」情狀（第 17 條），以及據此風險判定進行妥適應對管理措施之規劃。

之風險分類條文內容廣泛且未設有明確指引，特別是其所謂「與國際介接」的樣貌與取徑均待釐清（特別是考量前述各國不同治理模式與變動性，以及目前國際 AI 治理的碎片化問題），但數發部的治理作為，已初步形塑較為具體的規範內容。接下來，即可在此基礎上，清楚建構不同類型與程度之 AI 發展應用風險在我國實體法上分別代表的意義、判斷標準、程序要求與法律效果。

類似的廣泛授權與模糊指引，亦可見於本法其他涉及重要制度建立的條文。例如，本法第 17 條要求政府建立問責、救濟、補償或保險機制，第 13 條明定政府應建立資料開放、共享及再利用機制，第 8 條責令政府積極推動人才及技術之跨域合作、交流與基礎設施之建立，以及第 11 條規定各目的事業主管機關得針對人工智慧創新產品或服務，建立或完備人工智慧研發及應用服務之創新實驗環境。基本法寬泛要求相關主管機關形塑具體制度並非孤例，若仔細閱讀本法共計 20 條的內容即可發現，廣泛宣示與授權正是本法核心立法技術，特別是第 6 條至第 19 條一再出現「政府應……」或「某機關應……」的用語，凸顯立法者將 AI 治理的責任大量委諸行政部門，這在 AI 快速變動的年代下，確實有保留給相關機關審度時勢加以應對的必要性，但同時應注意清楚設定國家權力行使的框架，避免使行政權「球員兼裁判」，既無充分的法律限制，又獲得立法者廣泛授權，從而取得過於龐大的制度形塑權力。

若行政機關濫用廣泛授權而擁有過大制度創設權力，則下一個值得關注的問題是 AI 時代的新興權利是否應加以考量與肯認，如何形塑其內涵以及與既有權利之關係為何等。可惜的是，本法未能借力基本法之模式，羅列應受保障的新興基本權利清單，例如知情權、決策解釋權與人類審查權等與 AI 密切相關的新興基本權利類型。除本法第 4 條基本原則外，若能透過立法者的指引，敦促後續針對重要新興基本權利的優先討論與內涵補充，將能更快地在我國 AI 治理中生根發展。當然，這些新興基本權利主張無論在憲法、既有法律與 AI 治理體系中的定位、功能與內涵都需要再釐清，立法者未處理這些問題，一定程度上延後了有效基本權利保障機制的建立與相關討論；從另一方面來看，由於此等權利內涵發展的不確定性，留待未來由行政實踐與法院裁判基於既有的憲法基礎（如憲法第 22 條未列舉權概括條款）與基本權利理論加以發展，亦不失為一條穩健的發展路徑。這樣的兩難也凸顯 AI 發展迅速多變，法規範不易即刻因應挑戰，AI 治理如何走得穩健，仍有相當大的辯論與研究空間。

無論如何，《人工智慧基本法》的通過，象徵臺灣正式踏入全球 AI 治理競賽，但也如實揭示了我國科技治理的慣性——立法寬泛授權、依賴行政機關填補規範、缺乏權力分析與價值權衡機制，並傾向將複雜多面向的治理問題簡化

為經濟與產業創新議題。本法相對完整地考量了全球 AI 治理論述所凝煉出的重要原則與管制面向，羅列了創新、透明與可解釋、人類自主、公平與不歧視、問責、基本權利保障等核心，也涵蓋資料治理與個人資料保護、創新實驗環境、技術人才合作與基礎設施之建立、救濟補償或保險等重要配套制度，可說是已擘劃出基礎框架與價值。未來宜進一步納入全生命週期觀點、釐清各面向的關聯性與規範意涵、正面面對權力關係與社會結構的變遷，並確立價值衝突時的取捨準則，據此為治理機制的持續建構提供有效指引、建構面對快速科技變化相應的回應能力——這正是人文社會科學研究者在此刻不可缺席的根本理由。本法的通過，與其說是治理架構的完成，不如說是臺灣學習 AI 治理、在實踐中摸索切合在地脈絡之治理模式的真正起點。