研究論文
《中研院法學期刊》第 36 期
頁 1-119（2025 年 3 月）

歐盟《人工智慧法》之制度設計、規範內涵與治理侷限*

林勤富**

摘 要

歐盟《人工智慧法》於2024年8月1日正式成為歐盟法律，作為全球第一部針對人工智慧管制的全面性專法，除其制度設計與具體規定具參考價值外，其適用效力及規範影響力，亦將穿透歐盟疆界，對歐洲會員國外的其他國家或組織在人工智慧治理、標準制定與技術發展方向上產生重要影響，值得深入探討其制度框架、規範內涵與潛在侷限。本文回顧人工智慧帶來的各種挑戰，以及美國、英國、日本、中國與韓國等國的規範趨勢，並以此為基礎，研析《人工智慧法》之重要制度與規定，並從風險分級體系、人類監督、透明度與監理沙盒等面向評論可能的規範影響與疑義。最後，本文亦藉此比較我國人工智慧治理作為，指出潛在不足處與改善方向，以期政策制定者得藉由觀察歐盟立法模式理解國際規範趨勢，強化我國人工智慧之治理。

＊ 投稿日：2024年1月16日；接受刊登日：2024年8月19日。〔責任校對：陳沐恩〕。
本文研究受國科會研究計畫（計畫編號 NSTC 111-2423-H-007-002-MY3）補助支持，謹此致謝。作者感謝匿名審查委員對本文提出的評論與建議，惟一切文責當由作者自負。此外，作者特別感謝國立清華大學科技法律研究所研究生何冠霖於文稿撰擬及文獻蒐集時的研究協助，亦感謝研究生林耘生、洪言瑜對本文修改與校對過程之協助。

＊＊ 國立清華大學科技法律研究所教授兼所長。

穩定網址：https://publication.iias.sinica.edu.tw/54321142.pdf。

關鍵詞：人工智慧、歐盟、人工智慧管制、以風險為基礎取徑、透明度、可課責性。

目 次

壹、前言

自歐盟執委會（European Commission）於2021年4月份所提《人工智慧法律調和規則草案[1]》（Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised

1 *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 21, 2021) [hereinafter Draft Artificial Intelligence Act 2021 ver.].

Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts，以下簡稱「2021年版草案」）後，經歷歐盟理事會（Council of the European Union）於2022年12月6日推出理事會版本[2]、歐洲議會（European Parliament）於2023年6月14日對2021年版草案提出修正建議[3]（以下簡稱「2023年版草案」）[4]，以及歐洲議會、歐盟執委會及歐盟理事會三方協商（trilogue）後於2023年12月9日達成政治協議（political agreement）[5]，歐洲議會於2024年3月13日通過[6]，並於2024年7月12日正式發布於歐盟官方公報（Official Journal of the European Union），於同年8月1日正式生

2 Council General Approach No. 15698/22 of Dec. 6, 2022 [hereinafter Draft Artificial Intelligence Act 2022 ver.].

3 European Parliament, *Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, P9_TA(2023)0236 (June 14, 2023) [hereinafter Draft Artificial Intelligence Act 2023 ver.].

4 Adam Satariano, *Europeans Take a Major Step Toward Regulating A.I.*, N.Y. TIMES (June 14, 2023), https://www.nytimes.com/2023/06/14/technology/europe-ai-regulation.html.

5 European Council Press Release IP/986/23, Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World (Dec. 9, 2023), https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/.

6 European Parliament, *European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*, P9_TA(2024)0138 (Mar. 13, 2024); *see* Konrad Meier & Roger Spichiger, *The EU AI Regulation Is Coming. What Does It Mean for You and Your Business in Switzerland?*, EY (Mar. 15, 2024), https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business; *see also* Charlie Mitchell, *Senior EU Officials Say AI Act Adoption Expected in April Following Landmark Agreement*, INSIDE AI POLICY (Dec. 11, 2023), https://insideaipolicy.com/ai-daily-news/senior-eu-officials-say-ai-act-adoption-expected-april-following-landmark-agreement?s=na; Tambiama Madiega, *Artificial Intelligence Act*, at 1, EUR. PARL. RES. SERV., PE 698.792 (Sept. 2024), https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

效[7]，最終條文以2021年歐盟執委會最初提出之草案為核心，並綜合參採歐盟理事會與歐洲議會意見（以下簡稱《人工智慧法》）。《人工智慧法》對歐盟乃至於全球的人工智慧（artificial intelligence, AI）治理帶來重要規範啟示。相較於歐盟執委會提出之2021年版草案，政治協商後之版本新增了大幅度的修正，不僅是因為OpenAI公布了ChatGPT這項劃時代的生成式AI（generative AI）系統，挑戰原先規劃的管制框架[8]，更是因為AI迅速發展下，在推動創新與適當治理之間，不易取得平衡之故。AI雖被視為推動經濟發展之利器[9]，但其在公、私部門的廣泛部署運用，也造成各項新興風險與管制挑戰，使得各國在政治、經濟、法律等綜合考量下，各自發展出不同的治理方式。此外，由於歐盟法不同層次的「布魯塞爾效應」（Brussels Effect），亦即透過市場驅動力與跨國企業行為的規範擴散（norm diffusion）效果[10]，以及人工智慧發展與國際經貿實務

---

7 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 (Artificial Intelligence Act) (Text with EEA relevance), 2024 O.J. (L 2024/1689) [hereinafter Artificial Intelligence Act].

8 J. Scott Marcus, *Adapting the European Union AI Act to Deal with Generative Artificial Intelligence*, BRUEGEL (July 19, 2023), https://www.bruegel.org/analysis/adapting-european-union-ai-act-deal-generative-artificial-intelligence.

9 *Generative AI Could Raise Global GDP by 7%*, GOLDMAN SACHS (Apr. 5, 2023), https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.

10 布魯塞爾效應可分為兩層次，一方面是「事實上的布魯塞爾效應」（*de facto* Brussels Effect），即跨國企業依照歐盟法調整商業行為，確保能進入歐盟市場，而大型企業更可能以歐盟法規遵循為基準，在全球採用相同標準，要求上下游配合遵循，造成規範擴散效果。另一方面則是「法律上的布魯塞爾效應」（*de jure* Brussels Effect），指的是跨國企業在遵循歐盟法規要求後營運成本大幅提高，有商業的誘因遊說其母國政府採取相類標準，避免競爭劣勢。兩種層次之效應未必總是會發生，而發生後影響程度與內涵亦不同，而有論者另主張歐盟《人工智慧法》並不會如多數學者論述會自然觸發布魯塞爾效應，或有其他類型之影響（詳見本文第壹部分第三節更完整之討論）。Anu Bradford, *The Brussels Effect*, 107 NW. U. L. REV. 1, 3-6 (2012); ANU BRADFORD, THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD 2-4, 25 (2020).

的「管制相互依賴」（regulatory interdependence）[11]，乃至於各國政府基於國內管制需求、管制競爭或管制學習等因素，亦有參考歐盟在內的他國法規趨勢之動機。準此，本文並非主張以歐盟的《人工智慧法》作為直接法律移植或片面立法例，亦非敦促臺灣立法者全面採取歐盟管制模式，而係基於《人工智慧法》本身作為第一部AI專法之重要性、其規範示例（reference）以及透過跨國市場驅動與大型企業作為之規範擴散等間接效果，很可能將對非歐盟國家管制法規政策、國際組織倡議與跨國企業行為產生不同程度之影響，整體而言對未來全球AI發展與管制，形成不容忽視的規範力量。有鑒於歐盟AI法制潛在的全球性規範影響力，以及管制相互依賴、管制競爭或管制學習等實踐面向，本文雖不以採取歐盟《人工智慧法》全面性硬法管制模式為論述主張，仍強調深入剖析《人工智慧法》之制度設計、規範內涵與治理侷限的重要性，並宜進一步討論我國未來AI治理與歐盟法之互動關係、規範落差與調和等議題。

AI作為第四次工業革命的核心[12]，其標準制定、管制規範以及技術發展皆已為多數開發國家競爭之標的，本文以下將先梳理AI所帶來之挑戰，以及概覽歐盟以外之國家如何發展應對措施，以資與第貳部分詳述之歐盟管制作為相互參照。本文將續於第參部分自風險分級體系、人類監督、透明度與監理沙盒等面向評論歐盟《人工智慧法》之制度設計與規範重點。最後，本文將基於歐盟立法模式之研析，嘗試對我國現行AI治理之發展狀況進行比較與評論，提出初步建議，期政策制定者理解國際重要規範發展趨勢，系統性審視與形塑我國未來的AI治理。

---

11 Daniel Mügge, *Regulatory Interdependence in Artificial Intelligence*, ISRAEL PUB. POL'Y INST. (Apr. 11, 2022), https://www.ippi.org.il/regulatory-interdependence-in-artificial-intelligence/.

12 *See* KLAUS SCHWAB, THE FOURTH INDUSTRIAL REVOLUTION 14-15 (2016).

一、人工智慧帶來之挑戰

近年來，隨著公私部門採用各種AI系統，許多管制挑戰也逐漸浮現，例如何種AI系統應受到管制或禁止、公私部門的使用是否應受不同程度之管制、如何在管制與創新間取得平衡、以及如何避免與緩解不斷創新的各種AI系統對基本權利所生的風險與侵害等。在私部門部分，以近年荷蘭法院對歐盟《一般資料保護規則》（General Data Protection Regulation, GDPR）[13]中請求「自動化決策」解釋的判決具指標性意義。在該案中，數名使用Uber與Ola叫車平臺之司機主張，平臺不當使用自動化決策（automated decision-making, ADM）系統停止其工作資格（robo-firing，即透過機器人開除員工）、對其生計造成重大影響，並基於GDPR第13條至第15條之規定要求平臺提供有關ADM之相關資訊並解釋其運作方式，且獲得法院肯認[14]。惟依GDPR規定，資料主體僅能要求資料控管者就其個資及ADM系統之使用方式提出說明，而不能對ADM系統使用之妥適性提出挑戰，凸顯了現行規範不足之處[15]。

---

13 Regulation (EU) 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), arts. 5 & 32, 2016 O.J. (L 119) 1, 35-36 & 51-52 [hereinafter GDPR].

14 *See* Hof's-Amsterdam 4 April 2023, JAR 2023, 136 m.nt. C.M. Jakimowicz (Uber chauffeurs/Uber) (Neth.); Hof's-Amsterdam 4 April 2023, JBP 2023, 78 (chauffeurs/Ola Netherlands BV) (Neth.); Hof's-Amsterdam 4 April 2023, JAR 2023, 137 m.nt. C.M. Jakimowicz (Uber chauffeurs/Uber) (Neth.).

15 本案相關討論參見Sebastião Barros Vale & Gabriela Zanfir-Fortuna, *Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities*, FUTURE OF PRIVACY FORUM 22-24 (May 17, 2022), https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf; *see also* Wenlong Li & Jill Toh, *Data Subject Rights as a Tool for Platform Worker Resistance: Lessons from the Uber/Ola Judgments*, *in* DATA PROTECTION AND PRIVACY, VOLUME 15: IN TRANSITIONAL TIMES 119, 129-32 (Hideyuki Matsumi et al. eds., 2023).

另2022年美國司法部（Department of Justice）與臉書（Facebook）母公司Meta達成之和解協議，亦係因Meta不當使用AI演算法之故。該案中，司法部認為Meta使用演算法來決定房屋廣告在臉書頁面的投放對象，且演算法之判斷依據包括種族、膚色、宗教、性別、身心障礙、家庭狀況與國籍，將對臉書使用者產生歧視、違反《公平住房法》（Fair Housing Act）禁止因前述特徵而拒絕出租房屋等行為之規定，因而對Meta提起訴訟[16]。雙方後來達成和解協議，除支付罰款之外，Meta須停止使用其開發的「特定廣告受眾」工具（"Special Ad Audience" tool）來投放房屋廣告，並與司法部選擇獨立第三方審查者來調查與驗證Meta之系統是否符合規範[17]。雖然如同司法部官員表示，此案彰顯聯邦政府有能力在科技公司未有效處理演算法歧視問題時進行課責[18]，但另一方面，卻也顯示AI的使用可能與許多既有法規產生扞格，在缺乏專責治理框架與機構的情況下，只能仰賴各權責機構個案調查，難免掛一漏萬。

而近年因公部門使用AI而生的爭議，莫過於美國威斯康辛州的*State v. Loomis*一案[19]：州法院基於COMPAS系統的評估報告，認為被告Eric Loomis具有高再犯率評分，故予以重刑。被告主張COMPAS的演算法因營業秘密保護不對外揭露，使其無法挑戰系統評估方式之科學準確性，又該系統不當納入性別作為判定因素、參

16 *Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising*, U.S. DEP'T JUST. (June 21, 2022), https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known.

17 *Justice Department and Meta Platforms Inc. Reach Key Agreement as They Implement Groundbreaking Resolution to Address Discriminatory Delivery of Housing Advertisements*, U.S. DEP'T JUST. (Jan. 9, 2023), https://www.justice.gov/opa/pr/justice-department-and-meta-platforms-inc-reach-key-agreement-they-implement-groundbreaking.

18 *Id.*

19 *See* State v. Loomis, 881 N.W.2d 749 (Wis. 2016).

考群體資料而非單純以個人行為作為判斷基礎，已違反其受憲法所保障之正當法律程序權利等[20]。最高法院雖然承認該系統有許多問題[21]，但仍認為該演算法的量刑建議僅為參考之用，因此不認為該系統侵害Loomis的權利[22]。

另一則發生在休士頓的AI相關爭議，則是該州的獨立學區（Houston Independent School District）使用演算法評鑑教師表現，並基於演算法所產生之分數決定教師聘任與績效獎金等事項[23]，而遭到休士頓教師協會（Houston Federation of Teachers）提出訴訟，主張正當法律程序遭到侵害。不過法院認為該系統之使用有助於達成政府目的，仍具有合理關聯；至於演算法之具體運作方式雖然未經公開揭露，但因為教師已獲告知演算法之使用方式與概略評分依據，並非對該演算法的判斷標準一無所知、或完全缺乏評分標準，因此不違反「模糊標準」（vague standard）[24]，並未侵害正當程序。

---

20 *Id.* at 757.

21 *Id.* at 767-69.

22 雖最高法院認為量刑中使用COMPAS系統並未違反被告之正當法律程序權利，但指出未來若欲於量刑中使用COMPAS，應提供以下警示訊息：(1)由於營業秘密，COMPAS關於計算因素與權重資訊將不公開；(2)有研究指出，COMPAS不當地將少數族群定義為較高風險群體；(3) COMPAS所計算出的並非特定個人是否具有高再犯風險，而是特定群體；(4) COMPAS並未完成與威斯康辛州交叉驗證之研究；(5) COMPAS之開發目的並非用於量刑，而是用矯正機關為治療與假釋之相關認定。*Id.* at 769.

23 Mark A. Paige & Audrey Amrein-Beardsley, *"Houston, We Have a Lawsuit": A Cautionary Tale for the Implementation of Value-Added Models for High-Stakes Employment Decisions*, 49 EDUC. RESEARCHER 350, 352 (2020).

24 *See* Houston Federation of Teachers v. Houston Independent School District, 251 F. Supp. 3d 1168, 1180-84 (S.D. Tex. 2017); *see also* Ching-Fu Lin, *Sorting Teachers Out: Automated Performance Scoring and the Limit of Algorithmic Governance in the Education Sector*, *in* MONEY, POWER, AND AI: AUTOMATED BANKS AND AUTOMATED STATES 189, 189-204 (Zofia Bednarz & Monika Zalnieriute eds., 2023).

從以上案例可看出，AI雖可提高效率、減少人為疏失、協助處理重複性高之工作[25]，更能提升醫療診斷的準確度[26]、增進學生學習效果[27]等，帶來許多正面效益，但也衍生許多挑戰，包括不透明性[28]（即演算法黑盒子，詳見本文第參部分之討論）與可課責性（accountability）問題[29]、偏見（少數或弱勢族群尤其可能淪為受害者[30]）與正當法律程序之侵害[31]；政府若使用設計不良的ADM系統，更可能危害法治（rule of law）[32]。即使在刑事程序法體系中引進人工智慧或可增加司法人員工作效率，或有助於使司法變得更加透明、公平，但由於AI在面對個案時，將難以「審酌一切情狀」，

25 Rashi Maheshwari, *Advantages of Artificial Intelligence (AI) in 2024*, FORBES (Aug. 24, 2023), https://www.forbes.com/advisor/in/business/software/advantages-of-ai/.

26 例如AI能正確判讀X光片並發現病灶，進而減少放射科醫師之工作負擔。*See generally* Louis L. Plesner et al., *Autonomous Chest Radiograph Reporting Using AI: Estimation of Clinical Impact*, 307(3)e222268 RADIOLOGY (2023).

27 Wendy Kopp & Bo Stjerne Thomsen, *How AI Can Accelerate Students' Holistic Development and Make Teaching More Fulfilling*, WORLD ECONOMIC FORUM (May 1, 2023), https://www.weforum.org/agenda/2023/05/ai-accelerate-students-holistic-development-teaching-fulfilling/.

28 如李榮耕討論AI系統運作的不透明度以及黑盒子疑慮，見李榮耕，刑事程序中人工智慧於風險評估上的應用，政大法學評論，168期，頁153-160（2022年）；又如劉靜怡探討AI決策透明化實踐的困境，見劉靜怡，人工智慧時代的法學研究路徑初探，收於：李建良編，法律思維與制度的智慧轉型，頁102-104（2020年）。

29 Han-Wei Liu et al., *Beyond* State v. Loomis*: Artificial Intelligence, Government Algorithmization, and Accountability*, 27 INT'L J.L. & INFO. TECH. 122, 134 (2019); *see* Sylvia Lu, *Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence*, 23 VAND. J. ENT. & TECH. L. 99, 114-15 (2020); *see also* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 15 (2016).

30 林勤富、李怡俐，人工智慧時代下的國際人權法——規範與制度的韌性探索與再建構，收於：李建良編，法律思維與制度的智慧轉型，頁424-446（2020年），以及劉靜怡（註28），頁100-102。*See also* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1023-45 (2017).

31 *See* Danielle K. Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1 (2014).

32 *See generally* Monika Zalnieriute et al., *The Rule of Law and Automation of Government Decision-Making*, 82 MOD. L. REV. 425 (2019).

而僅會基於可量化的因素而進行判斷，恐將犧牲個案正義[33]，甚至引發新爭議而無助於解決問題[34]。

除前述案件外，某些類型的AI應用，特別容易侵害基本權利[35]。以臉部辨識科技（facial recognition technology, FRT）之運用為例，執法單位若透過FRT即時監控示威者[36]，將使得公眾擔心受到懲處或註記而不敢發聲，造成寒蟬效應，嚴重侵害言論自由與集會結社自由[37]，且此技術對某些族群的正確率特別低，亦有實質歧視之問題[38]。又如演算法操縱（algorithmic manipulation），係指運用AI分析個人資料、偵測個人弱點（vulnerabilities），並量身打造內容以直接或間接影響個人的決策流程[39]。最常見的便是個人化廣告，甚至可用來改變政治觀點、影響選舉[40]，過去高度爭議的劍橋分析（Cambridge Analytica）事件，即為經典案例[41]。面對這些新

33 薛智仁，初探人工智慧對刑法學的挑戰——以自動駕駛為例，台灣法律人，27期，頁107（2023年）。

34 李榮耕（註28），頁132-133。

35 AI之研發與運用對人權之影響，參見林勤富、李怡俐（註30），頁422-430。

36 例如美國巴爾的摩警察局曾於2016年遭揭露運用臉部辨識科技即時辨識抗議者。Russell Brandom, *Facebook, Twitter, and Instagram Surveillance Tool Was Used to Arrest Baltimore Protestors*, THE VERGE (Oct. 12, 2016), https://www.theverge.com/2016/10/11/13243890/facebook-twitter-instagram-police-surveillance-geofeedia-api.

37 Elizabeth Stoycheff, *Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring*, 93 JOURNALISM & MASS COMMC'N Q. 296, 306-07 (2016)；翁逸泓，面部自動辨識監控系統與個資保護問題研究——Bridges案之啟示，政大法學評論，168期，頁253-254（2022年）。

38 翁逸泓（註37），頁254-255。

39 Jon Whittle, *AI Can Now Learn to Manipulate Human Behaviour*, THE CONVERSATION (Feb. 11, 2021), https://theconversation.com/ai-can-now-learn-to-manipulate-human-behaviour-155031.

40 *See generally* Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J.L. & TECH. 106 (2019); *see also* Anu Bradford, *Whose AI Revolution?*, PROJECT SYNDICATE (Sept. 1, 2023), https://www.project-syndicate.org/onpoint/ai-regulation-us-eu-china-challenges-opportunities-by-anu-bradford-2023-09.

41 Alex Hern, *Cambridge Analytica Scandal 'Highlights Need for AI Regulation'*, THE GUARDIAN (Apr. 16, 2018), https://www.theguardian.com/technology/2018/apr/16/cambridge-analytica-scandal-highlights-need-for-ai-regulation.

興挑戰，各國無不開始敦促私部門開發與使用者正視負責任與可信任AI的重要性、以公權力介入AI發展及應用，確保AI的妥適運用。

二、各國法制因應作為

一般在討論新興科技治理途徑時，可概分為採取硬法（hard law）或軟法（soft law）為主的兩大取徑[42]（雖然此區分不免過於簡化）：前者係透過制定有強制拘束力的法律，規範AI之發展及應用，確保某些權利或利益之保障；後者則是以不具拘束力的準則（guidelines）或原則（principles）為主，強調市場機制與業者自律（self-regulation），藉此避免因過度管制而扼殺創新，僅由各主管機關就其權責範圍內訂定最低必要規範。綜觀各國目前作法，亦可粗略採用此一分類標準，例如美國、英國與日本可作為採取仍以軟法管制為主之代表；而就硬法部分，除本文分析主軸歐盟之外，中國與韓國等國亦已有制定AI專法之規畫或倡議[43]。以下將分別簡要討論前述國家之法制管制作為。

（一）目前以軟法管制為主之國家

1. 美國

美國雖然在AI發展處於領先地位，但出於經濟與政治利益考量，擔心將領先地位讓予中國，故不願過度介入管制[44]。截至目前

---

42 *See generally* ROGER BROWNSWORD & KAREN YEUNG EDS., REGULATING TECHNOLOGIES: LEGAL FUTURES, REGULATORY FRAMES AND TECHNOLOGICAL FIXES (2008); Julia Black & Andrew Douglas Murray, *Regulating AI and Machine Learning: Setting the Regulatory Agenda*, 10(3) EUR. J.L. & TECH. 1 (2019).

43 本文為便於討論，將歐盟與中國歸於相同類別，惟兩者管制思維大相逕庭，前者屬於「權利導向」（right-driven）之監管思維，旨在保障基本人權，而後者則是採取「國家導向」（state-driven）之態度，確保AI之發展能為國家掌控、為國家所用。Anu Bradford, *The Race to Regulate Artificial Intelligence: Why Europe Has an Edge Over America and China*, FOREIGN AFFAIRS (June 27, 2023), https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence?check_logged_in=1.

44 *Id.*

為止，美國雖在硬法層面已有若干AI相關草案提出，但均非全面性之法案。不過，美國總統行政辦公室（Executive Office of the President）與各聯邦機關於軟法（如基本原則、行為準則與政策指引）的制定，則相對積極且具體。美國政府早在2016年就已意識到AI帶來的挑戰與治理需求。當時由總統行政辦公室所發布之《為人工智慧的未來做好準備》（Preparing for The Future of Artificial Intelligence）報告[45]，認為既有法規已足以面對AI挑戰，不宜貿然制定廣泛法規（broad regulation）以免扼殺創新[46]。同時，由國家科學科技會議（National Science and Technology Council）公布的《國家AI研究與發展戰略計畫》（National Artificial Intelligence Research and Development Strategic Plan），則是建議透過標準與基準（benchmark）之建立，確保AI符合基本的安全與倫理期待[47]。

立法部門方面，聯邦層次上不乏有立法管制AI發展之草案提出，但仍未有全面性之立法通過[48]，例如2019年提出《2019年演算法課責法》（Algorithmic Accountability Act of 2019）草案[49]、針對演算法黑盒子的問題而提出的《2023年AI揭露法》（AI Disclosure Act of 2023）草案[50]等。整體而言，美國國會目前雖有數部AI相關

45 EXEC. OFF. OF THE PRESIDENT, NAT'L SCI. AND TECH. COUNCIL, NAT'L SCI. AND TECH. COUNCIL COMM. ON TECH., PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016).

46 *Id.* at 17.

47 SELECT COMM. ON ARTIFICIAL INTELLIGENCE OF THE NAT'L SCI. AND TECH. COUNCIL, NATIONAL ARTIFICIAL INTELLIGENCE RESEARCH AND DEVELOPMENT STRATEGIC PLAN 22-26 (2016).

48 美國眾議院於2023年12月22日通過《2024年國防授權法案》（National Defense Authorization Act for Fiscal Year 2024），其中部分條文提及應負責任地使用AI技術，以及推動數位浮水印與偵測生成式AI的使用等計畫。

49 Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. (2019-2020).

50 *U.S. Rep. Ritchie Torres Introduces Federal Legislation Requiring Mandatory Disclaimer for Material Generated by Artificial Intelligence*, RITCHIE TORRES (June 5, 2023), https://ritchietorres.house.gov/posts/u-s-rep-ritchie-torres-introduces-federal-legislation-requiring-mandatory-disclaimer-for-material-generated-by-artificial-intelligence.

之法案[51]，惟多仍處於草案提出階段、或僅針對特定的AI使用類型進行管制，尚無全面性之法案提出。不過，民主黨籍參議員Gary Peters與共和黨籍參議員John Cornyn於2023年7月共同提出法案[52]，要求各聯邦機構設置人工智慧長（Chief AI Officer）以確保聯邦政府妥善運用並管理AI[53]。同時，時任參議院多數黨領袖、民主黨籍的Chuck Schumer，亦曾與共和黨籍議員共同研擬因應AI威脅之法案[54]。這些跡象似乎顯示AI監管已成為兩黨共識，國會未來立法動向值得密切關注。在聯邦層次外，近年已有科羅拉多州與康乃狄克州等陸續針對自動化決策、公部門使用AI工具，乃至於針對AI進行全面管制的法案，不過仍屬少數作法[55]。

美國雖尚未於聯邦層級針對人工智慧的應用與風險有積極立法，但自2019年起即透過數則行政命令，一定程度地拘束與管制聯邦機關AI的使用。川普總統於2019年2月所簽署之行政命令《維持美國AI領導地位》（Maintaining American Leadership in Artificial Intelligence），同樣彰顯了美國對於AI監管之相對保守態度[56]。而行政管理與預算辦公室（Office of Management and Budget, OMB）於隔年針對聯邦機構對於AI應用設計、研發與運作之相關管制與非管

51 例如《人工智慧與生物安全風險評估法案》（Artificial Intelligence and Biosecurity Risk Assessment Act）、《演算法正義與線上平臺透明度法案》（Algorithmic Justice and Online Platform Transparency Act）、《消除演算法系統偏見法案》（Eliminating Bias in Algorithmic Systems Act）等。

52 AI LEAD Act, S. 2293, 118th Cong. (2023-2024).

53 *Committee Passes Peters & Cornyn Bipartisan Bill to Ensure Federal Government is Properly Using and Managing Artificial Intelligence*, HSGAC (July 27, 2023), https://www.hsgac.senate.gov/media/dems/committee-passes-peters-cornyn-bipartisan-bill-to-ensure-federal-government-is-properly-using-and-managing-artificial-intelligence/.

54 Oma Seddiq, *AI Threat Unites Bipartisan Senate Group Helmed by Schumer*, BLOOMBERG GOV'T (Sept. 7, 2023), https://about.bgov.com/insights/news/ai-threat-unites-bipartisan-senate-group-helmed-by-schumer/.

55 COLO. REV. STAT. § 6-1-1701 (2024); 2023 Conn. Pub. Acts No. 23-16.

56 José Vida Fernández, *Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State*, 30 IND. J. GLOB. LEGAL STUD. 65, 84 (2023).

制方法之研擬，發布之《人工智慧應用監管指引》（Guidance for Regulation of Artificial Intelligence Applications），為公部門面對「在聯邦政府外開發與部署之AI應用」應進行的政策考量設下指引[57]，但該報告仍認為AI應用並不必然會引發新興問題，應僅在必要情況下考量進行管制[58]。川普第一任總統期間所簽署最後一份有關AI之行政命令《促進聯邦政府使用可信任人工智慧的行政命令》（Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government），則是促使聯邦政府使用可信任的AI（trustworthy AI），並提出了聯邦政府在使用AI時應遵守之原則。

拜登政府上臺後，白宮科技與技術政策辦公室（White House Office of Science and Technology Policy）於2022年10月發布《人工智慧權利法案藍圖》（Blueprint for an AI Bill of Rights），對自動化系統的設計、使用與部署提出了五項原則，包括安全與有效的系統（Safe and Effective Systems）、對演算法歧視的保障措施（Algorithmic Discrimination Protections）、資料隱私（Data Privacy）、通知與說明（Notice and Explanation），以及提供人類替代方案、審酌與應變計畫（Human Alternatives, Consideration, and Fallback）等[59]，凸顯美國政府對於AI時代下的人民權利保護面向更加重視[60]。

但有論者認為，美國政府的監管思維，仍過於仰賴業者自律[61]。例如在2023年1月時，國家標準暨技術研究院（National Institute of

57 OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, M-21-06, MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: GUIDANCE FOR REGULATION OF ARTIFICIAL INTELLIGENCE APPLICATIONS (2020).

58 *Id*. at 3.

59 WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 13 (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

60 Fernández, *supra* note 56, at 85.

61 *See* Bradford, *supra* note 43.

Standards and Technology, NIST）發布了《AI風險管理框架》（Artificial Intelligence Risk Management Framework），期能藉此讓業者能設計出更可被信任、更負責任的AI[62]，但其性質仍屬於自願性標準。

不過，拜登總統基於此一自願性承諾而於同年10月簽署的第14110號行政命令《推動安全可信任的AI發展與運用》（Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence），內容涉及訂定人工智慧安全的新標準、保障美國人民隱私，以及強化平等與公民權利，具體作為則包括要求主要的AI開發者與政府分享其「安全檢測結果」（safety test result）與「關鍵資訊」（critical information，包括可能影響國家安全與經濟者），並指示NIST制定得以確保人工智慧系統安全與可信任（safe, secure, and trustworthy）的標準、工具與檢測方式等[63]。同年11月，NIST即建立「美國人工智慧安全研究院」（The U.S. Artificial Intelligence Safety Institute），致力於整合私部門與學術界之資源以促進人工智慧的安全與可信任[64]，顯示美國政府對於AI之監管似也正採取更加積極之態度。承前述拜登簽署之行政命令，OMB於2024年3月發布《促進聯邦機構使用人工智慧的治理、創新和風險管理》（Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence）政策[65]，要求各機關在不對個

62 U.S. NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, ARTIFICIAL INTELLIGENCE RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) (2023).

63 Exec. Order No. 14110, 88 Fed. Reg. 75191 (Oct. 30, 2023)（然此行政命令於2025年1月遭新任總統川普以阻礙私部門AI創新為由撤銷，Exec. Order No. 14148, 90 Fed. Reg. 8237 (Jan. 20, 2025)）; U.S. National Institute of Standards and Technology, *U.S. Artificial Intelligence Safety Institute*, https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute (last visited Nov. 20, 2023).

64 *See* Rick Weber, *NIST Kicks off AI Safety Institute by Stressing Long View for Establishing Ecosystem of Trustworthiness*, INSIDE AI POLICY (Nov. 20, 2023), https://insideaipolicy.com/ai-daily-news/nist-kicks-ai-safety-institute-stressing-long-view-establishing-ecosystem.

65 OFF. OF MGMT. & BUDGET, EXEC. OFF. OF THE PRESIDENT, M-24-10, MEMORANDUM

人隱私或政府運作上產生風險之前提下，公開其AI程式碼、模型與訓練資料，以及為掌握各級政府機關的AI使用情況，各機關每年應繳交其AI使用情況報告予行政管理與預算辦公室，並公告於自身網站。

2. 英國

英國早於2017年就將「人工智慧與巨量資料」列為四大影響產業發展的挑戰之一[66]，並建立人工智慧辦公室（Office for Artificial Intelligence）、資料倫理與創新中心（Centre for Data Ethics and Innovation）以及AI理事會（AI Council）等機構，其後亦針對演算法提出許多準則或標準，例如在2021年5月，英國資料倫理與創新中心、內閣辦公室（Cabinet Office）與人工智慧辦公室共同發布名為《自動化決策倫理、透明度與可課責性框架》（Ethics, Transparency and Accountability Framework for Automated Decision-Making）之七點框架指引，期能藉此協助政府在使用演算法決策系統時，能達成安全、永續與符合倫理的使用方式[67]。此外，為了增進公部門使用AI的透明度，英國中央數位與資料辦公室同年11月公布《演算法透明度標準》（Algorithmic Transparency Standard）[68]，

FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: ADVANCING GOVERNANCE, INNOVATION, AND RISK MANAGEMENT FOR AGENCY USE OF ARTIFICIAL INTELLIGENCE (2024).

66 其餘三者為潔淨成長（clean growth）、自駕車與人口高齡化。INDUSTRIAL STRATEGY: BUILDING A BRITAIN FIT FOR THE FUTURE, HM GOVERNMENT 7 (Nov. 27, 2017), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf (last visited Sept. 17, 2023).

67 *Guidance: Ethics, Transparency and Accountability Framework for Automated Decision-Making*, GOV.UK (Nov. 29, 2023), https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making#help-users-and-citizens-understand-how-it-impacts-them.

68 《演算法透明度標準》（Algorithmic Transparency Standard）後更名為《演算法透明度回報標準》（Algorithmic Transparency Recording Standard）。*See Guidance:*

揭露政府使用演算法工具與演算法協助之決策（algorithm-assisted decision）的相關資訊[69]，主要透過演算法登記制度（algorithmic register）來強化可課責性[70]，藉此增進公眾對於政府使用AI工具之信任度[71]。

不過，雖然英國政府對AI的透明度、倫理、安全性等法律議題均有所回應，觀察其於2022年針對《國家AI戰略》（National AI Strategy）所提出的《AI行動計畫》（AI Action Plan），似暫無制定AI專法之規畫[72]，而是以《AI與資料倫理框架》（AI and Data Ethics Framework）等軟法為主進行監管[73]。2023年3月公布的《支持創新的AI管制方法》白皮書（A Pro-innovation Approach to AI Regulation）中，同樣延續此一管制思維，指出現階段仍規劃透過公私部門合作而非制定專法應對AI發展，以避免倉促立法對企業與創新造成過度負擔[74]。

Guidance for Organisations Using the Algorithmic Transparency Recording Standard, GOV.UK (Jan. 5, 2023), https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard.

69 *Collection: Algorithmic Transparency Recording Standard Hub*, GOV.UK (Jan. 5, 2023), https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub.

70 Nigel Kingsman et al., *Public Sector AI Transparency Standard: UK Government Seeks to Lead by Example*, 2(1)2 DISCOVER A.I. 1, 2 (2022).

71 Marion Oswald et al., *The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives* 6 (July 7, 2022), https://ssrn.com/abstract=4155549.

72 *See Guidance: National AI Strategy - AI Action Plan*, GOV.UK (July 18, 2022), https://www.gov.uk/government/publications/national-ai-strategy-ai-action-plan/national-ai-strategy-ai-action-plan.

73 *See* Government Communications Headquarters, *Pioneering a New National Security: The Ethics of Artificial Intelligence* (Feb. 24, 2021), https://www.gchq.gov.uk/files/GCHQAIPaper.pdf.

74 Department for Science, Innovation & Technology, *A Pro-Innovation Approach to AI Regulation* 3 (Mar. 2023), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf.

3. 日本

日本亦屬於較早意識到AI發展重要性之國家，其總務省從2016年開始，年年召開AI網路化社會推進會議（AIネットワーク社会推進会議）討論AI造成的社會、經濟、倫理與法律議題，除發布報告書外，亦公布多項指引，例如2017年的《用於國際討論的AI研發準則草案》（国際的な議論のためのAI開発ガイドライン案）、2019年的《AI使用準則》（AI利活用ガイドライン）等[75]，惟究其性質，仍屬於不具拘束力、建議性質的軟法規範。

值得注意者為，日本非常積極參與制定AI國際標準，試圖在全球層次發揮規範影響力。例如日本內閣官房於2019年發布之《以人類為中心的AI社會原則》（人間中心のAI社会原則），成為經濟合作暨發展組織（Organisation for Economic Cooperation and Development, OECD）於2019年正式提出的《人工智慧建議》（Recommendation of the Council on Artificial Intelligence）之基礎[76]。

整體而言，雖然日本在國內外AI使用準則之制定相當積極，亦自2019年開始幾乎連年發布《AI戰略》（AI戦略）[77]，但其對於制定有拘束力規範之意願則相對消極。例如，在經濟產業省於2021年

---

75 AI Nettowāku Shakai Suishin Kaigi [The Conference toward AI Network Society], *Hōkokusho 2022 [2022 Report]*, SŌMU SHŌ [MINISTRY OF INTERNAL AFFAIRS AND COMMUNICATIONS] 1 (July 25, 2022), https://www.soumu.go.jp/main_content/000826564.pdf.

76 Ministry of Internal Affairs and Communications, *AI Utilization Guidelines* 3 (Aug. 9, 2019), https://www.soumu.go.jp/main_content/000658284.pdf; *see also* Hiroki Habuka, *Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency*, CSIS (Feb. 14, 2023), https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency; Organisation for Economic Cooperation and Development (OECD), *Recommendation of the Council on Artificial Intelligence*, at 7, OECD/LEGAL/0449 (May 22, 2019), https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

77 *AI Senryaku [AI Strategy]*, NAIKAKU FU [CABINET OFFICE], https://www8.cao.go.jp/cstp/ai/index.html (last visited Sept. 27, 2023).

7月發布之《AI治理報告書1.1版》（AI Governance in Japan ver. 1.1）中，便指出現階段尚無必要推出具有法律拘束力的規定[78]，直至2023年仍採取相同立場[79]，各主管機關現階段對AI發展也仍採取較為開放、支持產業創新的態度，僅要求企業採取適當措施並揭露風險資訊[80]。

（二）目前以硬法管制為主之國家

1. 中國

早在2017年時，中國國務院便公布《新一代人工智能發展規劃的通知》，定於2020年時在部分領域初步建立AI倫理規範與法規，並在2025年初步建立整體的AI「法律法規、倫理規範和政策體系」[81]。2019年，中國國家新一代人工智能治理專業委員會則是發表了《新一代人工智能治理原則——發展負責任的人工智能》文件，提出了AI發展所應遵循之原則，包括「和諧友好」、「公平公正」等[82]，究其性質，仍屬於不具拘束力之軟法規範[83]。到了2020

78 Expert Group on How AI Principles Should Be Implemented, *AI Governance in Japan ver. 1.1*, MINISTRY OF ECONOMY, TRADE AND INDUSTRY 28-29 (July 9, 2021), https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf.

79 Sam Nussey & Tim Kelly, *Japan Leaning toward Softer AI Rules than EU, Official Close to Deliberations Says*, REUTERS (July 4, 2023), https://www.reuters.com/technology/japan-leaning-toward-softer-ai-rules-than-eu-source-2023-07-03/.

80 *See* Habuka, *supra* note 76.

81 中華人民共和國國務院，國務院關於印發新一代人工智能發展規劃的通知，2017年7月20日，https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm（最後瀏覽日：2023年9月17日）。

82 中華人民共和國科學技術部，發展負責任的人工智能：新一代人工智能治理原則發布，2019年6月17日，https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html（最後瀏覽日：2023年9月17日）。

83 雖然其性質屬於軟法，但在中國的AI治理框架中，政府意見、技術標準與產業自律計畫都屬於治理工具；雖然這些軟法缺乏法律拘束力，但仍會透過自律採納（self-adoption）、在契約中參照、產業自治、輿論監督，以及市場競爭等方式，而獲得採用與執行。*See* Weixing Shen & Yun Liu, *China's Normative Systems for Responsible AI: From Soft Law to Hard Law*, *in* THE CAMBRIDGE HANDBOOK OF RESPONSIBLE ARTIFICIAL INTELLIGENCE: INTERDISCIPLINARY PERSPECTIVES 150, 153

年，中國信息通信研究院發布《人工智能治理白皮書》，指出應透過軟硬法兼施的管制作法來建構AI治理體系[84]。此後，中國從比較分散、反應式的治理，轉為集中管制模式，使用偏向硬法的管制措施來面對AI發展。

例如，中國於2021年底通過《互聯網信息服務算法推薦管理規定》，給予「應用算法推薦技術」明確定義，確保服務供應者對其演算法的監管與相應的法律責任、保障使用者避免受到不公平差別待遇，以及賦予使用者選擇停止演算法推薦的權利[85]。在爭議較大的深偽（deepfake）技術、人臉辨識與生成式AI部分，中國也已經著手研擬相關個別規範。例如中國國家互聯網信息辦公室（以下簡稱「網信辦」）於2022年11月偕同工業和信息化部與公安部發布《互聯網信息服務深度合成管理規定》[86]，據稱是全世界對於深偽技術最為全面的監管規定[87]。隔年7月，以網信辦為首的7個中央部會公布了《生成式人工智能服務管理暫行辦法》[88]，規定要求生成式AI之提供與使用應「堅持社會主義核心價值觀」、避免產生歧視、不得侵犯隱私，並提高透明度、準確性與可靠度等[89]。1個月

(Silja Voeneky et al. eds., 2022).

84 中國信息通信研究院、中國人工智能產業發展聯盟，人工智能治理白皮書，2020年9月，http://www.caict.ac.cn/kxyj/qwfb/bps/202009/P020200928368250504705.pdf（最後瀏覽日：2023年8月26日）。

85 中華人民共和國國家互聯網信息辦公室，互聯網信息服務算法推薦管理規定，2022年1月4日，http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm（最後瀏覽日：2023年8月26日）。

86 中華人民共和國國家互聯網信息辦公室，國家互聯網信息辦公室等三部門發布《互聯網信息服務深度合成管理規定》，2022年12月11日，http://www.cac.gov.cn/2022-12/11/c_1672221949318230.htm（最後瀏覽日：2023年8月26日）。

87 Asha Hemrajani, *China's New Legislation on Deepfakes: Should the Rest of Asia Follow Suit?*, THE DIPLOMAT (Mar. 8, 2023), https://thediplomat.com/2023/03/chinas-new-legislation-on-deepfakes-should-the-rest-of-asia-follow-suit/.

88 中華人民共和國國家互聯網信息辦公室，生成式人工智能服務管理暫行辦法，2023年7月13日，http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm（最後瀏覽日：2024年12月13日）。

89 生成式人工智能服務管理暫行辦法第4條，見前註88。值得注意的是，網信辦最後所採納的版本，較上半年公布的《生成式人工智能服務管理辦法（徵求意

後，網信辦則是公布了《人臉識別技術應用安全管理規定》的草案[90]，要求「只有在具有特定的目的和充分的必要性，並採取嚴格保護措施的情形下，方可使用人臉識別技術處理人臉信息」[91]。

2. 韓國

2019年10月，時任總統文在寅曾經宣布「總統AI倡議」（Presidential Initiative for AI），提及AI發展的重要性與急迫性，凸顯建立國家戰略之重要性[92]。而在年底公布的《國家AI戰略》（National Strategy for Artificial Intelligence）中，也建議韓國政府應制定全面性的監管創新策略[93]，包括基於「先批准、後監管」（approval first and regulate later）的取徑，以原則開放、負面表列之管制方式[94]，建立AI時代的未來導向（future-oriented）法律體系[95]，並建立全球層次的AI倫理規範[96]等作為，在AI治理方面展現積極態度。

《國家AI戰略》公布後，韓國分別在2020年底推出《AI倫理標準》（인공지능(AI) 윤리기준）[97]、2021年提出《實現可信任AI戰

見稿）》更為寬鬆，也顯示中國在生成式AI管制的力度仍持續調整。見張雅涵，中國對AI生成內容實施新規範 較草案寬鬆以維持競爭力，中央廣播電臺，2023年8月18日，https://www.rti.org.tw/news/view/id/2177146（最後瀏覽日：2024年12月13日）。

90 中華人民共和國國家互聯網信息辦公室，國家互聯網信息辦公室關於《人臉識別技術應用安全管理規定（試行）（徵求意見稿）》公開徵求意見的通知，2023年8月8日，http://www.cac.gov.cn/2023-08/08/c_1693064670537413.htm（最後瀏覽日：2023年8月26日）。

91 人臉識別技術應用安全管理規定（試行）第4條，見前註90。

92 Ministry of Science and ICT, *National Strategy for Artificial Intelligence* 3-4 (Dec. 17, 2019), https://wp.oecd.ai/app/uploads/2021/12/Korea_National_Strategy_for_Artificial_Intelligence_2019.pdf.

93 *Id.* at 24.

94 *Id.* at 24-25.

95 *Id.* at 26.

96 *Id.* at 47.

97 *Ingongjineung (AI) yunligijun [Artificial Intelligence (AI) Ethical Standards]*, MSIT, https://www.msit.go.kr/bbs/view.do?sCode=user&mPid=112&mId=113&bbsSeqNo=94&nttSeqNo=3179742 (last visited Sept. 28, 2023).

略》（Strategy to Realize Trustworthy Artificial Intelligence）[98]，以及2022年公布《人工智慧開發與應用之人權準則》（인공지능 개발과 활용에 관한 인권 가이드라인）[99]。韓國另已著手制定AI專法，國會「科學、資通、廣播與通訊委員會」（과학기술정보방송통신위원회）於2023年2月已通過《促進AI產業與建立可信任AI架構法案》（인공지능산업 육성 및 신뢰 기반 조성 등에 관한법률）草案[100]，企圖成為全球第一個制定AI專法之國家[101]，有望於韓國國會第21屆屆期滿前通過（2024年5月）[102]。韓國國會最終於2024年12月26日，在動盪政局下通過《人工智慧發展與建立信任基礎基本法》（인공지능 발전과 신뢰 기반 조성 등에 관한 기본법），並將於2026年1月22日正式施行[103]。

98 *MSIT Announced "Strategy to Realize Trustworthy Artificial Intelligence"*, MSIT, https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&pageIndex=&bbsSeqNo=42&nttSeqNo=509&searchOpt=ALL&searchTxt= (last visited Sept. 27, 2023).

99 *Ingongjineung Gaebalgwa Hwaryonge Gwanhan Ingwon Gaideurain [Human Rights Guidelines for the Development and Application of AI]*, Gukgaingwonwiwonhoe [National Human Rights Commission of Korea] (May 17, 2022), https://www.humanrights.go.kr/site/program/board/basicboard/view?boardtypeid=24&boardid=7607961&menuid=001004002001.

100 *South Korea: Legislation on Artificial Intelligence to Make Significant Progress*, KIM & CHANG (Mar. 6, 2023), https://www.kimchang.com/en/insights/detail.kc?sch_section=4&idx=26935.

101 Somang Yang, *Spurred by ChatGPT, South Korea Pushes World's First AI Bill*, PS-ENGAGE (Mar. 16, 2023), https://ps-engage.com/spurred-by-chatgpt-south-korea-pushes-worlds-first-ai-bill/.

102 백연식, 'AI 기본법' 올해는 빛볼까...과기정통부"21대 국회 종료 전 통과 목표" ['AI基本法'今年能夠見到曙光嗎... 科技情報通訊部 "目標在第21屆國會結束前通過"], DIGITAL TODAY (Jan. 2, 2024), https://www.digitaltoday.co.kr/news/articleView.html?idxno=499842.

103 《人工智慧發展與建立信任基礎基本法》（인공지능 발전과 신뢰 기반 조성 등에 관한 기본법）。此部法律是繼歐盟《人工智慧法》後，全球第二部全面性規範人工智慧發展的法律。

三、全球首部AI專法的布魯塞爾效應

從各國近年規範取徑之盤點可以發現，對AI採取積極的管制作為係大勢所趨，惟許多國家仍以軟法之管制手段為主，即使有制定專法之規畫，也多處在研擬階段，凸顯歐盟已完成制定硬法、以「全面性監管專法」[104]管制AI作法之迅速與特殊重要性。尤其考量歐盟法規的「布魯塞爾效應」，歐盟《人工智慧法》將透過其規範示例效果或市場機制驅動，對各國管制措施之形塑，以及相關企業實踐產生不同程度的影響。

布魯塞爾效應至少可從兩個層次觀察，一方面，跨國企業將調整其商業行為以符合歐盟規則，確保其商品與服務能進入歐盟市場並避免法律責任，而大型企業更可能以歐盟法規遵循為基準，在全球採用相同的標準，此即為「事實上的布魯塞爾效應」（*de facto* Brussels Effect）；另一方面，各國政府基於國際經貿實務、管制競爭或管制學習，固有參考歐盟在內的他國法規趨勢之動機，而跨國企業在採行歐盟法規的嚴格標準後，亦有商業上誘因回過頭遊說母國政府採取相類嚴格標準，確保其面對非出口導向、不用遵循歐盟嚴格規定的國內競爭者，不處於劣勢，因此逐漸形成「法律上的布魯塞爾效應」（*de jure* Brussels Effect），最終使若干國家採取與歐盟相近之管制作為[105]。近年歐盟GDPR的廣泛影響即為適證，作為一種被動但有力的市場驅動機制，藉由跨國企業的遵循與擴散，布魯塞爾效應時常比其他更主動之機制（如透過簽訂條約或國際組織合作機制）更有實質影響力。

不過，布魯塞爾效應亦有其侷限性，可能受政策領域不同或管制競爭脈絡變化而異，並非均有相同程度、相同性質的影響力。雖

104 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，全國律師月刊，27卷6期，頁25-27（2023年）。

105 Bradford, *supra* note 10, at 3-6; BRADFORD, *supra* note 10, at 2-4, 82-91.

然多數學者均認為布魯塞爾效應已實質形成並將持續發生規範擴散影響力，但亦有少數學者認為，在各國近年人工智慧技術與規範競爭趨白熱化的情況，前述美國、日本、中國等管制措施亦會參與其中，混雜地對其他國家產生不同方向、類型的規範影響力，加上歐盟本身硬法模式，在技術迅速發展且不確定性高的情況下，可能不易與時俱進，進而降低其參考價值。基於此等因素，歐盟《人工智慧法》未必會產生強烈的布魯塞爾效應，而可能如學者Ugo Pagallo所主張，逐漸形成所謂「拼布效應」（patchwork effect）而非以歐盟獨大的布魯塞爾效應[106]。

整體而言，歐盟搶先各國制定《人工智慧法》，如多數學者所論，預期將有引領跨國企業行為與各國立法選擇之效果[107]，雖然其

106 Ugo Pagallo, *Why the AI Act Won't Trigger a Brussels Effect*, *in* AI APPROACHES TO THE COMPLEXITY OF LEGAL SYSTEMS 1, 2, 4-5 (forthcoming). 詳言之，Pagallo提出歐盟《人工智慧法》可能不會當然產生布魯塞爾效應的主張，係基於外部因素（exogenous reasons）與內生因素（endogenous reasons）兩面向：外部因素指的是各國AI規範競爭激烈，並非僅會以歐盟獨大，美國、中國、日本近年均積極採取管制措施或標準制定程序，特別是日本在國際層次上的軟法影響力，已在ISO（International Organization for Standardization）與OECD發揮實質效應。內生因素則是指相對於其他國家，歐盟針對AI技術發展速度與不確定性極高的議題，領先全球採取全面性硬法管制模式，而硬法具較高之僵固性，相對於行為準則、標準等軟法，很有可能不久後即產生適用問題，甚至窒礙難行之問題，進一步限縮其跨國影響力。若進一步細究，考量AI的技術發展速度以及高度不確定性，過早以硬法模式進行全面性規範，確實可能產生（如生成式AI對歐盟《人工智慧法》2021年版草案造成的挑戰）法規未必能與時俱進的適用困境。不過，此一內生因素與《人工智慧法》規範內涵的條文彈性與主管機關對於解釋適用的取徑有關，而各國在目前時空下是否參考歐盟《人工智慧法》之考量因素，亦未必會圍繞在科技法的步調難題，是故不當然會限制布魯塞爾效應之產生。

107 *EU Engagement in Global AI Standard Setting*, REGULAITE, https://www.regulaite.eu/iv-the-eu-and-global-ai-standard-setting (last visited Oct. 18, 2023). *But see* Alex Engler, *The EU AI Act Will Have Global Impact, But a Limited Brussels Effect*, BROOKINGS (June 8, 2022), https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/. 管制過嚴也可能阻礙歐盟企業發展，或是讓企業選擇在歐盟以外的地區開發AI而使歐盟規則欠缺影響力或難以被執行。*See Theorizing Regulatory Interdependence in EU AI Policy*,

布魯塞爾效應可能受技術發展或管制競爭脈絡變化而異，並非均有相同程度、相同性質的影響力，但《人工智慧法》諸多重要制度設計與規範內涵，仍值得密切注意與研析。是故，本文將於以下第貳部分概述歐盟《人工智慧法》之背景與其重要規範，作為後續討論之基礎。

貳、歐盟人工智慧法之立法脈絡與規範內涵

一、制定背景

演算法需要巨量資料驅動[108]，而資料帶有龐大經濟價值、法規遵循成本亦高，以致供應商與部署者具有強烈動機以規避法規之方式與商業夥伴交易資料[109]，這也使得AI系統的應用導致許多隱私權侵害情形。此外，諸如在司法程序中對正當法律程序權利侵害、偏見與歧視的固化、生成式AI與深偽內容所生的假消息問題、剖繪（profiling）與不公平待遇、針對性廣告投放等演算法部署後開始衍生之問題，也陸續浮上檯面。

2016年時，歐盟執委會意識到快速的科技發展與全球化對於個資保護帶來更大挑戰[110]，尤其個資蒐集方式日益精巧複雜與難以偵測[111]，以及科技使得個人資訊變得更為公開且隨處可近用，均使個

REGULAITE, https://www.regulaite.eu/i-why-regulatory-interdependence-matters (last visited Oct. 18, 2023).

108 Manheim & Kaplan, *supra* note 40, at 120.

109 Sylvia Lu, *Data Privacy, Human Rights, and Algorithmic Opacity*, 110 CAL. L. REV. 2087, 2102 (2022).

110 *Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: A Comprehensive Approach on Personal Data Protection in the European Union*, at 2, COM (2010) 609 final (Nov. 4, 2010).

111 *Id.*

資利用之態樣超越立法者過去之理解與想像[112]，故於制定GDPR時，將資料保護的層級從指令（directive）提升到規則（regulation）。為因應科技發展對個人隱私與基本權之衝擊，GDPR賦予資料主體許多新興數位權利（digital rights），包括資料可攜權、被遺忘權、刪除權、免受剖繪權與解釋權等等[113]，對於資料的處理設下更多限制[114]，對於資料主體同意的規定也更為細緻[115]。有學者甚至認為，GDPR是最早意識到演算法不透明問題、並試圖解決AI可解釋性與倫理問題之法律[116]。例如，因應對於公司使用巨量個人資料的擔憂，GDPR第12條至第14條設下了更高的透明度標準，要求企業以透明與公平的方式管理資料[117]。第22條也規定，企業應就自動化決策、資料處理、運算邏輯與可能的結果提供有意義的資訊（meaningful information about the logic involved）[118]。

即使GDPR在隱私與個人資料保護方面仍堪稱先進，但其所規定有關AI與自動化決策相關之資料揭露方式，無法徹底解決AI帶來的問題[119]，因為GDPR之本質與規範目的，是在限制個人資料的處理與運用，而非AI的研發與應用[120]，相關條文的解釋與適用，則易受到限縮或有不相應（mismatch）問題。例如，由於先進複雜的AI

112 *Id.*

113 劉靜怡，淺談GDPR的國際衝擊及其可能因應之道，月旦法學雜誌，286期，頁6（2019年）。

114 *See, e.g.*, GDPR arts. 5 & 32.

115 *See, e.g.*, GDPR art. 7.

116 Alexander J. Wulf & Ognyan Seizov, *"Please Understand We Cannot Provide Further Information": Evaluating Content and Transparency of GDPR-Mandated AI Disclosures*, 39 AI & SOC'Y 235, 236 (2024).

117 GDPR arts. 12-14.

118 GDPR arts. 13, 14, 15 and 22.

119 Wulf & Seizov, *supra* note 116, at 252. 有關GDPR與AI的相關討論，詳見本文第參部分第二節。

120 林勤富、李怡俐（註30），頁435。*See also* Marijn Storm & Alex van der Wolk, *Privacy and the EU's Regulation on AI: What's New and What's Not?*, 4 J. ROBOTICS A.I. & L. 451 (2021).

受到營業秘密保護，外人難以檢驗企業對於個資的使用方式是否確實符合蒐集目的；像是Google利用演算法對個資進行的分析與運用，幾乎都以「強化使用者體驗」一語帶過而合理化，卻無從檢驗AI所造成之風險[121]。此外，GDPR看似能充分保護資料主體免受自動化決策影響，但實際上亦並非如此，解釋權之規定也因不夠明確而可能無法落實[122]。另一方面，從鼓勵創新的角度來看，若過度擴張這些權利的範圍與內涵，也可能會對AI的發展設下過多限制。例如被遺忘權的主張，若採寬鬆解釋，除可能衝擊知識傳播與言論自由外，亦會影響用於訓練人工智慧的資料，因而削弱人工智慧的安全性、可信度與預測能力[123]。因此，單純仰賴GDPR恐怕已經無法對AI進行適當管制，而需要新的制度設計與規範。

為了更有效、全面地管制AI發展，歐盟執委會於2018年4月之討論報告中，指出AI已經獲得廣泛運用，因此除須持續推動AI發展之外，亦應確保AI發展能保障重要價值、基本權利與倫理原則，例如可課責性與透明度[124]。為了協助執委會提供AI策略方面之建議，歐盟於同年6月成立了獨立的「人工智慧高級專家小組」（High-Level Expert Group on Artificial Intelligence, AI HLEG）。AI HLEG隨後陸續發布多項政策文件，包括《可信任AI的政策與投資建議》（Policy and Investment Recommendations for Trustworthy AI）[125]、

121 Lu, *supra* note 109, at 2118.

122 Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 96 (2017).

123 Matthew Humerick, *Taking AI Personally: How the E.U. Must Learn to Balance the Interests of Personal Data Privacy & Artificial Intelligence*, 34 SANTA CLARA HIGH TECH. L.J. 393, 405-08 (2018).

124 *Communication from the Commission to the European Parliament, The European Council, The Council, The European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe*, at 2, 14, COM (2018) 237 final (Apr. 25, 2018).

125 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, POLICY AND INVESTMENT RECOMMENDATIONS FOR TRUSTWORTHY AI (2019).

《可信任AI之倫理準則》(Ethics Guidelines for Trustworthy AI)[126]，以及《可信任AI自我評估清單》(The Assessment List for Trustworthy Artificial Intelligence for Self-Assessment)[127]等，尤其後兩者是歐盟現階段因應AI發展的具體法律框架[128]，對於AI法的制定有指標性的意義。

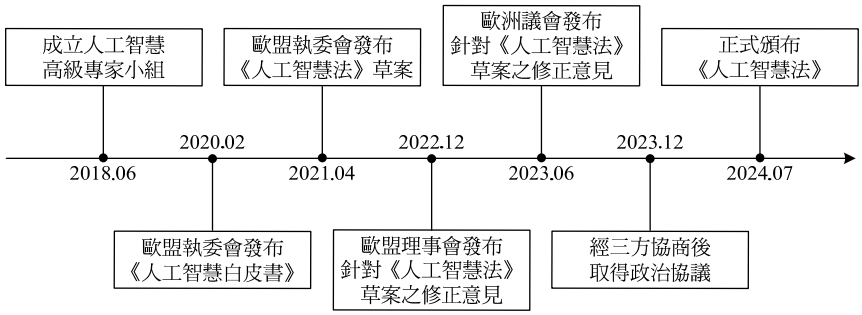

圖1 歐盟《人工智慧法》立法背景與發展時間軸

資料來源：本文作者繪製。

2020年2月，歐盟執委會進一步發布了《人工智慧白皮書》(White Paper on Artificial Intelligence)，其中提到臉部辨識等高風險的AI運用在上市前應執行「符合性評估」(conformity assessment)，並指出應建立相關程序，使權責單位可以確保高風險的AI系統符合基本人權[129]。同年7月，歐盟執委會發布了對於AI法制的評估分析報告，指出新的法規將會著重於(一)消費者／使用者保護；(二)隱私、言論自由等基本人權的保障與(三)AI所生非法歧

126 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, ETHICS GUIDELINES FOR TRUSTWORTHY AI (2019).

127 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, THE ASSESSMENT LIST FOR TRUSTWORTHY ARTIFICIAL INTELLIGENCE (ALTAI) FOR SELF-ASSESSMENT (2020).

128 Fernández, *supra* note 56, at 86.

129 *See Commission White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*, COM (2020) 65 final (Feb. 19, 2020).

視[130]。同年7月，歐洲議會亦建立了「數位時代AI特別委員會」（Special Committee on Artificial Intelligence in a Digital Age），特別委員會並發布了數個關於AI倫理的相關文件[131]，以及針對AI責任歸屬相關的建議報告[132]。歐盟執委會基於白皮書之政策建議，以及白皮書發布後向利害關係人與公眾就其內容所進行的公眾諮詢結果[133]，於2021年4月發布2021年版草案[134]，正式展開AI法之立法程序[135]。

130 *See* European Commission, *Artificial Intelligence – Ethical and Legal Requirements*, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Requirements-for-Artificial-Intelligence (last visited Sept. 1, 2023).

131 European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies (2020/2012(INL)), 2021 O.J. (C 404) 63.

132 European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Civil Liability Regime for Artificial Intelligence (2020/2014(INL)), 2021 O.J. (C 404) 107.

133 Draft Artificial Intelligence Act 2021 ver. explanatory memorandum 1.1 and 3.1.

134 Draft Artificial Intelligence Act 2021 ver.

135 除《人工智慧法》外，歐盟近年亦訂有若干與AI發展及管制相關的法規，如GDPR、《資料治理法》（Data Governance Act）、《數位服務法》（Digital Services Act）以及《資料法》（Data Act）。其中《資料法》對AI的發展有不小的影響，資料作為AI發展之根基，《資料法》確保資料的公平取得與使用以防止市場壟斷，以及促進私部門、公部門與個人間的資料共享，並允許公部門於緊急情況得近用私部門資料，俾使基本權利與公共利益受到保障。此外，歐盟執委會為使其所提出之2021年版草案更加完善並銜接歐盟《1985年產品責任指令》（Product Liability Directive 1985），另於2022年9月提出《人工智慧責任指令草案》（Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence），並指出2021年版草案與《人工智慧責任指令草案》相輔相成：前者旨在防止AI系統造成損害；後者則是建立AI相關損害發生時的責任歸屬與賠償機制。囿於直接關聯性及篇幅限制，《資料法》與《人工智慧責任指令草案》等相關法規，非本文目前集中討論《人工智慧法》制度設計之重點。Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance), 2022 O.J. (L 152) 1; Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance), 2022 O.J. (L 277) 1; Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on

針對歐盟執委會所發布之2021年版草案，歐盟理事會與歐洲議會先後提出修正版草案與意見。2022年12月6日，歐盟理事會表達其對於2021年版草案的意見與立場[136]，主要建議修正內容包含：限縮AI系統定義[137]、增加高風險AI系統供應鏈中各方角色的責任與義務[138]、增加真實世界測試之規範[139]，以及擴增通用目的人工智慧系統（general purpose AI systems）的章節[140]。2023年6月14日，歐洲議會表決通過2023年版草案[141]，主要建議增修之內容包括：增

harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA relevance), 2023 O.J. (L 2023/2854); *Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive)*, explanatory memorandum 1, at 1-2, COM (2022) 496 final (Sept. 28, 2022), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0496; Council Directive 85/374/EEC of July 25, 1985, on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products, 1985 O.J. (L 210) 29; European Commission, *Questions & Answers: AI Liability Directive*, point 6 (Sept. 28, 2022), https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_22_5793/QANDA_22_5793_EN.pdf.

136 Draft Artificial Intelligence Act 2022 ver.; *see also* Council General Approach No. 14954/22 of 25 Nov. 2022, https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf.

137 歐盟理事會建議將欲管制的AI系統定義限縮，將其中提及AI系統「所使用的方法」明確限定為「使用機器學習方法以及基於邏輯與知識方法」（using machine learning and/or logic- and knowledge based approaches）之系統。詳參 Council General Approach No. 14954/22 of 25 Nov. 2022, at 4, https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf; Draft Artificial Intelligence Act 2022 ver. art. 3 (“‘artificial intelligence system’ (AI system) means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;”).

138 Draft Artificial Intelligence Act 2022 ver. art. 23a.

139 Draft Artificial Intelligence Act 2022 ver. arts. 54a & 54b.

140 Draft Artificial Intelligence Act 2022 ver. arts. 4a, 4b, and 4c.

141 Draft Artificial Intelligence Act 2023 ver.

補立法目的[142]、擴大管制範圍[143]、強化透明度義務[144]、針對生物特徵相關AI系統設立更嚴格的管制[145]、納入更多禁止的AI使用情狀[146]、增加歐盟人工智慧局（European Artificial Intelligence Office，下稱AI Office）相關規範[147]，以及加重違反本法之罰鍰金額[148]。

三方經協商並於2023年12月達成政治協議後[149]，《人工智慧法》最終版條文以2021年版草案為核心，並綜合參採歐盟理事會與歐洲議會意見，於2024年7月12日正式發布於歐盟官方公報[150]。《人工智慧法》最終版條文架構有九大部分：定義與適用範圍（第1條至第4條）、禁止使用的AI系統（第5條）、高風險AI系統（第6條至第49條）、針對特定系統／情狀之透明度義務（第50條）、通用目的AI模型（第51條至第56條）、監理沙盒與中小企業（第57條至第63條）、組織架構與管理措施（第64條至第94條）、行為準則（第95條與第96條）、保密、罰則與其他（第97條至第113條）。

《人工智慧法》於歐盟官方公報發布20日後正式生效（即2024年8月1日）。不過，《人工智慧法》實際上並不會於2024年8月1日起立即施行，而是會依第113條之規定分階段性施行各項規範，直至

142 Draft Artificial Intelligence Act 2023 ver. art. 1.
143 Draft Artificial Intelligence Act 2023 ver. art. 2.
144 Draft Artificial Intelligence Act 2023 ver. art. 52(1) & (3).
145 Draft Artificial Intelligence Act 2023 ver. art. 5(1)(db), (dc), and (dd).
146 Draft Artificial Intelligence Act 2023 ver. art. 5.
147 Draft Artificial Intelligence Act 2023 ver. arts. 56-58.
148 Draft Artificial Intelligence Act 2023 ver. art. 71.
149 European Parliament Press Release, Artificial Intelligence Act: Deal on Comprehensive Rules for Trustworthy AI (Dec. 9, 2023), https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.
150 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 (Artificial Intelligence Act) (Text with EEA relevance), 2024 O.J. (L 2024/1689).

2026年8月2日才會全面施行各項規範。此一分階段施行之設計考量包括基本原則的重要規範效果，以及禁止的AI使用情狀的管制急迫性等因素。依施行順序來看，2025年2月2日起將先施行第一章（通則）與第二章（禁止的AI使用情狀）、同年8月2日施行組織管理、保密、罰則以及通用目的AI模型等相關條文。而就第6條第1項與相關義務等針對高風險AI系統特定分類規範部分，則延至2027年8月2日始施行[151]。

二、重要規範與制度設計

從制度設計角度觀之，《人工智慧法》採取以風險為基礎的管制取徑（risk-based regulatory approach），依據風險高低，制定不同程度的管制門檻及權利義務規範[152]。然值得注意者為，《人工智慧法》中風險高低的判斷依據，不同於傳統以風險為基礎的管制取徑、聚焦在特定的技術本質或科學證據等，而係針對AI系統作為經濟市場中的產品或服務，在不同場域的「使用情狀」（practices）所可能產生的風險[153]。此外，為強化對AI系統之管制以避免對人民產生的不利影響，《人工智慧法》於AI系統設計開發階段與使用階段，即定有AI系統的透明度的規範與義務。相較於GDPR以肯認各項新興「數位權利」（digital rights）為核心，如被遺忘權與資料主體近用權等[154]，《人工智慧法》並未明示處理各項管制措施與基本權利之直接關係，而係以「使用情狀」作為AI管制規範設計之主軸。此種管制作法在實行時，有可能在具體風險、管制措施、以及

151 Artificial Intelligence Act art. 113.

152 Giulia Geneletti, *[Interview] On the Ongoing EU AI Act Trilogues and Specific Provisions Covering the Regulation of Foundation Models (FMS): Questions to Professor Hacker*, SCIENCESPO (Nov. 28, 2023), https://www.sciencespo.fr/public/chaire-numerique/en/2023/11/28/interviewtthe-ongoing-ai-act-negotiations-questions-to-prof-hacker/.

153 European Council Press Release IP/986/23, *supra* note 5.

154 *See, e.g.*, GDPR arts. 12-22.

其欲保障之基本權利之三角關係間，造成不小的解釋空間與適用模糊地帶，可能因此賦予主管機關或法院更多裁量空間，更將進一步提高前言（recitals）之重要性與影響力[155]。本文以下將討論聚焦《人工智慧法》之正式條文內涵[156]，探討其重要規範與制度設計之意義，並於必要時補充或比較歐盟理事會與歐洲議會在立法過程中所提出之建議，以資參照。

（一）定義與適用範圍

參酌OECD對「AI系統」的定義[157]及近年生成式AI之新興發展，《人工智慧法》將AI系統定義為「一個具備不同程度自主性而運作並能於部署後展現適應性的機器（machine-based）系統[158]，基

155 前言在歐盟法規解釋適用中的功能與地位毋須贅述，惟AI領域之技術複雜、發展迅速（從生成式AI對於歐盟《人工智慧法》立法過程中帶來的挑戰可見一斑），如何在法規施行實務中借力前言以取得管制平衡（提供適當管制彈性空間，但避免裁量範圍過大而造成管制不確定性），實為歐盟《人工智慧法》未來重要挑戰之一。

156 另補充說明，關於AI的產品責任管制，雖為重要的管制面向，但並非《人工智慧法》直接規範對象，因此不落入本文探討之範圍。AI的產品責任將由歐盟執委會後續透過「指令」而非「規則」來管制，主要應係考量到當今歐盟產品責任即依據《1985年產品責任指令》並仰賴各會員國各自實行。此外，有學者指出歐盟使《1985年產品責任指令》與「人工智慧責任指令草案」並行適用並非最佳選擇，而應修訂一部涵蓋AI的產品責任指令。*See* Philipp Hacker, *The European AI Liability Directives—Critique of a Half-Hearted Approach and Lessons for the Future* 1, 66 (Working Paper, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4279796.

157 OECD, *supra* note 76, at 7 應注意者為，OECD《人工智慧建議》初於2019年發布，並於2024年修訂，二版本對AI定義不盡相同。2019年版本對AI之定義為：“An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.”；而2024年版則為：“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.”

158 Artificial Intelligence Act recital 12 (“... The term ‘machine-based’ refers to the fact that AI systems run on machines. ...”).

於明示或隱含的目標，從其接收之輸入（input）推論如何產出如預測、生成內容、建議或決策等能影響實體或虛擬環境之結果（output）[159]」。相較於2021年版草案對於AI系統相對具體且限縮的定義[160]，《人工智慧法》最終條文擴張AI系統所涵蓋的範圍，文字上從單純的「軟體」擴張為機器系統、將「人類訂定的目標」改為「明示或隱含的目標」、從「環境」修改為「實體與虛擬環境」，以及新增不同程度的自主性運作、適應性與生成內容等用語。

因應大型語言模型（large language model, LLM）與其他各種生成式AI應用的興起，本法新增「通用目的人工智慧模型」（general-purpose AI model）與「通用目的人工智慧系統」（general-purpose AI system）的定義。「通用目的人工智慧模型」的定義為「不論銷售方式為何，均具有顯著通用性及勝任廣泛不同任務的能力，且可整併至各種下游系統或應用之AI模型（包括使用巨量資料進行大規模自我監督訓練之AI模型），但不涵蓋用於銷售前之研究、開發或原型設計活動之AI模型」[161]。而「通用目的人工智慧系

159 Artificial Intelligence Act art. 3(1) (“‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”). 應注意者為，其中自主性（autonomy）、機器（machine-based）、適應性（adaptiveness），乃至於推論（infer）等用語，均在技術上與法律上有待進一步明確化，單從條文文字並無法有精確清楚的指涉範圍。

160 Draft Artificial Intelligence Act 2021 ver. art. 3(1) (“‘artificial intelligence system’ (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with”).

161 Artificial Intelligence Act art. 3(63) (“‘general-purpose AI model’ means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of

統」則指「基於通用目的人工智慧模型，具備應用於多種用途之能力，既可直接使用，亦可整併至其他人工智慧系統中之人工智慧系統」[162]。此二者概念雖有相似之處，但定義與適用上並不相同。本法依據AI系統之應用目的是否特定，以及是否具備廣泛用途應用能力，將其區分為一般常見具特定目的之AI系統（通常因其訓練資料來源與範圍之故而限應於特定專業領域及任務類型），以及無特定目的且可廣泛應用之「通用目的人工智慧系統」[163]。「通用目的人工智慧模型」則非AI系統，而是可進一步開發成AI系統的模型。值得留意者為，AI系統之風險高低係取決於其使用情狀（見本節後述），AI系統是否具特定目的與其風險高低並無直接關聯。「通用目的人工智慧系統」依據其使用情狀，可能分類為高風險AI系統、亦可能為禁止的AI使用情狀或其他非高風險AI系統[164]。

《人工智慧法》最終條文採納歐洲議會2023年之修訂建議，於第1條增列歐盟在AI發展中所應維護的核心價值，即AI應具備「以人為本」與「可資信賴」等特質，且在推動AI創新發展的同時，應避免AI對安全、健康、基本權、民主、法治與環境造成有害影響，並強調對於中小企業發展之重視[165]。關於法規適用範圍，本法第2條建立在2021年版草案基礎[166]上並綜合歐盟理事會與歐洲議會之建

---

downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market").

162 Artificial Intelligence Act art. 3(66) ("'general-purpose AI system' means an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems").

163 需注意者為，此所稱「通用目的人工智慧系統」(general purpose AI system)，並非等同可通過圖靈測試的強AI或通用AI（strong AI/artificial general intelligence)。多數認為強AI或通用AI尚未被發明，本法「通用目的人工智慧系統」中的「通用」係相對於具明確特定目的的AI系統，「通用目的人工智慧系統」得廣泛地應用於不同情境。

164 Artificial Intelligence Act recitals 85 & 161.

165 Artificial Intelligence Act art. 1.

166 2021年版草案之相關條文規定：「所有於歐盟上市或提供服務的AI系統供應

議，予以詳細規範：1.於歐盟上市銷售或提供服務的AI系統與通用目的AI模型供應商，不論該供應商是否設立於或位於歐盟境內；2.設立於或位於歐盟境內的AI系統部署者（deployer）[167]；3.位於歐盟境外但其AI系統產出之結果運用於歐盟境內之供應商與部署者；4.AI系統的進口商（importers）與分銷商（distributors）；5.以自己名義或商標將其產品與AI系統一同上市銷售或提供服務的產品製造商（product manufacturers）；6.設立於歐盟境外之供應商授權代表（authorised representatives）以及7.於歐盟境內受影響者，均有本法之適用[168]。第2條另針對各個會員國就國家安全事項的AI應用，以及第三國政府或國際組織之AI應用設有排除適用規定[169]。最後，為鼓勵創新發展，在以下情狀亦排除適用，包含AI系統上市銷售前或提供服務前的研究、測試與開發活動（若是真實世界測試則不排除適用），以及非通用目的AI模型的免費開源AI組件（components）[170]。

關於《人工智慧法》與其他可能涉及AI管制之相關法規的關係，本法對於已存在保障制度之權利多未直接強化規範，而是競合適用，如GDPR及其他歐盟個人資料保護規範、消費者保障與產品

商、在歐盟境內的AI系統使用者，以及在歐盟境外但其AI系統所產出之結果位於歐盟境內之供應商與使用者」。Draft Artificial Intelligence Act 2021 ver. art. 2(1).

167 關於「部署者」（deployer），最初於2021年版草案中係以「使用者」（user）稱之，在採納2023年歐洲議會之修正建議後，2024年正式文字改為「部署者」（deployer）。雖未修改使用者／部署者之定義（‘deployer’ means a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity），但應可理解歐盟係為避免「使用者」被限縮在終端使用者（end-user），因此將用字修改為「部署者」，使其概念更全面地包含所有使用AI系統的人與單位。

168 Artificial Intelligence Act art. 2(1).

169 Artificial Intelligence Act art. 2(3) & (4).

170 Artificial Intelligence Act art. 2(8), (10), and (12). 關於免費且開源的AI組件（components），若其被作為以下類型AI系統之部分或作為該AI系統時仍適用本法：高風險AI系統、禁止的AI使用情狀以及本法第四部分（Chapter IV）於特定系統／情狀具透明度義務的AI系統。

安全相關規範，以及各會員國對於勞工權利更加有利之保障規範等[171]。此外，就生物識別資料（biometric data）部分，本法援引GDPR對個人資料、特殊類型之個人資料與剖繪的定義，並新增了數項因應近期新興科技發展的用語與定義，包括深偽、廣泛侵害（widespread infringement）、監理沙盒、關鍵基礎建設、真實世界測試（testing in real-world conditions）等[172]。

（二）AI Office之建立與組成

為發展歐盟在AI領域的專業與能力，歐盟執委會於2024年1月通過決議建立AI Office[173]，旨在有效落實《人工智慧法》的施行、推動可信任AI的發展，以及促進國際與跨機關之合作[174]，各會員國則應協助AI Office之運行[175]。此外，AI Office接手了歐盟執委會於2023年11月發起的AI承諾書（AI Pact）計畫，鼓勵私部門於本法施行前自願試行本法與提供意見，截至2024年9月已有逾百家科技、電信、醫療、銀行、汽車等領域的跨國與中小型企業簽署承諾書以依循《人工智慧法》相關原則、辨識可能被歸類為高風險的AI系統，以及強化相關人員的AI識讀能力（literacy）[176]。除AI Office外，為避免歐盟會員國間碎片化的管制，並協助各會員國落實相關規範，《人工智慧法》在第七章（治理機構）增建三個歐盟層級之單位，分別為：人工智慧委員會（European Artificial Intelligence

171 Artificial Intelligence Act art. 2(7), (9), and (11).

172 Artificial Intelligence Act art. 3.

173 Commission Decision of 24 January 2024, Establishing the European Artificial Intelligence Office, 2024 O.J. (C 2024/1459).

174 European Commission, *European AI Office*, https://digital-strategy.ec.europa.eu/en/policies/ai-office#ecl-inpage-the-structure-of-the-ai-office (last visited July 20, 2024).

175 Artificial Intelligence Act art. 64.

176 European Commission, *AI Pact*, https://digital-strategy.ec.europa.eu/en/policies/ai-pact (last visited July 20, 2024); and European Commission Press Release, Over a Hundred Companies Sign EU AI Pact Pledges to Drive Trustworthy and Safe AI Development (Sept. 25, 2024), https://ec.europa.eu/commission/presscorner/detail/en/ip_24_4864.

Board)、諮詢論壇(Advisory Forum)以及「獨立專家科學小組」(Scientific Panel of Independent Experts)[177]。

首先,人工智慧委員會由各會員國之代表組成,歐盟資料保護監督機關(European Data Protection Supervisor)則為常設觀察員,AI Office亦應出席委員會會議但無投票權,而其他機構、組織或專家則視個案情狀受邀參與委員會會議[178]。此委員會主要任務為協調與輔助各會員國權責機構落實本法、處理執行過程所生的疑問、提出對本法的修正意見以及相關準則與規範的建議、製作本法執行成效的年度報告、向歐盟執委會建議關於禁止的AI使用情狀,與修訂高風險AI系統清單[179]。諮詢論壇則由產業、新創企業、中小型企業、公民團體與學術機構等利害關係團體之多方代表組成,主要任務係提供歐盟執委會與人工智慧委員會所需之專業知識,且每年應至少舉行兩場會議並公開其年度活動報告[180]。最後,獨立專家科學小組則是由歐盟執委會負責選任具備相關專業與公正客觀立場之專家,主要負責協助AI Office並提供諮詢,特別是通用目的AI模型之風險管制與檢測、跨境市場監測及相關法律遵循事項[181]。

(三)人工智慧系統的風險分級規範

為將管制資源集中於高風險AI系統且確保管制必要性與效率,《人工智慧法》採取以風險為基礎的管制取徑進行制度設計,依據AI系統風險高低分級並分別課予不同程度義務:1.具不可接受的風險(unacceptable risk)之AI系統使用情狀為「禁止的AI使用情狀」(prohibited AI practices),訂有鉅額違反罰則;2.具高風險(high risk)之AI系統為「高風險AI系統」(high-risk AI systems),訂有諸

177 Artificial Intelligence Act recital 148.
178 Artificial Intelligence Act art. 65(2).
179 Artificial Intelligence Act art. 66.
180 Artificial Intelligence Act art. 67.
181 Artificial Intelligence Act art. 68.

多AI系統上市銷售前後或使用前後的義務；以及3.具低或微小風險（low or minimal risk）之AI系統屬「其他非高風險AI系統」（other than high-risk AI systems），未直接訂有具強制拘束力的管制措施[182]。需注意者為，《人工智慧法》於相關條文特別區分「使用情狀」與「系統」之不同，例如第5條明文禁止特定部署者或使用場域下AI系統的特定AI「使用情狀」，並非禁止特定AI「系統」。換言之，即使是同一個系統，仍會因使用情狀不同而決定其是否落入第5條的禁止範圍、以及是否屬高風險AI系統之判定。舉例而言，遠端生物辨識系統本身雖屬於高風險AI系統[183]，然若一即時遠端生物辨識系統基於執法目的而於公開場所使用、且欠缺嚴格必要性時，則屬於第5條所禁止的AI使用情狀[184]。又如，用於執法單位或移民與邊境管制單位之AI測謊系統，可能被歸類為高風險AI系統[185]，但若係由前述單位以外之人所使用，則非屬高風險AI系統。

---

182 關於AI系統之風險分級，歐盟執委會在其關於《人工智慧法》網頁上之說明將風險分為四個層次：「不可接受的風險」（unacceptable risk）、「高風險」（high risk）、「有限風險」（limited risk）與「微小或無風險」（minimal or no risk），並分別將其對應至《人工智慧法》的第5條「禁止的AI使用情狀」、第6條「高風險AI系統」、第50條「供應商與部署者於特定AI系統之透明度義務」與第95條「其他非高風險AI系統」。但若按此四個層次風險分級方式，將可能產生第50條透明度義務（有限風險）獨立於高風險AI系統與其他非高風險AI系統之誤解，與法條原意不盡相同。細究《人工智慧法》條文設計與2021年版草案解釋備忘錄（explanatory memorandum）之脈絡，《人工智慧法》法條適用範圍中之風險分級應為三個層次，另加上受第50條透明度義務所規範之特定AI系統。即歐盟執委會所稱之「有限風險」非指風險之高低，係指特定AI系統在使用時可能產生操縱或捏造訊息之風險，而並非因其風險分級介於「高風險」與「微小或無風險」之間。因此《人工智慧法》於第50條透明度義務所適用之特定AI系統同時可能歸類於高風險系統或其他非高風險系統，而有相應規範之遵守。基此，為避免對《人工智慧法》之條文適用產生誤解，本文以法條文本為據，以三個風險層次區別加上透明義務之框架進行討論。European Commission, *AI Act*, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai (last visited Sept. 26, 2024); Draft Artificial Intelligence Act 2021 ver. explanatory memorandum 5.2.2; Artificial Intelligence Act recital 26.

183 Artificial Intelligence Act annex III point 1(a).

184 Artificial Intelligence Act art. 5(1), first subparagraph, point (h).

185 Artificial Intelligence Act annex III point 6(b) & point 7(a).

綜觀歐盟執委會於2021年版草案、歐盟理事會與歐洲議會分別於2022年與2023年所提出意見與立場，以及《人工智慧法》之最終條文，可發現在短短幾年的立法過程中，歐盟對於禁止的AI使用情狀之管制立場愈趨嚴謹，並積極應對近來新興AI發展動態。例如，《人工智慧法》將禁止的AI使用情狀種類自最初的4種增至最終的8種情狀。同時，考量到實務上對於AI系統使用需求增加，《人工智慧法》針對禁止的AI使用情狀進一步加以細分，增訂例外可使用之情狀，以試圖在「管制AI風險」與「善用AI功能」間取得平衡，如「於工作場所或教育機構使用情緒推論AI系統」雖歸類於禁止的AI使用情狀，然若基於醫療或安全目的而使用，則屬例外可使用之高風險AI系統。本文以下將進一步闡明《人工智慧法》中「禁止的AI使用情狀」、「高風險AI系統」與「其他非高風險AI系統」之各項具體情狀。

1. 禁止的 AI 使用情狀

依據《人工智慧法》第5條，禁止的AI使用情狀共有八大情狀（詳見下表1），據此明文劃設管制紅線，以維護歐盟在AI發展與應用中最核心的原則與價值。相對於最初2021年版草案僅概略列出禁止之使用情狀[186]，《人工智慧法》除使用語更臻精確外，更大幅擴張管制範圍，如人類生物數據之系統，包括剖繪、情緒辨識、生物分類，以及臉部辨識資料庫等。

第一種禁止情狀，為上市銷售、提供服務或使用運用潛意識技術、操控或欺騙之技術，以削減特定人或群體為知情同意之能力並藉此扭曲特定人或群體行為，且使其做出原本不會做之行為，並對

186 2021年版草案所訂禁止的AI使用情狀僅包含以下四種情狀：運用潛意識技術扭曲人類行為並致生損害之系統、傷害特定弱勢群體之系統、可能產生不平等待遇之公部門社會信用評分系統，以及即時遠端生物辨識系統。Draft Artificial Intelligence Act 2021 ver. art. 5(1).

其造成或可能造成重大損害之AI系統[187]。第二種情狀，為上市銷售、提供服務或使用利用特定人或群體之年齡、身心障礙、特定社會上或經濟上之弱勢，以扭曲特定人之行為，並因此對特定人或第三者造成或可能造成重大損害之AI系統[188]。第三種情狀，為上市銷售、提供服務或使用基於特定人或群體的社會行為，或已知的、推測的或預測的個人或人格特徵，進行評估或分類，並基於此社會評分（social score）而對特定人或群體施加有害或不利的對待，且該對待與最初資料生成或蒐集無關、欠缺正當理由或不符比例原則之AI系統[189]。第四種情狀，則是上市銷售、提供服務或使用完全基於剖繪或對於個人特徵之評估，以進行犯罪風險評估或預測之AI系統；但若此AI系統係用於輔助（support）進行人類評估（human assessment）以確認特定人是否涉及特定犯罪，且已存在經證實之客觀事實而知該特定人與該特定犯罪有直接關聯時，則不在此限[190]。第五種情狀，係指上市銷售、提供服務或使用從網路或閉路監視器畫面中，無差別抓取臉部影像以建立或擴展臉部辨識資料庫之AI系統[191]。第六種情狀，乃為於工作場所或教育機構所使用之情緒推論（infer emotions）AI系統，但若係基於醫療或安全目的而使用，則不在此限[192]。第七種情狀，則是上市銷售、提供服務或使用「針對特定群體之生物資料對其進行分類，以推測其種族、政治觀點、工會成員身分、宗教與哲學信念，或性生活與性傾向」之AI系統；但若係針對依法取得之生物資料庫進行標記與過濾，則不在此限[193]。最後的第八種情狀，為基於執法目的而於公開場所使用之「即時」遠端生物辨識系統（the use of “real-time” remote biometric

187 Artificial Intelligence Act art. 5(1), first subparagraph, point (a).
188 Artificial Intelligence Act art. 5(1), first subparagraph, point (b).
189 Artificial Intelligence Act art. 5(1), first subparagraph, point (c).
190 Artificial Intelligence Act art. 5(1), first subparagraph, point (d).
191 Artificial Intelligence Act art. 5(1), first subparagraph, point (e).
192 Artificial Intelligence Act art. 5(1), first subparagraph, point (f).
193 Artificial Intelligence Act art. 5(1), first subparagraph, point (g).

identification systems in publicly accessible spaces for the purposes of law enforcement），僅限用於以下目的並具嚴格必要性（strictly necessary）[194]：(1)為搜尋特定犯罪被害者或失蹤者；(2)為避免對生命或安全特定具體且迫切之威脅，以及現時或可預見之恐怖攻擊威脅；及(3)為尋找或辨識涉犯本法附件二所列犯罪類型且依據該會員國刑法法定刑4年以上有期徒刑之犯罪嫌疑人[195]。

表1 《人工智慧法》「禁止的AI使用情狀」與例外規定

禁止的AI使用情狀	
一、潛意識技術	上市銷售、提供服務或使用運用潛意識技術、操控或欺騙之技術，以削減特定人或群體為知情同意之能力並藉此扭曲特定人或群體行為，且使其做出原本不會做之行為，並對其造成或可能造成重大損害之AI系統
二、針對弱勢族群	上市銷售、提供服務或使用利用特定人或群體之年齡、身心障礙、特定社會上或經濟上之弱勢，以扭曲特定人之行為，並對特定人或第三者造成或可能造成重大損害之AI系統
三、社會評分與不利對待	上市銷售、提供服務或使用基於特定人或群體的社會行為，或已知的、推測的或預測的個人或人格特徵，進行評估或分類，並基於此社會評分而對特定人或群體施加有害或不利的對待，且該對待與最初資料生成或搜集無關、欠缺正當理由或不符比例原則之AI系統

194 Artificial Intelligence Act art. 5(1), first subparagraph, point (h).

195 值得注意者為，《人工智慧法》針對「基於執法目的而於公開場所使用『即時』的遠端生物辨識系統」之例外可使用之情狀，額外訂有更嚴謹的要求，執法單位須事前經各會員國之司法或獨立行政機關授權始得為使用之，且該授權應以系統已完成本法第27條之「高風險AI系統基本權利影響評估」（fundamental rights impact assessment for high-risk AI systems），並依據本法第49條註冊至歐盟資料庫為前提。但若於緊急情況，縱未取得授權亦可使用，但應最晚於24小時內向各會員國之司法或獨立行政機關請求授權，若於緊急情況且未有無故拖延之情形時，縱使系統未完成註冊亦可使用。Artificial Intelligence Act art. 5(3), first subparagraph & art. 5(2), second subparagraph.

四、完全基於剖繪之犯罪風險評估	上市銷售、提供服務或使用完全基於剖繪或對於個人特徵之評估，以進行犯罪風險評估或預測之AI系統 ＊例外：若此AI系統係用於輔助進行人類評估以確認特定人是否涉特定犯罪，且已存在經證實之客觀事實而知該特定人與該特定犯罪有直接關聯時，則不在此限
五、臉部辨識資料庫	上市銷售、提供服務或使用從網路或閉路監視器畫面中，無差別地抓取臉部影像以建立或擴展臉部辨識資料庫之AI系統
六、情緒識別	於工作場所或教育機構使用之情緒推論AI系統 ＊例外：若係基於醫療或安全目的而於上述場域使用，則不在此限
七、生物分類	上市銷售、提供服務或使用「針對特定群體之生物資料對其進行分類，以推測其種族、政治觀點、工會成員身分、宗教與哲學信念，或性生活與性傾向」之AI系統 ＊例外：若係針對依法取得之生物資料庫進行標記與過濾，則不在此限
八、即時遠端生物辨識	基於執法目的而於公開場所使用之「即時」遠端生物辨識系統 ＊例外：但基於以下目的且具嚴格必要性者，可使用之：(1)為搜尋特定犯罪被害者，或失蹤者；(2)為避免對生命或安全特定具體且迫切之威脅，以及現時或可預見之恐怖攻擊威脅；(3)為尋找或辨識涉犯本法附件二所列犯罪類型且依據該會員國刑法法定刑4年以上有期徒刑之犯罪嫌疑人

資料來源：本文作者編製。

2. 高風險 AI 系統

《人工智慧法》第3條並未針對「高風險AI系統」進行一般性的定義，而係由第6條列舉屬於「高風險AI系統」的應用情狀及特徵，並分為兩大類型，其一為本法附件一所列之歐盟調和立法（Union harmonisation legislation）所涵蓋、且用於產品或其安全元

件之AI系統[196]；其二為落入附件三所列情狀之AI系統[197]，包括(1)用於「關鍵基礎設施」與「司法與民主程序」者，因其涉及重大公共利益風險並可能對社會產生嚴重危害；(2)用於「社會福利或必要服務」、「法律執行」與「移民與邊境管制」等涉及公部門權力行使者，因其可能對基本權利造成不當影響；以及(3)透過AI系統進行「生物特徵識別」、「教育與職業訓練」與「員工管理」者，因其涉及個人資料與剖繪而可能於私領域產生不平等待遇[198]。另需注意，用於以上三種情狀之AI系統雖稱為高風險AI「系統」，實際上是否構成高風險則與其特定運用「情狀」直接相關（詳參下表2）。

考量到附件三所列之高風險AI系統實際上未必會對本法所欲保障之權利產生顯著風險、或對決策結果產生實質影響力[199]，《人工智慧法》最終條文於第6條第3項增訂高風險AI系統之例外[200]：當附件三所列AI系統不會對自然人之健康、安全與基本權利產生顯著風險或對決策結果產生實質影響，且符合以下任一情形時，則應認其非屬高風險AI系統，包括用於執行狹義的程序性工作、用於改進人類已完成之成果、用於偵測決策模式以及決策是否偏離先前的決策模式、用於進行附件三所列AI系統評估之前置作業。然須注意者為，第6條第3項第3款又增訂例外中之例外，將使用剖繪技術之AI系統排除前述例外規定之適用，並將其歸類為高風險AI系統。

此外，可能由於技術的複雜性與應用多樣性，或是立法技術問題，附件三所列之高風險AI系統中，亦包括部分非屬高風險AI系統的例外情狀。舉例而言，用於進行「生物特徵識別」（biometrics）之AI應用情狀泛指依據歐盟與各會員國法規而獲允許使用之遠端生

196 Artificial Intelligence Act art. 6(1).
197 Artificial Intelligence Act art. 6(2).
198 Artificial Intelligence Act annex III.
199 Artificial Intelligence Act recital 53.
200 Artificial Intelligence Act art. 6(3) & recital 53.

物辨識系統（remote biometric identification systems）、以敏感或受保護特徵等推論之生物分類（biometric categorisation）系統，以及情緒辨識（emotion recognition）系統，但其中遠端生物辨識系統若僅用於辨識特定個人是否為其所宣稱之身分，則不被歸類為高風險AI系統。又如「司法與民主程序」（administration of justice and democratic processes）類別中雖包括司法機關用於解釋與適用法律之系統，以及用以影響選舉或公投之系統，但後者在其所產出的結果並未對任何人產生直接影響時，則不屬於高風險AI系統（如政黨用於行政或改善其組織活動之工具）。此外，在「社會福利或必要服務」類別中用於評估並建立信用分數之AI系統，若其使用目的係偵測金融詐騙，亦非屬高風險AI系統。

表2 《人工智慧法》附件三高風險AI系統

高風險AI系統	
一、生物特徵識別（以歐盟與會員國相關法規允許使用為限）	遠端生物辨識系統（不包括僅用於辨識特定個人是否為其所宣稱之人）
	以敏感或受保護特徵等推論之生物分類AI系統
	情緒辨識AI系統
二、關鍵基礎設施	用於交通或關鍵基礎數位設施管理與運作之安全元件，或被用於供應水、電、瓦斯或暖氣之安全元件的AI系統
三、各級教育與職業訓練機構應用	用於決定是否錄取個人至特定教育或職業訓練機構之AI系統
	用於評價個人學習成果及促進學習／職業訓練進程之AI系統
	用於將個人分級以決定其能接受或近用教育程度之AI系統
	用於監督與偵測學生於測驗時違規行為之AI系統
四、員工管理	用於招募特定職位時用以投放招募廣告、篩選與評價應聘者之AI系統
	用於做出影響勞雇關係、升遷、終止等契約條款相關決策之AI系統

	用於基於受僱者行為或特徵而進行工作分配之AI系統
	用於監控與評價受僱者表現之AI系統
五、必要服務與社會福利	公部門用於評估人民能否近用必要服務或社會福利之AI系統（包括醫療服務），以及授予、減免、撤銷、收回相關福利或服務之作為
	用於評估並建立信用分數之AI系統（不包括用於偵測金融詐騙者）
	用於人壽保險與健康保險核保之風險與金額評估之AI系統
	用於評估與排序緊急救助優先順序之AI系統（如消防、警察、醫療急救等緊急服務）
六、法律執行（執法機關或受其委託行使公權力者／以歐盟與會員國相關法規允許使用為限）	用於評估特定人成為犯罪被害者風險之AI系統
	用於測謊或類似功能之AI系統
	用於偵查階段評估犯罪證據的可信度之AI系統
	用於「非完全」基於剖繪的犯罪風險評估或再犯風險評估之AI系統
	用於偵查、調查與起訴階段之剖繪AI系統
七、移民與邊境管制（權責機關或受其委託行使公權力者／以歐盟與會員國相關法規允許使用為限）	用於測謊或類似功能之AI系統
	用於評估意圖進入或已進入成員國領土之人所生安全風險、非法移民風險或公共衛生風險之AI系統
	用於協助審查庇護、簽證、居留許可以及相關申訴之AI系統
	用於處理移民、庇護或邊境管制事務時，檢測、識別或辨識自然人之AI系統
八、司法與民主程序	司法機關或受其委託提供輔助之其他單位用於確認事實、解釋與適用法律等AI系統，包含訴訟外紛爭解決機制之類似應用
	用於影響選舉或公投結果或具體投票行為之AI系統（不包括系統所產出的結果並未對任何人產生直接影響者）

資料來源：本文作者編製。

值得注意者為，其中最敏感的面向，亦即生物特徵識別的AI系統與使用情狀管制，若比較分析不同草案用語及範圍，可見歐盟立法者間似乎並未達成高度共識，因而有諸多折衷空間。整體而言，2021年版草案與2022年歐盟理事會建議版本之管制力道相對寬鬆，歐洲議會於2023年之建議草案版本則最為嚴格，《人工智慧法》最終條文則係折衷綜合三者之建議（比較表參下表3）。首先，2021年版草案與2022年歐盟理事會之建議版本大致相同，皆將「遠端生物辨識系統」與「執法機關用於偵測情緒狀態之AI系統」訂為高風險AI系統，並禁止在欠缺嚴格必要性之情形下基於執法目的而於公開場所使用即時遠端生物辨識系統。2023年版草案則嚴格許多，除將「生物辨識系統」訂為高風險AI系統、且規定「情緒辨識AI系統」不限使用主體皆屬高風險AI系統之外，額外禁止所有「於公開場所使用即時遠端生物辨識系統」，以及新增數項禁止的AI使用情狀，包括(1)未取得司法單位事前授權且不具嚴格必要性之情況下，使用非即時遠端生物辨識系統分析錄影畫面；(2)於執法、邊境、工作場所或教育機構使用情緒推論AI系統，以及(3)生物分類系統。

《人工智慧法》最終條文則綜合前述三個版本之建議，特點有四：其一，由於歐盟或其會員國可能針對生物特徵識別相關AI系統設有更嚴格之限制，為避免《人工智慧法》與其他法規產生矛盾，最終條文將生物特徵識別之高風險AI系統條文限縮於「以歐盟與會員國相關法規允許使用為限」。其二，就生物辨識系統部分，不論即時或非即時的「遠端生物辨識系統」最後皆被列為高風險AI系統，並禁止「於欠缺嚴格必要性且基於執法目的於公開場所使用即時遠端生物辨識系統」；2023年版草案之建議亦獲採納，於第26條增訂部署者應事先取得授權之義務，意即部署者於使用「非即時遠端生物辨識系統」前，應具嚴格必要性且事前取得司法或相關行政單位授權。

其三，2021年版草案原先規定，當執法機關或受其委託行使公權力之機關使用情緒識別系統時，始歸類為高風險AI系統；相較之下，《人工智慧法》最終條文採取更嚴格的取徑，擴大管制範圍且不限使用主體，將所有情緒識別系統納為高風險AI系統，並禁止於工作或教育場所使用情緒推論系統。其四，所有生物分類系統皆列為高風險AI系統，並禁止使用針對特定群體之生物資料對其進行分類並推測其種族、政治觀點、工會成員身分、宗教與哲學信念或性生活與性傾向之AI系統。

表3 「生物識別」相關之高風險AI系統與禁止的AI使用情狀——各版本草案與最終條文比較

「生物識別」相關之AI系統與使用情狀——草案與最終條文比較				
	2021年	2022年	2023年	2024年
高風險AI系統	即時與非即時的遠端生物辨識系統 Annex III point 1(a)	~~即時與非即時的~~遠端生物辨識系統 Annex III point 1(a)	~~即時與非即時的遠端~~生物辨識系統 Annex III point 1(a)	~~即時與非即時的~~遠端生物辨識系統 Annex III point 1(a)
	執法單位用於偵測情緒狀態之系統 Annex III point 6(b)	執法單位用於偵測情緒狀態之系統 Annex III point 6(b)	情緒識別等個人特徵與生物特徵推論系統 Annex III point 1(aa)	情緒識別系統 Annex III point 1(c)
	生物分類系統 Annex III point 1	-	-	以敏感或受保護特徵等推論之生物分類系統 Annex III point 1(b)
禁止的AI使用情狀	欠缺嚴格必要性之情形下，基於執法目的於公開場所使	欠缺嚴格必要性之情形下，執法機關或受其委託行使公	~~使用欠缺嚴格必要性且基於執法目的~~於公開場所使用即時遠端生物辨識系	欠缺嚴格必要性之情形下，基於執法目的於公開場所使用即時遠

	用即時遠端生物辨識系統 Art. 5(1)(d)	權力者基於執法目的於公開場所使用即時遠端生物辨識系統 Art. 5(1)(d)	統 Art. 5(1)(d)	端生物辨識系統 Art. 5(1)(h)
	-	-	欠缺嚴格必要性之情形下，使用非即時遠端生物辨識系統分析錄影畫面 Art. 5(1)(dd)	欠缺嚴格必要性之情形下，使用非即時遠端生物辨識系統 Art. 26(10)
	-	-	於執法、邊境、工作場所或教育機構使用情緒推論系統 Art. 5(1)(dc)	於工作場所或教育機構使用情緒推論系統 Art. 5(1)(f)
	-	-	以敏感或受保護特徵等推論之生物分類系統 Art. 5(1)(ba)	針對特定群體之生物資料對其進行分類，以推測他們的種族、政治觀點、工會成員身分、宗教與哲學信念，或性生活與性傾向之系統 Art. 5(1)(g)

資料來源：本文作者編製。

此外，有關「執法機關或受其委託行使公權力者基於執法目的所使用之AI系統與使用情狀」之分類亦值得注意。綜觀2021年至2024年之草案、建議與最終條文，可見此部分之變動係考量實務上的執法迫切性，在避免剖繪技術對基本權利造成侵害的前提下，相當程度放寬管制以利執法單位運用。例如，《人工智慧法》最終條

文將「偵測深度偽造的AI系統」與「使執法機關能從不同資料來源或格式中，針對大量相關或無關的複雜大型資料集進行搜尋，以從中找出未知的模式或隱藏的關係，並用於分析自然人犯罪之AI系統」自高風險AI系統類別中移除，換言之，此二類系統之使用毋須符合本法高風險AI系統之各項要求。又如，相對於全面禁止「基於剖繪的犯罪風險評估或再犯風險評估之AI系統」，最終條文選擇對之進行細分，將「非完全」基於剖繪的犯罪風險評估或再犯風險評估之AI系統歸類於高風險AI系統，並禁用「完全」基於剖繪或對個人特徵分析以進行犯罪風險評估或預測之AI系統。

3. 其他非高風險 AI 系統

至於其它非高風險AI系統，即「未被禁止或列於高風險之AI系統之AI系統」[201]，《人工智慧法》並未訂定具體管制措施，而係責成歐盟執委會推動行為準則的訂定，以使此類非高風險AI系統亦有相關規範可資遵循，藉此鼓勵所有AI系統落實可信任AI等原則[202]。不過，當非高風險AI系統在符合第50條所列系統或情狀等條件時，仍負有一定透明度義務（詳參下一節說明）。

（四）透明度義務

由於AI系統的各個生命週期都將面臨透明度議題，為避免AI系統對人民產生不利影響，本法特別重視透明度原則，於AI系統設計開發階段與特定使用情狀，對供應商與部署者訂有資訊提供、解釋、揭露與告知的義務。本法明文以「透明度」為主旨之規定，包含第13條針對高風險AI系統的「透明度與對部署者之資訊提供」，以及第50條非針對高風險AI系統的「供應商與部署者於特定AI系統之透明度義務」。此外，《人工智慧法》第86條為請求解釋權，使受

201 Artificial Intelligence Act art. 95.
202 Artificial Intelligence Act art. 95 & recital 165.

到基於高風險AI系統所作成決策影響者得向部署者「請求針對個人決策提供解釋之權利」，亦與理論上所討論之「透明度」直接相關。第13條與第50條規範分別對應到AI系統的設計開發與使用時的透明度議題，第86條則為保障受AI決策影響者（affected person）所設的透明度義務，以下說明之。

首先，第13條「透明度與對部署者之資訊提供」為本法針對高風險AI系統供應商的數項要求之一。高風險AI系統的供應商應提供部署者清楚的系統使用說明，以確保部署者了解該系統的目的、準確度，以及已知或可能的使用情狀。相較於2021年版草案，《人工智慧法》最終條文經彙整歐洲議會與歐盟理事會之建議後，將本條透明度規範更加具體化，針對提供資訊之具體內容有更加完整的規定，使系統的結果具備（一定程度的）可解釋性，俾使供應商與部署者能完成本法制定之各項義務。例如，系統提供解釋等相關資訊之技術能力與特點、使部署者解釋該系統產出結果之所需資訊，以及如何存取與解釋系統運作之紀錄等[203]。

另一方面，第50條「供應商與部署者於特定AI系統之透明度義務」，並非針對高風險AI系統之要求，而係考量到AI系統的冒充與捏造能力在使用時可能衍生大規模錯誤訊息與操縱風險（risks of misinformation and manipulation at scale）[204]而額外制定之規範。因此，本條與第13條不同，其透明度義務並非針對高風險AI系統，而係列舉可能產生操縱風險的情狀（如使用生成影音文字內容之通用目的AI系統、情緒識別系統、生物分類系統等），並要求供應商或部署者確保人類得到適當告知。第50條所涵蓋情狀與相應透明度義務如下[205]：1.若AI系統係用於與人類互動，供應商應確保AI系統與

203 Artificial Intelligence Act art. 13(3)(b)(iv), (vii), and (f).
204 Artificial Intelligence Act recitals 133 & 134.
205 Artificial Intelligence Act art. 50.

人類的互動過程均符合其揭露之方式；2.生成式AI系統與生成式通用目的AI系統之供應商，應確保其生成之內容將以「機器可讀之形式」（machine-readable format）且可被偵測的方式，標示其內容係經AI生成或捏造，且生成之內容包括聲音、圖像、影片與文字；3.情緒辨識與生物分類AI系統的部署者，應向暴露於系統偵測之人進行告知；4.生成或捏造可能產生誤導之深偽內容的部署者，應揭露該內容係由AI生成或捏造，且捏造之內容包括聲音、圖像與影片；以及5.於公布涉公共利益之事務，使用AI生成或捏造之文字之部署者，應揭露該內容係由AI生成或捏造[206]。

而就第86條「請求針對個人決策提供解釋之權利」部分，立法者考量到高風險AI系統可能會做出個人權益產生負面影響之決策結果，因而賦予受影響者向高風險AI系統部署者請求提供清楚且有意義（clear and meaningful）解釋之權利。該法所稱「影響」，係指因該決策對當事人產生法律效果或相似之重大影響，而足認對該當事人之健康、安全等基本權利產生負面影響。部署者應提供之解釋，應包含該AI系統於決策過程中所扮演的角色，以及作成該決策的主要因素（main elements of the decision taken）[207]。前言並指出此解釋應以受影響者得行使其權利之方式為之[208]，藉此使高風險AI系統具備保障受影響者權利所需的透明度。另如同前述，考量到部署者

206 值得注意者為，2023年歐洲議會所提出許多修正建議，嘗試使透明度義務更為具體有效，然並未採納於最終條文中。相關建議之內容包括：應揭露AI決策的負責人身分（若設有人類監督機制）、告知受影響者得行使之請求解釋權利與相關司法救濟程序。以及，若AI系統係用於製作可能產生誤導之深偽多媒體內容，應揭露利用系統製作內容者之姓名，且揭露必須以內容接收者清楚可見之方式標示該內容為非真實（inauthentic）。Draft Artificial Intelligence Act 2023 ver. art. 52(1) & (3).

207 Artificial Intelligence Act art. 86(1). 值得注意者為，關於部署者應提供之清楚且有意義的解釋，2023年歐洲議會之建議修正內容中強調應包括決策過程中的功能、決策主要參考的參數、相關的輸入資料，但並未採納至《人工智慧法》之最終條文。Draft Artificial Intelligence Act 2023 ver. art. 68c(1).

208 Artificial Intelligence Act recital 171.

並非供應商，且並未擁有AI系統的所有設計與開發資訊，同法第13條要求供應商應提供系統相關資訊予部署者，以確保後者能了解系統運作與資料處理方式，並向受影響者提供有意義的解釋[209]。

最後，雖本法僅於第13條與第50條明文規定「透明度」義務，然透明度之規範內涵實並不限於此二條之規範，如前述第86條「請求針對個人決策提供解釋之權利」即與理論上所討論之「透明度」直接相關。本法與透明度概念相關之條文尚包括第43條之符合性評估、第49條之登記制度以及第53條之通用目的AI模型的技術資訊提供。本文將於第參部分第三節一併討論更廣泛的透明度規範內涵、其如何影響AI系統透明度的實踐，以及可能產生之問題。

（五）高風險AI系統供應商、部署者與其他行為者的義務

關於高風險AI系統的要求，主要見於《人工智慧法》第三章第二部分之「高風險AI系統之要求」（requirements for high-risk AI systems）與第三部分之「高風險AI系統供應商、部署者以及其他行為者之義務」（obligations of providers and deployers of high-risk AI systems and other parties），前者規範高風險AI系統在上市銷售或提供服務「前」於開發、建構以及規格方面應符合之各項要求；後者則著重在系統上市銷售或提供服務「後」之應用層面，規範不同行為者在高風險AI系統之生命週期中應負之各項法定義務，相關行為者包括供應商、部署者、進口商、分銷商與授權代表。基此，各行為者之法定義務，可依據上市銷售或提供服務之前後而分為兩種類型，而「供應商」身處銷售及提供服務的節點，乃為兩種類型規範對象之「交集」，因此將相對於其他行為者負擔更多的法律遵循義務。此外，本文認為，風險管理系統、技術文件（technical documentation）、系統紀錄保存、符合性評估、基本權利影響評估

209 Artificial Intelligence Act art. 13(3)(b)(iv), (vii), and (f).

(fundamental rights impact assessment)與人類監督機制(human oversight)等上市銷售或提供服務前之各項管制措施,為《人工智慧法》重要的制度設計,值得進一步分析。

首先,為更全面地掌握高風險AI系統的風險,本法要求供應商於此類系統上市銷售或提供服務前設有風險管理系統,藉此辨識該AI系統於生命週期中所有已知與合理可得而知的風險,並評估是否可能有其他顯著風險的產生,藉此採取相應的風險管理措施[210]。同時,AI系統運作紀錄之保存乃為執行其他條文要求之根本,在本法亦有相關規定,要求供應商在設計與開發高風險AI系統的過程即應開始記錄系統運作過程,以確保所有可能導致風險的情狀、可能使AI系統產生重大修改的各種情狀等均獲確實記錄[211]。此外,供應商應製作與保存技術文件,並提供予國家權責機關(national competent authorities)備查[212],內容應說明該AI系統就本法第三章第二部分所定各項要求之法遵情形,包括資料管理、風險管理系統、資訊的提供、人類監督機制、系統準確度、健全性與網路安全措施等。

再者,為確保AI系統符合本法所定之要求且具有持續依循本法之能力與意願,除前述第三章第二部分所定各項要求外,供應商另應建立「品質管理系統」(quality management system),說明該AI系統之設計、管理、法遵、意外處理、資源管理與可課責性框架之計畫[213]。為確保AI系統於上市銷售或提供服務前符合本法規定,供應商應於系統上市銷售或提供服務前向受認證之第三方符合性評估機構(notified body)提供必要資訊以進行「符合性評估」[214],使

210 Artificial Intelligence Act art. 9. AI系統上市前應接受測試,以決定最佳的風險管理措施。
211 Artificial Intelligence Act art. 12(1) & (2).
212 Artificial Intelligence Act art. 11.
213 Artificial Intelligence Act art. 17.
214 高風險AI系統應先通過符合性評估,始得上市或提供服務。關於符合性評估

其「品質管理系統」之運作及「技術文件」之完整性獲得審查[215]，並登記註冊至歐盟資料庫。

除課予供應商建立前述兩項系統之義務外，《人工智慧法》亦針對高風險AI系統的部署者規定上市銷售或提供服務前之義務，要求在第一次啟用系統前應針對系統進行「基本權利影響評估」[216]，內容涵蓋AI系統的使用目的、使用時間、頻率、可能受到影響的族群或個人及其受到侵害之風險，以及應對與管理措施（包括消弭前述影響與風險的詳盡計畫，以及人類監督與申訴處理等機制）。完成基本權利影響評估後，部署者應將評估結果通知「市場監督機關」（market surveillance authority）。若供應商已先行完成評估且該AI系統實際使用情況並無不同，則部署者可沿用該評估結果；但部署者若於系統開始使用後發現該評估未能反映實際使用情狀，則應重新進行影響評估。

前述義務主要係高風險AI系統於上市銷售前或開始提供服務前所應達成之事項。而就系統上市銷售或開始提供服務後的法規要求部分，首先介紹通報義務。高風險AI系統於上市銷售或開始提供服務後，若產生嚴重意外而違反歐盟保障基本權之相關法規，供應商

的施行，《人工智慧法》要求各會員國設有認證機關（notifying authority），由該認證機關管理各國符合性評估制度的運作。認證機關將受理第三方符合性評估機構之申請，評估其是否符合本法對於「受認證之第三方符合性評估機構」（notified body）的要求並由受認證之第三方符合性評估機構負責受理供應商的申請，並對其高風險AI系統進行符合性評估。對於未通過符合性評估認證之高風險AI系統，各會員國應訂有相應之救濟制度確保供應商的權利。Artificial Intelligence Act arts. 28, 29, 31, and 44.

215 Artificial Intelligence Act art. 43 & annex VII. 供應商亦可採取內部控制的方式，即自行確認其AI系統之品質管理系統與技術文件之要求皆符合本法之要求。Artificial Intelligence Act art. 43 & annex VIII.

216 Artificial Intelligence Act art. 27. 符合以下情狀者，無需進行高風險AI系統基本權利影響評估：本法附件三第二點之使用情狀、部署者為公部門或提供公共服務之私部門、用於評估並建立信用分數之AI系統，以及用於人壽保險與健康保險核保之風險與金額評估與之AI系統。

負有立即向市場監督機關通報之義務[217]。為更有效地減少高風險AI系統可能產生的風險與負面影響，除供應商之義務外，《人工智慧法》採納歐洲議會與歐盟理事會對於部署者義務之增訂意見，要求部署者應指派特定人員進行人類監督[218]，並於有合理理由認為AI系統可能對健康安全或基本權利產生侵害時，應通知供應商、分銷商與市場監督機關[219]，以及在發現重大意外時，先後通知供應商以及與該意外相關之進口商、分銷商與市場監督機關[220]。當部署者對於系統握有完全控制權時，則應確保輸入系統之資料具有足夠代表性[221]。而若系統係使用於工作場所，部署者應與受僱者協商並告知可能受影響的受僱者，且須進行個人資料保護影響評估[222]。另為調查特定犯罪嫌疑人而使用非即時遠端生物辨識AI系統（AI system for post-remote biometric identification）時，部署者應事先取得司法機關或相關權責行政機關之授權，並告知受系統決策影響者[223]。

由於高風險AI系統之價值鏈（value chain）複雜，相關法規要求亦相當嚴格，實務上可能無法要求所有涉及AI系統之行為者均負擔沉重義務。考量供應商為銷售及提供服務的交集點，故制度設計上選擇由「供應商」負擔主要義務，藉此將管制資源集中於單一節點、提高管制效率。不過，由於AI系統生命週期中的各行為者亦可能影響AI系統是否符合本法對於高風險AI系統之義務規定[224]，且為確保最有能力或最適合為特定AI系統負責之人負擔供應商義務，本法因此擬制供應商身分的轉換：於特定情狀下，進口商、經銷

217 Artificial Intelligence Act art. 73(1).
218 Artificial Intelligence Act art. 26(2).
219 Artificial Intelligence Act art. 26(5).
220 若未能通知到供應商，則準用（*mutatis mutandis*）本法第73條，即供應商向市場監督機關通報之義務。Artificial Intelligence Act arts. 26(5) & 73.
221 Artificial Intelligence Act art. 26(4).
222 Artificial Intelligence Act art. 26(7).
223 Artificial Intelligence Act art. 26(10).
224 Artificial Intelligence Act recital 84.

商、部署者以及其他行為者擬制（be considered）為供應商並負擔與高風險AI系統供應商相同之義務[225]，原供應商（initial provider）則不再屬本法所規範之供應商（參下表4）、亦無需再負擔相關義務。相關情狀包括：1.進口商、經銷商、部署者或其他行為者在已上市銷售或提供服務的高風險AI系統上附加其商標或名稱時；2.前開人員對該系統進行重大修改且修改後仍屬高風險AI系統時，以及3.前開人員對非高風險AI系統進行重大修改使其成為高風險AI系統時。雖然原供應商在前述情狀下免除其供應商義務，但仍應負責移交必要資訊，即「原供應商」應提供「擬制供應商」技術文件以及其他與AI系統運作相關之資訊，並應提供擬制供應商完成本法所定義務所需之技術權限與協助[226]。如此制度設計，係考量AI系統生命週期及價值鏈之複雜性，旨在有效集中管制資源而減少行政成本，而潛在挑戰則在於認定過程的技術模糊空間，特別是對所謂「重大修改」的判斷基準。

表4 《人工智慧法》第25條第1項——供應商義務移轉

進口商、經銷商、部署者以及其他行為者之行為	義務移轉	
	原供應商	擬制供應商（進口商、經銷商、部署者或其他行為者）
(1)(a)於已上市銷售或提供服務的高風險AI系統上「附加其商標或名稱」	不再屬本法所定供應商，但仍負	負供應商義務

225 Artificial Intelligence Act art 25.

226 若高風險AI系統為本法附件一第A部分中列出的歐盟調和立法所涵蓋的產品的安全組件，則產品製造商（product manufacturer）亦擬制為高風險AI系統的供應商，並在符合以下情狀時，應負本法第16條之供應商義務：(a)高風險AI系統與產品共同使用該產品製造商的名稱或商標上市，或(b)於該產品上市後，高風險AI系統使用產品製造商的名稱或商標提供服務。Artificial Intelligence Act art. 25(3). 若事實上供應商已清楚聲明其AI系統並未打算被改變為高風險AI系統，則無移交必要資料之義務。Artificial Intelligence Act art. 25(2).

(1)(b)對已上市銷售或提供服務的高風險AI系統進行「重大修改」且修改後仍屬高風險AI系統	提供擬制供應商必要資訊之義務	
(1)(c)對已上市銷售或提供服務的「非」高風險AI系統進行「重大修改」使其成為高風險AI系統		

資料來源：本文作者編製。

（六）通用目的AI模型

「通用目的AI系統」依其使用情狀可能列為高風險AI系統、禁止的AI使用情狀，或其他非高風險AI系統，而「通用目的AI模型」則非本法所定義之AI系統。然而，考量到通用目的AI模型發展成AI系統之類型與自我管制的不確定性[227]，《人工智慧法》在2023年的三方協商後，強化了對於此模型的管制與資訊透明要求。通用目的AI模型之供應商應擬定相關技術文件，內容應包括模型訓練、測試與結果評估之資訊，並於AI Office或國家權責機關要求時提供之。同時，考量到通用目的AI模型可能整合至其他供應商之AI系統，該模型之供應商應擬定相關資訊與文件，俾使AI系統之供應商清楚了解該模型之能力與極限，並得以善盡其依本法所負之義務[228]。惟若通用目的AI模型係以免費開源之方式發布，則毋須負上述義務[229]。

此外，為因應通用目的AI模型可能衍生的系統性風險[230]，本法特針對「具有系統性風險之通用目的人工智慧模型」（general-purpose AI models with systemic risk）加以管制。此類模型係指具有「高度影響能力」（high impact capabilities）[231]、且經歐盟執委會認

227 Artificial Intelligence Act recital 97.
228 Artificial Intelligence Act art. 53(1).
229 Artificial Intelligence Act art. 53(2).
230 Artificial Intelligence Act recital 111; *see* European Council Press Release IP/986/23, *supra* note 5.
231 若通用目的AI模型訓練累計之浮現運算數大於10的25次方，推定為具有高度

定或經獨立專家科學小組提出警示之通用目的AI模型。「具有系統性風險之通用目的人工智慧模型」即使以免費開源之方式發布，其供應商仍應負與通用目的AI系統供應商相同之義務[232]，並另負有進行辨識、評估與消弭模型所生系統性風險之義務；當產生重大意外時，亦應及時通報AI Office與國家權責機關[233]。

（七）AI監理沙盒與真實世界測試

監理沙盒（regulatory sandbox）為近年來許多政府面對新興科技與破壞式創新商業模式時常採取的管制手段，期能在鼓勵創新與完善法律監管間取得平衡[234]。監理沙盒於本法亦扮演重要的角色：《人工智慧法》除要求各會員國應至少設有一國家級的監理沙盒（會員國間亦得合作建立跨國監理沙盒），並至遲於2026年8月前開始運作[235]，亦從各個層面強化監理沙盒的功能與效果，包括要求歐盟執委會對各會員國之專家團體提供協助以設立監理沙盒、規定監理沙盒主管機關對可能開發高風險AI系統之供應商提供法遵指引，以及針對監理沙盒增設免責條款等[236]。此外，為鼓勵潛在供應商申請進入監理沙盒，本法參考2023年歐盟議會之建議，開啟另一條高風險AI系統取得上市許可進入市場的途徑：高風險AI系統離開沙盒時，若經主管機關評估符合本法對該系統的各項要求，則推定其已符合本法規定，受認證之第三方符合性評估機構在進行符合性評估時並應主動將其沙盒結案報告（exit report）納入評估依據[237]。為

影響能力。Artificial Intelligence Act art. 51(2).

232 Artificial Intelligence Act art. 53(2).

233 Artificial Intelligence Act art. 55.

234 Chang-Hsien Tsai, Ching-Fu Lin & Han-Wei Liu, *The Diffusion of the Sandbox Approach to Disruptive Innovation and Its Limitations*, 53 CORNELL INT'L L.J. 261, 266 (2020); *see* Dirk A. Zetzsche et al., *Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation*, 23 FORDHAM J. CORP. & FIN. L. 31, 39-43 (2017).

235 Artificial Intelligence Act art. 57(1).

236 Artificial Intelligence Act art. 57 & recital 139.

237 Artificial Intelligence Act art. 57(7).

避免歐盟會員國間監理沙盒管制的碎片化，本法亦要求歐盟執委會訂定監理沙盒施行法，統一規範監理沙盒的建立與運作[238]。

此外，為加速高風險AI系統開發與上市銷售之過程[239]，本法採納了歐盟理事會於2022年所提出之修正建議：除申請進入監理沙盒，供應商與潛在供應商在AI系統上市銷售或提供服務前，亦可選擇向市場監督機關申請「AI監理沙盒外之高風險AI系統真實世界條件測試」（testing of high-risk AI systems in real world conditions outside AI regulatory sandboxes）[240]。除了應取得市場監督機關的許可，供應商與潛在供應商應確保可能受影響之弱勢群體獲得適當保護、已取得所有受測者的知情同意[241]，以及確保該AI系統之預測、建議或決策可得推翻或使之無效。測試期間原則上為6個月，至多能延長至12個月。

（八）其他規範

如同GDPR之監管模式，《人工智慧法》亦定有鉅額罰則。從最初（2021年版草案與2022年歐盟理事會之建議）較為寬鬆之版本，到2023歐洲議會之建議相對嚴格之罰則[242]，《人工智慧法》最

238 Artificial Intelligence Act art. 58.

239 Artificial Intelligence Act recital 139.

240 Artificial Intelligence Act art. 60.

241 Artificial Intelligence Act art. 61.

242 2021年版草案關於違反本法之罰則：AI系統違反第5條或第10條，處最高3,000萬歐元或前一會計年度營業總額6%罰鍰；AI系統違反第5條與第10條以外之規範，處最高2,000萬歐元或前一會計年度營業總額4%罰鍰；提供受認證之第三方符合性評估機構或國家權責機關錯誤、誤導或不完整資訊，處最高1,000萬歐元或前一年會計年度營業總額2%罰鍰。Draft Artificial Intelligence Act 2021 ver. art. 71. 2023年版草案關於違反本法之罰則：AI系統違反第5條，處最高4,000萬歐元或前一會計年度營業總額7%罰鍰；AI系統違反第10條與第13條者，處最高2,000萬歐元或前一會計年度營業總額4%罰鍰；AI系統或基礎模型違反第5、10或13條以外之規範，處最高1,000萬歐元或前一會計年度營業總額2%罰鍰；提供受認證之第三方符合性評估機構或國家權責機關錯誤、誤導或不完整資訊，處最高500萬歐元或前一年會計年度營業總額1%罰鍰。Draft Artificial Intelligence Act 2023 ver. art. 71.

終條文於三者間折衷取得罰鍰金額之共識[243]。簡要而言，針對以下三種違規態樣，《人工智慧法》之罰鍰分別為：1.違反第5條禁止的AI使用情狀者，處最高3,500萬歐元或前一會計年度營業總額7%罰鍰；2.各行為者違反依本法所應負之義務者[244]，處最高1,500萬歐元或前一會計年度營業總額3%罰鍰；以及3.提供受認證之第三方符合性評估機構或國家權責機關錯誤、誤導或不完整資訊者，處最高750萬歐元或前一年會計年度營業總額1%罰鍰。此外，相較各機關於2021年至2023年間所提出之草案與修正建議，《人工智慧法》最終條文似更擔憂違反通用目的AI模型相關規定可能衍生之廣泛後果，而將通用目的AI模型之罰則獨立於AI系統之罰則，明定若供應商違反本法關於通用目的AI模型之規定，且歐盟執委會認其違反行為係出於故意或過失，將面臨最高1,500萬歐元或前一會計年度營業總額3%之罰鍰[245]。

為緩解AI系統可能產生的風險與影響，《人工智慧法》除規定AI系統相關行為者事前與事後的義務外，亦保障可能受AI系統影響者的救濟權利，可分為二者：向市場監督機關提出申訴（complaint）[246]，以及請求高風險AI系統部署者提供清楚且有意義的解釋[247]。前者之申訴機制，允許任何人在認為有違反本法之情事時，得向市場監督機關提出申訴。後者則係指部署者於利用高風險AI系統進行決策並因此產生法律效果或相似之重大影響時，受該決策影響者如認該決策對其健康、安全等基本權利產生負面影響，得向部署者請求提供清楚且有意義的解釋。

---

243 Artificial Intelligence Act art. 99.

244 此所列違反各行為者依本法所應負之義務之條文包括：第16、22、23、24、26、31、34、50條，以及第33條第1、3與4項。Artificial Intelligence Act art. 99(4).

245 Artificial Intelligence Act art. 101.

246 Artificial Intelligence Act art. 85.

247 Artificial Intelligence Act art. 86.

此外，為促進AI產業的發展，《人工智慧法》特別減輕中小企業[248]與新創企業在各種層面的義務，並試圖緩解渠等與大型企業協商時可能遭受的風險。就申請進入監理沙盒的費用部分，則要求各會員國應特別考量中小企業或新創企業之權益與需求（interests and needs），酌予減少與免除費用，且當中小企業或新創企業設立於歐盟境內時，於申請時得享有優先權（priority access）[249]。此外，同法亦要求各會員國在訂定罰則時，應考量到中小企業與新創企業的權益與其經濟能力[250]，顯示立法者並未因推動AI系統管制而忽略促進AI產業創新之必要。

參、歐盟《人工智慧法》之許諾與失落

面對AI技術與應用的持續發展，歐盟已意識到AI對個人資料與隱私造成之問題已無法僅仰賴GDPR加以解決，尤其相關問題並不限於此，尚包含其應用所帶來各種基本權利風險、對司法體系、民主制度的影響，以及其他科技發展的不確定性。此外，「非個人資料」（non-personal data）亦為AI發展與應用之重要基礎，雖然其不涉及個人資料之蒐集、處理或利用，而未衍生個人資料保護法律適用問題，但其中仍牽涉技術倫理（如自駕車道路訓練與判斷）、社會資源分配、企業公平競爭，乃至於國際經貿對於資料傳輸或在地

248 本法「中小企業」之定義係援引歐盟2003年對中小企業規模定義之建議書（Commission Recommendation of 6 May 2003, Concerning the Definition of Micro, Small and Medium-Sized Enterprises, annex, art. 2, 2003 O.J. (L 124) 36, 39），中小企業係指公司僱用之員工人數少於250人，且年營業額不超過5,000萬歐元或資產負債淨額不超過4,300萬歐元。Artificial Intelligence Act recital 143.

249 Artificial Intelligence Act art. 62 & recital 143.

250 Artificial Intelligence Act art. 99; 另為緩解大型企業與中小型企業協商能力的不對等問題，並考量到契約自由原則，2023年版草案建議增訂關於不平等單方合約的定義與法律效果，以協助中小企業與新創企業在AI產業的發展，然最終版本並未採納。Draft Artificial Intelligence Act 2023 ver. recital 60a & art. 28a.

化（data localization）[251]與國家安全例外措施等管制問題。歐盟於2022年通過的資料治理法（Data Governance Act）及2023年通過的資料法（Data Act）均已對非個人資料設下相關規範[252]，試圖處理部分管制問題，《人工智慧法》則藉由以風險為基礎的管制框架將管制量能集中於高風險AI系統，並賦予高風險AI系統生命週期的各行為者相應的義務，以強化系統的風險管理、品質管理、紀錄完整度、系統健全性、人類監督與透明度等。此外，為輔助產業的法遵與創新，《人工智慧法》亦訂有監理沙盒與對中小企業的優待措施。

值得注意者為，除了《人工智慧法》各條規定外，其前言的作用亦不容忽視。或許因立法時間倉促且混合多方意見，又面對技術迅速發展而須更新調整用語，導致最終條文不乏法條結構鬆散、文字不精確甚至有潛在矛盾之問題，未來執行時恐將存在許多模糊空間與不確定性。因此，屆時除有待執委會發布補充性指引進一步闡明重要條文內涵外，法院對於條文的解釋與適用，勢必亦將援引立法宗旨與規範目的來支持其論理，使得前言將扮演重要功能。不過，由於立法倉促且多方折衝，目前《人工智慧法》的前言內容極為繁雜且指向多元，亦未必有利於法律明確性與穩定性。此外，歐盟以全面性專法針對AI具體使用情狀進行管制之措施，雖可一定程度地緩減當今面臨的AI挑戰，但面對AI技術的快速發展與不確定性，似尚需一般性之準則、指引及倫理原則作為補充性規範，以使AI管制透過適當運用軟法而有能力回應對於新興基本權利保障之需

251 *See, e.g.*, Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 Nov. 2018 on a Framework for the Free Flow of Non-personal Data in the European Union, 2018 O.J. (L303) 59.

252 Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance), 2022 O.J. (L 152) 1; and Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA relevance), 2023 O.J. (L 2023/2854).

求[253]。在《人工智慧法》立法過程中，雖然AI HLEG發布的多項文件與2023年版草案皆強調一般性倫理原則的功能與重要性[254]，但在最終條文中已被散落至各段前言內，並無一般性倫理原則的柔性支撐。若《人工智慧法》能輔以基本人權為核心（human rights-based）之取徑，並透過執委會與各部門彈性調整管制策略、發布標準、指引與準則，或許能使AI的發展與管制更具韌性[255]。

無論如何，《人工智慧法》於2024年順利通過後，將於生效日兩年後（即2026年8月2日）全面施行[256]。基於其廣泛的適用範圍與拘束力，歐盟的管制取徑與價值定位將影響全球人工智慧管制的發展方向。其中重要之制度設計包含以風險為基礎的管制框架、人類監督機制、透明度規範及監理沙盒，均屬近年來學術與實務上探討AI所帶來的核心治理問題中最受重視者，應予深入分析。首先，以風險分級為基礎的管制框架，似為當今相對適當且有一定共識的管制取徑。此框架係設定風險區分標準與方法，並據以衍生出對應之不同管制措施，包含禁用「存在不可接受的風險」之AI使用情狀，或是原則上由企業自律管制的非高風險AI系統等。然實務上如何執行、風險如何認定、以及對於所謂非高風險之AI系統之管制是否足夠等問題，均仍有疑義，本文將於第一節討論之。其次，高風險AI系統應遵循之各項法定要求中，人類監督機制與透明度尤為重要，且攸關人類如何妥善控制AI系統並掌握可能影響：前者已於GDPR

253 *See* Hannah Ruschemeier & Jascha Bareis, *Searching for Harmonised Rules: Understanding the Paradigms, Provisions and Pressing Issues in the Final EU AI Act* 25-26 (June 28, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4876206.

254 Draft Artificial Intelligence Act 2023 ver. art. 4a.

255 *See* Stephen K. Kwan & Jim Spohrer, *Reducing Industry Complexity with International Standards: Current Efforts for Services, E-commerce, Artificial Intelligence*, *in* ADVANCES IN THE HUMAN SIDE OF SERVICE ENGINEERING 67 (Christine Leitner et al. eds., 2021).

256 《人工智慧法》於2026年8月2日全面施行，然考量管制急迫性與相關行為者回應《人工智慧法》所定義務所需時間，部分條文提前與延後施行，詳見本文第貳部分說明。

第22條有所規定，要求用於AI系統輔助或取代人類決策之應用場域設有相關機制，《人工智慧法》則進一步將適用範圍具體延伸到特定政府公權力執行乃至於數位平臺監管；而透明度之要求，則為各國近年法規政策、基本原則、最佳實踐，或是國際組織採納的AI準則與宣言等，均視為重要治理原則者。本文將分別於第二、三節分析之。最後，《人工智慧法》為在管制與創新之間取得平衡，要求各會員國至少設有一國家級AI監理沙盒。監理沙盒為平行於其他管制措施之制度設計，將直接影響歐盟會員國在地管制成效，以及歐盟AI治理的一致性與有效性，而會員國間如何調和（harmonize）AI監理沙盒的規範與施行，實為不可忽略之問題，將於第四節探討之。

一、以風險為基礎的管制及其侷限

《人工智慧法》以風險作為管制AI系統之分級標準，如前述，分為「不可接受的風險」、「高風險」與「低或微小的風險」等三種級別。然而，條文並未明定「不可接受的風險」與「高風險」之定義，而係透過具體情狀的方式來說明：第5條例示存在不可接受的風險而禁止的AI使用情狀，第6條與附件三則列舉存在高風險而可能需符合各項要求之AI系統。值得注意的是，被禁止的AI使用情狀並不限於《人工智慧法》第5條所列之情狀，還包括其他歐盟法規所禁止者（如GDPR或反歧視法等）[257]；而高風險的AI系統則限於《人工智慧法》附件三所列舉之情狀，並得由歐盟執委會修訂附件三之涵蓋範圍[258]。

考量到AI系統應用的多樣性與其發展速度，對所有AI系統進行管制之成本高昂且不具實務可行性[259]，立法者必須基於社會的風險

257 Artificial Intelligence Act art. 5(8).
258 Artificial Intelligence Act art. 7(1) & (2).
259 Anja Folberth et al., *Tackling Problems, Harvesting Benefits - A Systematic Review*

認知與基本權利保護必要性，據以設定對應的判斷標準、劃下適當分界。是故，《人工智慧法》將主要的管制客體限縮於高風險AI系統，並透過符合性評估與基本權利影響評估，強化供應商與部署者對風險的管理，應係當今相對適宜且可行的作法[260]。尤其基於特定社會與政治經濟脈絡，或出於對基本人權保障等考量，透過立法明確劃定出不可接受的風險，進而禁止特定AI系統與應用情狀，係面臨AI技術迅速發展及高度不確定之情況下，值得各國參考之作法，此亦與歐盟於不同新興科技管制領域中所採納的預防原則（precautionary principle，或稱預警原則）取徑一致。然即便如此，在以風險為基礎之制度設計與法規執行上，《人工智慧法》仍存在待釐清與改善之處，包括高風險AI系統認定的歧異、供應商與部署者以及其他行為者間義務模糊，以及「使用情狀」為管制標的可能欠缺彈性而無法應對技術快速發展帶來的挑戰等，本文以下評析之。

（一）高風險AI系統的認定

相較於2021年版草案高風險AI系統之定義係指「落入附件三所列情狀」[261]，2023年歐洲議會版本似考量到僅透過附件三所列客觀情狀無法確實反映系統是否具備「高風險」，因此在高風險AI系統的判斷標準建議新增具相當裁量空間的判斷標準，將高風險AI系統的定義修改為「落入附件三所列情狀且對健康、安全或基本權利構成顯著風險者」[262]，供應商並得向主管機關提交附理由的通報（submit

---

of the Regulatory Debate Around AI 17 (KIT Scientific Working Papers No. 197, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4212447.

260 Anja Folberth等人的研究指出，目前多數AI的管制措施亦採以風險為基礎的管制架構。*Id.* at 3.

261 Draft Artificial Intelligence Act 2021 ver. art. 6(2) (“In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk.”).

262 Draft Artificial Intelligence Act 2023 ver. art. 6(2) (“…AI systems falling under one or more of the critical areas and use cases referred to in Annex III shall be considered high-risk if they pose a significant risk of harm to the health, safety or fundamental

a reasoned notification）說明其AI系統雖落入附件三所列情狀，但並未對前述基本權利構成顯著風險，故非高風險AI系統[263]。《人工智慧法》最終條文承2023年歐洲議會版本之建議，使高風險AI系統分類設有非屬高風險之例外，並使認定之標準與程序更加具體，即若附件三之AI系統未對前述基本權利構成顯著風險、或未對決策結果產生實質影響，並符合以下任一情形，則非屬高風險，包括：用於執行狹義的程序性工作、用於改進人類已完成之成果、用於偵測決策模式以及決策是否偏離先前的決策模式，或用於進行附件三所列AI系統評估之前置作業[264]。供應商若認其AI系統非屬高風險AI系統，應在上市銷售前或提供服務前記錄其對於AI系統之評估並登記於歐盟資料庫，以及應國家權責機關之要求提供該評估紀錄[265]。

《人工智慧法》雖考量到供應商與主管機關對於風險的認定可能有異，而使供應商得於符合上述情形時主張其系統不構成顯著風險，但制度上仍由各國主管機關作為認定AI系統「是否構成顯著風險」而屬高風險系統的把關者。此制度設計，雖然將風險認定權限交予主管機關，減少私部門在程序中的積極影響，相對能避免私部門對主管機關造成管制俘虜的問題，但是在高度技術專業的議題上恐怕仍有挑戰。此外，即使《人工智慧法》指出各國權責機關應相互合作[266]，然會員國間對於「是否構成顯著風險」的判斷標準，恐存在不少差異；會員國對於不同風險的接受程度，取決於其社會、經濟與國家發展等因素，無法一概而論。因此，歐盟如何透過進一步提供明確判斷標準與方法論，調和或統一會員國對於顯著風險的

---

rights of natural persons….").

263 Draft Artificial Intelligence Act 2023 ver. art. 6(2a).

264 Artificial Intelligence Act art. 6(3) ("By derogation from paragraph 2, an AI system referred to in Annex III shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.").

265 Artificial Intelligence Act art. 6(4).

266 Artificial Intelligence Act art. 74(10).

認定、以避免產生高風險AI系統管制漏洞，將是未來需要面對的問題之一。就此，歐盟執委會於2024年1月24日採納決定（Commission Decision of Establishing the European Artificial Intelligence Office）並於2月21日正式設立歐盟AI Office[267]，授權該局對於《人工智慧法》相關事項發布指引（guidance）或採取其他行動，如監測《人工智慧法》的施行與遵循狀況、開發用於評估通用目的AI模型之工具、方法與基準、針對監理沙盒之運作提供技術建議與協調、維護資料庫，並促進會員國主管機關間的資訊交流、歐盟執委會內的跨機關合作及國際合作等。其中關於技術支援與利害關係者之交流合作，有機會將私部門的技術專業正式納入相關決策程序，作為促成基本權利保障與數位科技研發間平衡的施力點。

綜言之，《人工智慧法》所設想最理想的適用結果，即供應商確實認定其所開發之系統是否屬高風險AI系統，並盡其所應負之對應義務。實際上，《人工智慧法》雖要求歐盟執委會應於2026年2月2日前提供關於判斷AI系統是否屬高風險之參考指引[268]，然並未建構完善機制協助主管機關積極審視供應商之自我認定結果；若供應商認定其系統並非高風險AI系統，而不踐行相關義務（如符合性評估與事前登記），主管機關往往僅會在發生重大事件時，才有機會審視之。此外，當主管機關與供應商見解不同時，現行管制框架係以AI Office或各歐盟會員國主管機關之判斷為尊，除可能忽略了不同利害關係人審議參與之重要性外，亦可能排除私部門在AI風險管理上的潛在專業貢獻（雖然會有利益衝突考量）；此外，各歐盟會員國間對於高風險甚至是禁止使用情狀的認定若有無法忽略的差異，亦可能衍生諸如管制競爭（regulatory competition）[269]、管制

267 European Commission, *Commission Decision Establishing the European AI Office* (Jan. 24, 2024), https://digital-strategy.ec.europa.eu/en/library/commission-decision-establishing-european-ai-office.

268 Artificial Intelligence Act art. 6(5).

269 *See generally* Nathalie A. Smuha, *From a "Race to AI" to a "Race to AI Regulation":*

套利（arbitrage）[270]或是類似挑選法庭（forum shopping）等問題。

（二）供應商、部署者與其他行為者的義務模糊

除前述主管機關與供應商之間的風險認定落差外，高風險AI系統相關行為者間義務範圍與關係區分不明，亦不利於此一風險管理框架獲得有效實踐。舉例而言，《人工智慧法》第25條規定，供應商以外之其他行為者若符合特定情狀（例如對已上市銷售或提供服務的非高風險AI系統進行重大修改使其成為高風險AI系統[271]），擬制為供應商並應負供應商義務，而原供應商則不再屬本法所定供應商。此規範設計之優點在於，各AI行為者對於AI風險之注意程度，理論上將因此提高；然而，此亦可能使高風險AI系統的注意義務與法律關係更趨複雜不穩定。

為了使系統不被認定為高風險AI系統（以減少法遵成本），供應商有強烈動機自行評估其是否「構成顯著風險」、並登記於歐盟資料庫，主張其即使落入附件三所列情況仍非屬高風險AI系統[272]。然而，在《人工智慧法》以供應商為主要管制對象，且其他行為者並無動機主動揭露並成為「擬制供應商」之情況下，僅有在損害結果產生後、探究責任歸屬時才可能將非供應商之行為者納入，有可能衍生許多問題，如：重大修改的認定標準，以及若AI系統產生損害或違反本法規定時，原供應商與擬制供應商卻尚未依本法移交相關資訊與完成相關義務，其責任應如何歸屬等[273]。從條文之形式規

Regulatory Competition for Artificial Intelligence, 13 LAW INNOVATION & TECH. 57 (2021).

270 *See generally* Tambiama Madiega & Anne Louise Van De Pol, *Artificial Intelligence Act and Regulatory Sandboxes*, EUR. PARL. RES. SERV., PE 733.544 (June 2022), https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.

271 Artificial Intelligence Act art. 25(1)(c).

272 Artificial Intelligence Act art. 6(4).

273 關於供應商與其他行為者間義務的劃定，另值得注意者為歐盟執委會於2023

定看來，這些問題都無法獲得明確答案。

（三）其他非高風險AI系統的管制不足

在《人工智慧法》風險分級的架構下，除禁止的AI使用情狀與高風險AI系統外，尚包括其他非高風險AI系統[274]，然而本法對此類型AI系統幾乎無任何強制規定與義務，而僅鼓勵業者採納建議性行為準則並自願遵循本法規範。此雖有助於中小企業與新創產業的發展，但若欠缺對此類（暫時屬非高風險的）AI系統之掌握，倘於相關系統蒐集更廣泛使用者資料並進一步訓練強化後而有不同應用情狀時，如何在適當時間內變更風險認定？行為者是否有誘因通報？持續性的變動又應如何監測？種種問題，本法並無著墨。此設計或許能大幅減少對其他非高風險AI系統風險之管理成本，但未來恐肇生許多管制盲點。例如，生成式AI於2023年帶來的改變與風險，即無從預期，且欠缺相應的治理措施，究屬禁止的AI使用情狀、高風險AI系統或其他非高風險AI系統，並不易一刀明確劃分。因此，從管制合理性與法律遵循成本等角度來看，雖不宜將針對高風險AI系統的所有管制要求全部套用至其他非高風險AI系統，但對於目前尚未被重視的AI應用與風險，或是AI系統在真實世界應用的變遷與風險不確定性，應可考慮設有低強度但更全面的管理制度（例如提供獎勵或優惠措施以鼓勵業者自願登記之制度），或許將優於在辨識出新風險後而亡羊補牢，更能達妥善治理之效。

年發布《公部門採購人工智慧合約條款建議範本》（Proposal for Standard Contractual Clauses for the Procurement of Artificial Intelligence (AI) by Public Organisations），意圖使公部門與AI系統提供者（supplier）之間，對《人工智慧法》中針對AI系統的行為者的各項要求與義務分配更加清楚。European Commission, *Proposal for Standard Contractual Clauses for the Procurement of Artificial Intelligence (AI) by Public Organisations, Version September 2023 (draft) – High Risk version*, https://public-buyers-community.ec.europa.eu/system/files/2023-10/AI_Procurement_Clauses_template_High_Risk%20EN.pdf.

274 Artificial Intelligence Act art. 95(1).

此外，《人工智慧法》刻意將AI系統於軍事、國防及國家安全（military, defence or national security）用途排除在本法適用範圍外[275]，並於前言指出依據《歐盟條約》（Treaty on European Union）第4條第2項，歐盟應尊重會員國對於軍事、國防及國家安全的自治與責任[276]，故於《人工智慧法》中排除AI於軍事、國防及國家安全上之應用係合理的法規設計[277]。惟本文認為，無法在歐盟立法層次上涵括軍事、國防及國家安全的AI應用規範，除未把握作為全球規範領頭羊而禁止發展及使用AI自動化武器的重要規範時刻外[278]，亦忽略各國逐漸將AI工具用於國家安全監測與自動化決策所產生的法治及人權問題，而未能藉此契機至少對其高風險或不可接受的風險情狀予以敘明，如此管制立場，實屬可惜。

二、確立人類監督機制的必要性

為控管AI所造成的風險，《人工智慧法》除了直接禁止使用具不可接受的風險之AI使用情狀外，對於存在高風險之AI系統則是定有多項義務，已如前述。其中，針對人類監督機制之要求，實為歐盟自1995年《個人資料保護指令》（Data Protection Directive, DPD）[279]以來，對於控管演算法對人類權利之負面影響、以及確認人類在自動化系統中所應扮演角色與責任等方面，在制度設計上堪稱最為明確、完善之版本[280]。

---

275 Artificial Intelligence Act art. 2(3).

276 Treaty on European Union, art. 4(2).

277 Artificial Intelligence Act recital 24.

278 *See* Ruschemeier & Bareis, *supra* note 253, at 13.

279 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 15(1), 1995 O.J. (L 281) 31, 43 [hereinafter DPD].

280 *See* Guillermo Lazcoz & Paul de Hert, *Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities*, 50(105833) COMPUTER L. & SEC. REV. 1, 6-7 (2023).

概觀歐盟過往作為，除在訂定DPD時嘗試藉由該指令第15條限制ADM與大型資料庫對於資料主體個人隱私之影響[281]，也在GDPR納入此一規定，於第22條規定「不受完全基於自動化處理（包括剖繪）所做成而對其產生法律效果（legal effects）或受類似重大影響（significantly affects）之決策所拘束」之權利[282]。此般機制之設計理念，是為了避免資料主體的個人資料遭到濫用，使得人類免於淪為機器決策之奴隸[283]，尤其是在系統可能對人類日常生活造成不當侵害、產生歧視性結果時，更應該要有人類監督以控管風險[284]。不過，DPD與GDPR的規範重心實為個資而非ADM與AI，使得人類監督機制的規範與操作出現了難以理解、易於規避，實務上也難以適用等問題[285]，作為AI治理措施仍有失完善。相較於DPD與GDPR之不足，《人工智慧法》延續並強化人類監督的具體規範，以避免人類監督流於形式而欠缺有意義的介入，並提升對基本權利之保障。以下將簡要介紹就GDPR在此方面之規範與不足，以及《人工智慧法》為了強化此機制所做之規定。

---

281 DPD art. 15(1) ("Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."); *see* Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1460 (2001).

282 GDPR art. 22(1) ("The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.").

283 Lazcoz & de Hert, *supra* note 280, at 4.

284 Citron & Pasquale, *supra* note 31, at 7-8; *but see* Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, 12 REG. & GOVERNANCE 505, 516 (2018).

285 Isak Mendoza & Lee A. Bygrave, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW: REGULATION AND ENFORCEMENT 77, 78 (Tatiana-Eleni Synodinou et al. eds., 2017).

（一）GDPR有關人類干預（human intervention）之規定

GDPR在第22條特別針對包括剖繪在內的自動化個人決策（automated individual decision-making, including profiling）的相關問題進行處理，是因為此類自動化決策的使用可能對個人權利與自由產生顯著風險，包括不透明、強化刻板印象、加深社會隔離等等，因此需要適當的防護措施（safeguards）[286]。若某項決定將對資料主體產生法律效果或類似的重大影響時，除為履行契約所必須、或已獲得資料主體之同意，否則在決策過程中應有人類干預之設定[287]。

根據歐盟「第29條工作小組」（Article 29 Working Party）所發布之《關於第2016/679號規則中的自動化個人決策和剖繪之指引》（Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679），GDPR第22條所要求的人類干預，必須要確保人類對於決策的監督是有意義的、而非只是橡皮圖章（token gesture）；人類干預必須由具有權限與能力改變決策之人所執行，且在進行分析時，該人必須考量所有相關資料[288]。由此可見，歐盟在GDPR之管制框架下，已著手嘗試處理（部分）演算法之影響，避免個人權利因為自動化決策系統而遭到侵害。

不過，GDPR第22條雖建立了資料主體「不受完全基於自動化處理所做成而對其產生法律效果或類似之重大影響之決策所拘束」之權利，然在部分條件下，自動化決策仍具拘束力，例如已取得資料主體的同意，或該自動化決策程序為締結或履行資料主體與控管者間之契約所必要者等[289]。在解釋上，GDPR第22條多被認為是一

---

286 Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 (wp251rev.01)* 5 (Feb. 6, 2018), https://ec.europa.eu/newsroom/article29/items/612053/en.

287 GDPR art. 22.

288 Article 29 Data Protection Working Party, *supra* note 286, at 21.

289 Michael W. Monterossi, *Algorithmic Decisions and Transparency: Designing Remedies in View of the Principle of Accountability*, 5 ITALIAN L.J. 711, 722 (2019).

種自由（liberty）或一般性的禁止規定，而非資料主體的權利（right）[290]。相對於其他條文賦予資料主體可主張的被遺忘權、修正權等權利，第22條規範設計的方式對於資料主體而言很難在法律上有所主張，因此有學者指出，該條本質上更像一項對於業者的禁令而非資料主體的權利[291]。換句話說，資料主體在行使此權利時無需（更精確地說是「無法」）採取積極行為，僅有資料控制者被要求避免對於此權利之干預[292]。是以，表面上資料主體雖享有此一權利，實務上卻似乎難以主張。

此外，GDPR第22條管制的對象為對資料主體產生法律效果或類似「重大影響」的演算法決策，使得線上廣告與個人化內容可能不會直接落入此範圍[293]。而由於缺乏對於「重大效果」之定義，個人也可能無法免於持續受到刻意設計且充滿偏見與極端的內容之影響[294]。換言之，只要演算法決策不產生法律效力、或有人類干預決策，就不需要負揭露義務[295]，這會使得有許多可能侵害隱私的演算法不受本條限制[296]。最後，GDPR第22條僅針對完全自動化的決策程序，並未對於自動化處理背後的演算法類型加以區分；以最簡單的決策樹與高階的深度學習系統為例，即令其背後技術複雜程度相差甚遠且「黑盒子」問題完全不同，仍均適用相同法律規範，並非妥適設計。

---

290 Article 29 Data Protection Working Party, *supra* note 286, at 19.

291 Mendoza & Bygrave, *supra* note 285, at 85.

292 Giovanni De Gregorio, *The Rise of Digital Constitutionalism in the European Union*, 19 INT'L J. CONST. L. 41, 64 (2021).

293 Lu, *supra* note 109, at 2114. 翁逸泓，社群媒體不實訊息之治理——以個資保護模式為選項，中研院法學期刊，30期，頁193（2022年）。

294 Lu, *supra* note 109, at 2114.

295 *See* Bryan Casey et al., *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 143 (2019).

296 Maja Brkan & Grégory Bonnet, *Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas*, 11 EUR. J. RISK REG. 18, 21-22 (2020).

（二）《人工智慧法》對人類監督相關規範之深化

相對於GDPR第22條對於演算法規範不足、資料主體也難以行使該條所賦予之權利，《人工智慧法》則是在第14條明確要求高風險AI系統必須有讓人類得以監督之設計，對於人類監督的內容也在該條第3項有更為細緻的要求，實為對《關於第2016/679號規則中的自動化個人決策和剖繪之指引》更進一步的回應與實踐。在監督者資格方面，第14條除規定其須完全瞭解高風險AI系統的能力與限制、具備發現異常之能力，還須認知到人類會過分依賴AI決策之問題、有能力解讀AI系統產出結果、有能力決定何時不應使用高風險AI系統或何時應推翻其決定，以及有能力介入高風險AI系統之運作，甚至有權按下「停止」按鈕（a “stop” button）[297]，藉此確保有意義的人類監督、有權限在必要時採取相應作為，而非只是形式上的觀察AI系統運作。

《人工智慧法》最終條文考量人類監督之目的，以及對個人健康、安全或基本權利產生危害之風險，進一步地禁止「完全基於即時遠端生物辨識系統的決策對個人產生法律上負面影響」[298]、禁止「執法機關完全基於非即時錄影畫面而做成之決策對個人產生法律上負面影響」[299]，以及禁止「上市銷售、提供服務或使用完全基於剖繪或對於個人特徵之評估，以進行犯罪風險評估或預測之AI系統」[300]。《人工智慧法》此些規範中「完全基於」之用語與GDPR第22條第1項之規定高度雷同，似也反映了GDPR對此規範不足，因而有必要在以AI為管制中心的法規中強化人類監督之規定[301]，並透過第13條有關透明度義務之規定監督法遵情形[302]，也凸顯了透明

297 Artificial Intelligence Act art. 14(4)(a)-(e).
298 Artificial Intelligence Act art. 5(3), second paragraph.
299 Artificial Intelligence Act art. 26(10), third subparagraph.
300 Artificial Intelligence Act art. 5(1), first subparagraph, point (d).
301 *See* Lazcoz & de Hert, *supra* note 280, at 6.
302 *See* Lazcoz & de Hert, *supra* note 280, at 7.

度規範在本法的重要性，本文將於下一小節分析之[303]。

然而，《人工智慧法》雖透過強化人類監督的具體規範，以提高對基本權利之保障，但卻忽略了在AI發展與應用之生命週期中，不同行為者之能力、權責、可獲得之必要資訊、對最終決策之影響力，以及技術上得否真正進行有意義人類監督及介入之程度，均不盡相同。《人工智慧法》僅以一般性規範設計處理，可能擴大在AI系統應用生命週期中各種具備足夠智識而能裁決、介入系統運作之人的自由裁量權力，如軟體開發者和演算法分析／轉譯者等，但卻未清楚規範應如何衡量渠等是否有適當能力或是否具可課責性。基此，決策過程中人類監督機制的性質與具體內涵，或許應更清楚地判別、劃分，並與傳統（GDPR所規範的）決策者的代理權相一致，從而針對相關決策者在AI系統生命週期中相應位置之個別權責與角色，設計更具技術可行性與經濟及社會理性之規範內涵。

三、力有未逮的透明度規範

早在AI黑盒子的問題受正視前，透明度在管制研究領域已為重要的概念與政策工具，透過要求權力行使應具備足夠透明度，可強化公眾對於該權力行使體制的監督，進而提升對受影響者權利的保障[304]。近年隨著AI的發展、資料利用樣態與演算法設計愈趨複雜，透明度在AI倫理與管制上也開始被尊為重要原則之一[305]。例如，歐

303 值得一提的是，2023年版草案於第29條第6項規定，高風險AI系統使用者應使用第13條要求提供之資訊，執行GDPR第35條所規定之資料保護影響評估（data protection impact assessment），凸顯了GDPR與2023年版草案之間的互補關係。

304 *See* Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885 (2006); *see also* Andrea Bianchi, *On Power and Illusion: The Concept of Transparency in International Law*, *in* TRANSPARENCY IN INTERNATIONAL LAW 1, 1 (Andrea Bianchi & Anne Peters eds., 2013); *see also* Anoeska Buijze, *The Six Faces of Transparency*, 9(3) UTRECHT L. REV. 3, 4 (2013).

305 Anna Jobin, Marcello Ienca & Effy Vayena, *The Global Landscape of AI Ethics Guidelines*, 1 NATURE MACHINE INTELLIGENCE 389, 391 (2019); *see also* Mateusz

盟AI HLEG於2019年發布的《可信任AI之倫理準則》[306]即將透明度列為實現可信任AI的七大準則之一。透明度之概念不僅反映於《人工智慧法》中的各項行為者義務，本法前言更進一步說明透明度之定義：「AI系統開發與使用具適當的可追溯性與可解釋性，以及當AI系統與人類溝通互動時應使人類知悉，並告知受AI系統影響者該AI系統的能力與限制[307]」。整體而言，《人工智慧法》關於透明度的規定，可依據作用不同而進一步區分為「AI系統本身」的透明度與「與AI系統互動」的透明度，前者俾使AI系統的運作與決策可被解釋，後者則係為確保AI系統之存在與互動能被與之互動者知悉。本文肯定《人工智慧法》明文定義透明度並設有相關法律要求之規範取徑，惟作為全球第一部管制AI的全面性法律，可發現歐盟在透明度要求之落實似乎力有未逮，僅針對人類與AI系統互動時之透明度予以規範，而未針對因營業秘密保護所生之「法律黑盒子」與深度學習演算法欠缺可解釋性之「技術黑盒子」制定相應且足夠之透明度規範，顯示仍有相當改善空間，以下進一步評析之。

（一）AI系統本身的透明度

「AI系統本身」的透明度係指AI系統的可近用性與可解釋性，亦即法律黑盒子與技術黑盒子[308]：前者係指AI系統演算法模型與原

Grochowski et al., *Algorithmic Transparency and Explainability for EU Consumer Protection: Unwrapping the Regulatory Premises*, 8 CRITICAL ANALYSIS L. 43, 58 (2021).

306 HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, *supra* note 126.

307 Artificial Intelligence Act recital 27 (“Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability, while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights.”). 關於透明度之定義，2023年版草案係建議將其增訂於第4a條「人工智慧一般性原則」（general principles applicable to all AI systems），然《人工智慧法》最終條文並未採納，並將其條文稍做修改並增修至前言第27點。Draft Artificial Intelligence Act 2023 ver. art. 4a(1)(d).

308 Liu et al., *supra* note 29, at 135.

始碼因受營業秘密所保護而得以不公開，使相關利害關係人無法近用並試圖知悉其基本運作邏輯[309]；後者則源於AI系統的技術本質，因為機器學習與深度學習演算法的運作特性，縱使是專業領域之人亦無法針對個別決策過程提供人類得以理解之解釋[310]。為了「打開」這些不同類型的黑盒子，《人工智慧法》要求供應商應使AI系統具備透明度與可解釋性（第13條），並透過符合性評估（第43條）確保系統上市銷售前或提供服務前符合第13條透明度與可解釋性要求，以及允許受影響者透過請求部署者提供有意義解釋之權利（第86條），使系統在事前與事後皆具備可解釋性。然而，前述條文雖可協助打開法律黑盒子與技術黑盒子，然此是否使AI系統具備真正透明度，則仍需從透明度之不同層面探討之。

首先，依據AI系統行為者的不同，法律黑盒子應可再細分為「供應商的法律黑盒子」與「部署者的法律黑盒子」等二層次進行討論。首先，「供應商的法律黑盒子」指供應商基於營業秘密保護而不公開其AI系統的設計與開發資訊。對此，《人工智慧法》第13條規定高風險AI系統之供應商在上市銷售前或提供服務前，應進行符合性評估，評估基準則為供應商繳交之技術文件與品質管理系統[311]，其中包含演算法、訓練資料、測試資料與AI系統運作邏輯等。於此，供應商雖毋須對公眾或受影響關係人公開其系統設計與開發資訊，但須相當程度揭露其系統資訊予第三方符合性評估審查機構，限縮營業秘密之保護程度。另外，通用目的AI模型雖非屬AI「系統」而無第13條進行符合性評估之義務，然第53條要求通用目的AI模型之

309 參見林勤富，智慧法院之發展與界限（下）——演算法、科技治理與司法韌性，月旦法學雜誌，324期，頁111（2022年）；亦參見林勤富，人工智慧自動決策系統的透明度難題，收於：李建良編，人工智慧的法制基礎：破壞式創新的建構式法制，頁162-164（2024年）（以下簡稱：林勤富，2024）。

310 *See generally* PASQUALE, *supra* note 29; *see also* Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 981 (2018).

311 Artificial Intelligence Act annex IV & VII.

供應商應備齊且即時更新其模型之技術文件，並在AI Office或國家權責機關要求時提供之[312]。基此，通用目的AI模型雖未如高風險AI系統受到嚴謹的審查，但於必要時仍具有一定程度的透明度。而另一層次「部署者的法律黑盒子」則指部署者亦有保護其營業秘密而得不公開AI系統具體使用之相關資訊。對此，《人工智慧法》第86條賦予受AI系統影響者向AI系統部署者，針對其所作決策或行為請求提供解釋之權利。考量到部署者並非供應商、並未擁有AI系統的所有設計與開發資訊，《人工智慧法》要求供應商於提交符合性評估時，應備齊使用說明予部署者，以促其了解系統運作與數據處理方式並有能力向受影響者提供有意義的解釋[313]。綜上，《人工智慧法》分別透過事前的專家審查（第三方符合性評估）、備齊技術文件以及要求供應商提供部署者必要資訊，俾對受影響者提供有意義的解釋，使得原先受營業秘密保護而毋須公開之AI系統相關內容得以揭露，也就此打開了法律黑盒子，符合學者Frank Pasquale所主張之「有條件的透明度」（qualified transparency）模式[314]。

至於技術黑盒子，則源於AI系統的演算法技術本質與複雜設計，其運作過程是人類無法充分理解的，因此即便完全揭露AI系統之技術資訊，仍存在相對於人類理解決策過程之「可說明性」或「可解釋性」問題。究其原因，實因於人類知性理解力中，「解釋（explain）物理現象與理解（understand）行動（或決策）之意義（the meaning of an action）是兩個不同層次的知性活動」[315]。對

312 Artificial Intelligence Act art. 53(1).

313 Artificial Intelligence Act art. 13 & annex IV 1(g). 有學者亦稱專業知識之於外行人的理解障礙亦為黑盒子的成因之一，而《人工智慧法》第13條要求供應商應提供使用說明予部署者，以利部署者提供受影響者有意義的解釋，即應有助於緩解此專業知識所生之黑盒子問題。*See* Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3(1) BIG DATA & SOC'Y 1, 9-10 (2016).

314 *See* PASQUALE, *supra* note 29, at 142 & 161-63.

315 陳弘儒，論法律人工智慧系統與法治理念的緊張關係——初探技術開發的道

此，《人工智慧法》要求供應商在符合性評估中提供的資料並非單純的資訊揭露，而係包含演算法、訓練數據與系統運作邏輯等資料，可說《人工智慧法》在處理AI黑盒子的管制挑戰時，雖形式上觸及技術黑盒子問題，然就真正實質「可解釋性」之問題則力有未逮[316]。

事實上，若今要求AI系統須「完全透明」（亦即其原始碼等資訊不受營業秘密與其他法規保護、所有人皆得近用並了解系統如何運作，且在系統作出個別決策時均可充分解釋等），由於受限於機器學習與深度學習之技術特性，大多數AI系統將必然犧牲高效能與高準確性之表現，因此無論從產業創新抑或對於使用者而言，「完全透明」之要求並不利AI之發展。是以，當立法者試圖解決AI系統不透明問題時，除應探討造成問題之黑盒子類型外，亦應考量透明度之目的以及管制措施是否足以達到該目的，相關面向包括：是否／何時要求透明度（何種應用場域、系統上市銷售與提供服務前或後）、對誰實現透明度（一般大眾、專家或受影響之個人）以及透明的程度（簡單揭露或充分解釋）等[317]。

本文認為，《人工智慧法》第43條符合性評估之機制，即為意識到透明度不同面向之制度設計。符合性評估為「事前、對專家、系統運作邏輯資料的揭露」，雖僅對專家（第三方符合性評估機構）、而非對一般大眾與受影響之個人揭露AI系統詳細資料，但因評估之目的在於滿足AI系統的風險管制，此系統性透明度（systematic transparency）[318]堪屬能滿足風險評估所需之透明度。

德許可性，台灣法理學會，2019年12月31日，https://ivrtaiwan.wordpress.com/2019ailaw/（最後瀏覽日：2024年12月13日）。

316 關於專業技術上AI系統是否具備可解釋性，需考量到技術領域中對於可解釋性的定義與可解釋AI技術的發展，囿於篇幅限制，非本文能延伸討論之範圍。

317 林勤富，2024（註309），頁155-170。

318 同前註。

相對地，同法第86條肯認針對AI系統作成個別決定請求解釋之權利，雖有其突破但仍有不足之處。相較於GDPR僅規定個人能免受完全自動化決策影響[319]，缺少受影響者真正請求解釋權之具體規範[320]，《人工智慧法》第86條明確制定解釋權，賦予「受高風險AI系統影響之個人」要求部署者提供「針對影響該特定個人決定的解釋」的權利，且不限於完全自動化決策。雖《人工智慧法》試圖強化系統解釋的主觀可理解性[321]，然條文並未說明該解釋之目的與效果，若此解釋之目的係作為提供受影響者事後請求救濟時之依據，則此解釋應具備使個人可控訴之透明度（individually actionable transparency）[322]，但本法第86條「清楚且有意義的解釋」並未確保受影響者可取得事後進行救濟或訴訟所需之詳細資訊，讓受影響之人能以此有效挑戰該不利益之決定，若欲使第86條的解釋權發揮其救濟之作用，尚需仰賴前言對於第86條的說明，即部署者所提供的解釋應以受影響者得行使其權利之方式為之[323]。再者，該由誰判斷所提供之解釋是否夠清楚且有意義？此外，此請求解釋之權利與GDPR第22條免受完全自動化決策影響之差別與適用關係為何？以上問題皆凸顯《人工智慧法》在請求個案解釋權利的設計有失完善，使得請求解釋權之作用與效果仍不明確。然而，考量《人工智慧法》已禁止使用無法接受風險之AI使用情狀，對於高風險AI系統可解釋性之法規要求，是否須全面維持如此（技術上不一定能達成之）高標準，或是依不同應用透過指引或標準另定要求，也不無商榷空間。

319 GDPR art. 22.

320 *See* Wachter et al., *supra* note 122.

321 2023年版草案前言第84b點建議AI系統使用者所提供的解釋，應考量當事人對於專業知識的理解能力，然最後並未採納至《人工智慧法》最終條文。

322 同前註。

323 Artificial Intelligence Act recital 171.

（二）與AI系統互動的透明度

《人工智慧法》另有針對「與AI系統互動」之揭露規定，亦即對於「AI系統的存在、被使用以及與之互動」皆需向受影響者告知。在生成式AI與各種AI（偵測）系統的應用快速發展下，AI系統已全面走入人們日常生活、無孔不入，對話式、生成式AI在未來更將成為大多數AI系統與人類互動之介面。在此情況下，唯有認知到我們正與AI系統互動，才有可能避免受到系統的操縱或混淆。《人工智慧法》的相關制度設計包含第49條之登記制度與第50條特定AI系統供應商與部署者之透明度義務。在這個面向，《人工智慧法》目前的制度設計應有助於實現與AI系統互動的透明度要求。

《人工智慧法》第49條的登記制度之雛形，於2021年版草案即可窺見。原規範主體僅限於高風險AI系統之供應商與設立於歐盟境外之供應商授權代表[324]，《人工智慧法》最終條文則參考2023年版草案，新增公部門部署者的登記義務。詳言之，高風險AI系統的供應商應於上市銷售前或提供服務前於歐盟資料庫進行登記，登記內容包含供應商名稱、系統名稱、系統目的、系統是否上市銷售或正在提供服務、符合性評估效期，以及系統上市銷售的國家等資訊[325]，且登記至歐盟資料庫之內容應公開大眾近用[326]。除了供應商有登記義務外，若高風險AI系統部署者為公部門，該部署者亦應至資料庫進行登記[327]，非公部門之高風險AI系統部署者則無強制登記要求。至於非公部門之高風險AI系統部署者是否得自願登記至資料庫，條文未明定，而係於前言說明其他非公部門之高風險AI系統部署者亦得自願登記[328]。前述登記資訊將提供一般大眾取用，且公

---

324 Draft Artificial Intelligence Act 2021 ver. art. 51.

325 Artificial Intelligence Act annex VIII.

326 Artificial Intelligence Act art. 71(4).

327 Artificial Intelligence Act art. 49(3).

328 關於高風險AI系統部署者可自願登記之規定，2023年版草案建議增訂於登記制度的條文中，但《人工智慧法》最終條文並未採納，而僅於前言說明之。

開的資訊僅為AI系統的狀態而非內容（如：符合性評估效期、是否正在提供服務以及於哪些國家提供服務）[329]。其次，第50條規定特定AI系統供應商與部署者之透明度義務：若AI系統用於與人互動、情緒偵測或生成式AI時，AI系統部署者應告知與AI系統互動或可能受影響之人。雖然前述規定未打開法律或技術黑盒子，而僅使一般大眾知道黑盒子的「存在」，但確有助於強化AI系統資訊的流通、維持人民與AI部署者之資訊對等，以及避免AI系統所衍生的操縱與混淆風險[330]。

四、監理沙盒的管制碎裂化問題

新興科技持續發展與創新的特性使其管制存在許多挑戰[331]，如法規更新滯後、對科技創新過多管制束縛、消極管制所生風險與危害等難題。監理沙盒則有助於取得管制與創新間的新平衡，視監理沙盒的取徑與目的，管制者得於過程中確認當前管制框架的不妥適處（不管是過於嚴格或過於寬鬆）或擬定新的法規以彌補缺漏，參與企業則得於沙盒內獲得一定程度法規鬆綁、進行測試或強化其法遵措施。基此，監理沙盒的實驗性與管制環境為當今許多管制者面對新興科技採取的措施，包含歐盟會員國在內的許多國家已有施行監理沙盒之經驗，如：英國[332]與荷蘭[333]的金融監理沙盒，以及德國

Artificial Intelligence Act recital 131; Draft Artificial Intelligence Act 2023 ver. art. 51(1a) & (1b).

329 Artificial Intelligence Act annex VIII.

330 *See generally* Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1 (2019).

331 *See* Kenneth W. Abbott, *Introduction: The Challenges of Oversight for Emerging Technologies*, *in* INNOVATIVE GOVERNANCE MODELS FOR EMERGING TECHNOLOGIES 1 (Gary E. Marchant et al. eds., 2013).

332 United Kingdom Financial Conduct Authority, *Regulatory Sandbox*, https://www.fca.org.uk/firms/innovation/regulatory-sandbox (last visited Sept. 30, 2023).

333 De Nederlandsche Bank & Dutch Authority for the Financial Markets, *Continuing Dialogue—Innovation Hub and Regulatory Sandbox: Lessons Learned after Three Years* (2019), https://www.afm.nl/~/profmedia/files/rapporten/2019/innovationhub-rapport-eng.pdf.

自駕車、區塊鏈與能源轉型等監理沙盒[334]。AI為第四次工業革命的關鍵之一，已為許多開發國家必爭之地[335]，且已逐漸從單純的AI技術競爭，擴大為AI技術發展、標準制定以及管制法規間之競爭[336]。各國為了在AI領域取得領導地位，勢必酌其經濟社會情狀提出對自身有利的AI技術標準與管制措施以分配資源、管理風險、鼓勵或限縮AI的發展，而AI監理沙盒近年成為其中一種重要手段（雖然背後仍某程度反映鼓勵產業創新及企業發展之預設政策價值）。如OECD於2019年的《人工智慧建議》即指出政府可打造一管制環境供AI系統測試與規模化[337]，英國與挪威並分別於2019年設立資料創新使用之監理沙盒（其中包含使用AI的準則）[338]，以及於2020年成立關於資料使用與AI應用的監理沙盒[339]，西班牙則於2023年與歐盟執委會合作進行AI沙盒試行計畫[340]，均屬於這個面向的政策嘗試。

如前述，若歐盟對於監理沙盒欠缺適當調和，將可能衍生歐盟會員國間管制碎裂化等問題。基此，《人工智慧法》提出AI監理沙

334 Germany Federal Ministry for Economic Affairs and Energy, *Making Space for Innovation: The Handbook for Regulatory Sandboxes*, GERMAN FEDERAL MINISTRY FOR ECONOMIC AFFAIRS AND CLIMATE ACTION (July 1, 2019), https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/handbook-regulatory-sandboxes.pdf?__blob=publicationFile&v=1.

335 *See* SCHWAB, *supra* note 12.

336 *See generally* Smuha, *supra* note 269.

337 OECD, *supra* note 76, at 9.

338 United Kingdom Information Commissioner's Office, *Regulatory Sandbox*, https://ico.org.uk/for-organisations/advice-and-services/regulatory-sandbox/ (last visited Dec. 13, 2024); *see also* Jon Truby et al., *A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications*, 13 EUR. J. RISK REG. 270, 287-89 (2022).

339 Norwegian Ministry of Local Government and Modernisation, *National Strategy for Artificial Intelligence* (2020), https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf; *Regulatory Privacy Sandbox*, DATATILSYNET, https://www.datatilsynet.no/en/regulations-and-tools/sandbox-for-artificial-intelligence/ (last visited Sept. 30, 2023).

340 Spanish Ministry of Economy, Digital Transformation & European Commission, *Spain Proposes to Pilot an Artificial Intelligence Sandbox to Implement Responsible AI with a Human-Centric Approach* (June 27, 2022), https://ec.europa.eu/newsroom/dae/redirection/document/87973.

盒的規範，並要求歐盟執委會訂定監理沙盒施行法，以避免歐盟會員國間管制碎片化，調和內容包括監理沙盒的設立、運作、高風險AI系統申請進入與離開監理沙盒的要求、相關程序與法律效果等[341]。惟各式監理沙盒分散於各會員國，渠等仍掌握對於沙盒的實際運作的細節，其間仍可能因管制競爭而產生取徑相異的監理沙盒管制框架[342]而造成前述的種種問題，並影響歐盟AI治理的一致性與有效性。

監理沙盒間的差異可能產生的影響包括沙盒管理者管制套利[343]以及沙盒參與者挑選法庭[344]。管制套利的概念源於金融交易管制，形容金融機關透過降低交易成本與障礙以吸引投資與交易的作法[345]，並可能導致管制規範的「逐底競爭」（race to the bottom）[346]，而《人工智慧法》對於AI監理沙盒的設計即可能產生管制套利的現象。詳言之，考量AI開發與應用為各國競爭之領域，且本法強制各會員國皆至少設有一國家級AI監理沙盒，並明定AI系統離開沙盒時的評估報告將影響其符合性評估[347]，對此，各國將可能藉由其降低監理沙盒的進入與離開之條件，以吸引AI系統供應商申請進入其沙盒，因而導致歐盟AI監理沙盒規範以及AI系統的風險管理失去一致性。挑選法庭則為國際私法上常見的現象，即訴訟當事人選擇對其最有利的法院為管轄法院[348]，有利的原因可能包含適用之法律、判

341 Artificial Intelligence Act art. 58.

342 *See* Truby et al., *supra* note 338, at 279, 284.

343 Madiega & Van De Pol, *supra* note 270, at 3.

344 Organisation for Economic Cooperation and Development (OECD), *Regulatory Sandboxes in Artificial Intelligence* 8 (OECD Digital Economic Papers No. 356, 2023), https://doi.org/10.1787/8f80a0e6-en.

345 *See* Frank Partnoy, *Financial Derivatives and the Costs of Regulatory Arbitrage*, 22 J. CORP. L. 211, 227 (1997); *see also* Annelise Riles, *Managing Regulatory Arbitrage: A Conflict of Laws Approach*, 47 CORNELL INT'L L.J. 63, 65 (2014).

346 Joel F. Houston, Chen Lin & Yue Ma, *Regulatory Arbitrage and International Bank Flows*, 67 J. FIN. 1845, 1846 (2012).

347 Artificial Intelligence Act art. 57(7).

348 *See generally* Christopher A. Whytock, *The Evolving Forum Shopping System*, 96 CORNELL L. REV. 481 (2011).

例與訴訟程序等。雖挑選法庭為當事人訴訟策略之一部分，未必會產生負面影響[349]，惟挑選法庭的現象若發生於沙盒參與者，將影響AI監理沙盒的運作與功能。沙盒參與者挑選法庭與沙盒管理者管制套利現象所產生的影響可能相互加成，考量到不同沙盒規範對其欲開發或使用AI系統之利弊，沙盒參與者將會選擇有利於其系統開發與法遵之沙盒，並可能因而刺激沙盒管制者隨之改變其措施，使得管制套利進一步加劇。

AI監理沙盒的規範是否適當、與AI法以及各國法律是否有適當調和，亦為潛在問題。《人工智慧法》允許會員國間共同設立一監理沙盒，此雖能促進會員國之間的合作與協調，然同時也伴隨不同法制框架的複雜性[350]。此外，AI的管制規範與AI的技術標準具高度關聯性[351]，因此AI的管制應與各項技術標準有更密切溝通。而監理沙盒作為具實驗性的管制措施，應利用其彈性的特質使AI管制規範與技術標準有更好的互動與影響。然觀本法對於沙盒的規範設計，其目的似僅著重於輔助AI系統供應商進行法遵（即符合《人工智慧法》對於高風險AI系統的各項要求）[352]，而欠缺回應或改善AI管制規範的功能。因此，《人工智慧法》之AI監理沙盒規範雖有助於促進高風險AI系統上市銷售前或提供服務前的法遵，然可能因欠缺沙盒間的調和而對歐盟AI管制增加額外的成本與風險[353]。

---

349 *See* Kevin M. Clermont & Theodore Eisenberg, *Exorcising the Evil of Forum-Shopping*, 80 CORNELL L. REV. 1507, 1514-17 (1995).

350 *See* Datasphere Initiative, *Sandboxes for Data: Creating Spaces for Agile Solutions Across Borders* (2022), https://www.thedatasphere.org/wp-content/uploads/2022/05/Sandboxes-for-data-2022-Datasphere-Initiative.pdf (last visited Sept. 30, 2022).

351 OECD, *supra* note 344, at 9.

352 Artificial Intelligence Act art. 57(5).

353 OECD, *supra* note 344, at 14.

肆、全球規範趨勢下我國人工智慧治理困境與未來

本文對於歐盟《人工智慧法》之評析，凸顯AI治理各種層次之不同挑戰，除須處理重要議題如風險控管判準與界線之拿捏、人類監督機制與具體作法，亦須考量AI系統在法律面與技術面不同的透明度問題，乃至於鼓勵創新之監理沙盒的設計，亦有諸多管制碎裂化、競爭與套利之考量，種種問題絕非於短時間內便可取得完美平衡與良善治理模式，而是需要長期的研究、整體性的管制策略、部門立法試誤（trial and error）與各方對話。近年來，隨著各種生成式AI工具走入大眾生活，臺灣各界對於AI應用與管制的討論也大量增加。就具體管制作為部分，多以各種「原則」、「指引」與「芻議」，或是各種特設性（*ad hoc*）的軟法，尚未形成足夠之共識與立法動能以制定具有拘束力之規範，使得AI治理仍顯不足且碎裂化。

歐盟採取硬法取徑，並已通過全球第一部全面性AI專法，考量AI發展與應用之跨國界性質、國內管制需求、國際經貿實務、管制競爭或管制學習等原因，該法對於我國現有或未來制定相關規範時，不論是採硬法或軟法取徑，均可能具有參考示例價值（reference power）；我國近年AI相關立法草案，也都明確提及參酌歐盟AI法制[354]。又縱使相關部會暫無立法規劃、或不採取《人工智慧法》之管制途徑，然布魯塞爾效應之潛在影響仍可能使《人工智慧法》對於我國AI供應商或部署者等產生直接或間接之拘束力或影響力，是以仍有需要進一步討論我國未來法治與歐盟法的互動關係、落差與調和等議題。以下將簡要回顧我國近年之AI治理作為，提出若干觀察與初步建議，並與歐盟作法相互參照，期藉本文初步見解而充實相關研究與對話。

354 例如民眾黨黨團於2022年提出之《人工智慧發展法草案》，以及於2023年提出之《人工智慧發展及管理條例草案》均是。

一、偏倚產業創新而忽視管制需求

我國過去對於特定AI議題之監管需求回應速度堪稱迅速，但多半集中在鼓勵創新與測試的面向。例如，經濟部2018年3月提出草案，同年11月經立法院通過之《無人載具科技創新實驗條例》（下稱「無人載具條例」），被稱為「全球第一個針對自駕車、無人機及自駕船建立無人載具監理沙盒」[355]。2019年時，科技部所提出的《人工智慧科研發展指引》，據稱亦為「世界第一個給科研人員的人工智慧研發準則」[356]，乃自願性準則性質的規範。隨著AI應用逐漸成熟與範圍擴張，各部門亦開始訂定相關規範，例如衛福部於2020年訂定之《人工智慧／機器學習技術之醫療器材軟體查驗登記技術指引》，以及訂於2022年、復於2023年修正之《人工智慧／機器學習技術之電腦輔助偵測（CADe）及電腦輔助診斷（CADx）醫療器材查驗登記技術指引》，要求醫療器材業者於登記時，提供產品描述、安全與效能驗證等資訊，在透明度方面亦有初步要求。

而隨著ChatGPT引發大眾關注，行政院已於2023年8月通過《行政院及所屬機關（構）使用生成式AI參考指引》[357]，禁止業務承辦人使用生成式AI撰寫機密文書，亦不得以生成式AI產出資訊「取代業務承辦人自主思維、創造力及人際互動」、以未經確認之產出內容直接作為行政處分之依據等[358]，但適用範圍僅限於行政院所屬機關。此外，金管會亦提出《金融業運用人工智慧（AI）之核心

355 財團法人台灣經濟研究院，2022產業技術白皮書，頁161（2022年）。

356 簡惠茹，避免AI濫用 科技部訂出世界第一個給研發人員的AI指引，自由時報，2019年9月23日，https://news.ltn.com.tw/news/life/breakingnews/2923939（最後瀏覽日：2024年12月13日）。

357 行政院，行政院及所屬機關（構）使用生成式AI參考指引，2023年8月31日，https://www.ey.gov.tw/Page/448DE008087A1971/40c1a925-121d-4b6b-8f40-7e9e1a5401f2（最後瀏覽日：2023年9月10日）。

358 公共政策網路參與平臺，行政院及所屬機關（構）使用生成式AI參考指引草案，歡迎您提供建議與看法，2023年7月18日，https://join.gov.tw/policies/detail/9f21a1e6-edd0-45e9-8ad6-b92354abb1fa（最後瀏覽日：2023年7月18日）。

原則與相關推動政策》，列舉包括「建立治理及問責機制」在內之6項核心原則，期藉此控管金融機構對AI使用之風險控管[359]。最後，數位發展部並於2023年底成立「AI產品與系統評測中心」[360]，鼓勵AI業者藉此平臺驗證其AI系統之「準確性、可靠性、公平性、可解釋性及安全性等」[361]。

在AI相關立法方面，除了無人載具條例之外，則曾有許毓仁委員於2019年5月提出之《人工智慧發展基本法》草案（以下簡稱2019基本法草案），其中提及政府應在推動AI發展的同時「促成人工智慧發展倫理規範之制定」，並訂有沙盒機制，以及成立人工智慧特別委員會專職處理AI的研發與倫理爭議[362]；隔年立院換屆，由鄭麗文委員所提出的新版本中（以下簡稱2020基本法草案），更納入了「網路人權」之概念[363]。而民眾黨黨團則是分別於2022年與2023年提出《人工智慧發展法草案》與《人工智慧發展及管理條例草案》，兩者均要求主管機關對AI發展制定相關倫理審查機制[364]。

359 金融監督管理委員會，金管會公布金融業運用人工智慧（AI）之核心原則及政策，2023年10月17日，https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0%2C2&mcustomize=news_view.jsp&dataserno=202310170002&dtable=News&fbclid=IwAR0jVkJDOaeu4_NAWO_mfPA9RYYPuzWhzQuZR3sv4FAnen80BFcJNbRVsMo_aem_AdBnlFwGHM1tdrnnG7MNgljFCDGT7yEFrMP00Ew5Dlkk9-zAZpDKRnRrNzRFN4fOCuY（最後瀏覽日：2023年10月18日）。

360 AI產品與系統評測中心，數位發展部「AI產品與系統評測中心」啟動 推動我國AI評測制度與可信任AI環境發展，2023年12月6日，https://www.aiec.org.tw/News/Content?type=bulletin（最後瀏覽日：2024年12月14日）。

361 徐子苓、歐宇祥，數位部年底成立AI評測中心 確保AI安全性 鼓勵產業導入應用，自由時報，2023年6月9日，https://ec.ltn.com.tw/article/paper/1587490（最後瀏覽日：2024年12月13日）。

362 立法院第9屆第7會期第14次會議議案關係文書，院總第1021號委員提案第23371號，2019年5月15日，https://ppg.ly.gov.tw/ppg/download/agenda1/02/pdf/09/07/14/LCEWA01_090714_00037.pdf。

363 立法院第10屆第2會期第1次會議議案關係文書，院總第1021號委員提案第24935號，2020年9月16日，https://lis.ly.gov.tw/lygazettec/mtcdoc?PD100201:LCEWA01_100201_00025。

364 立法院第10屆第6會期第2次會議議案關係文書，院總第1021號委員提案第28966號，2022年9月28日，https://lis.ly.gov.tw/lygazettec/mtcdoc?PD100602:LCEWA01_100602_00064。以及立法院第10屆第8會期第11次會議議案關係文

不過，以上法案最終都未獲通過。就行政部門方面，行政院原規劃於2023年9月提出《人工智慧基本法》院版草案，但因AI發展腳步快速，行政院已決定延後立法[365]，以研擬公部門使用AI指引為優先[366]，後於2024年7月正式提出《人工智慧基本法》國科會版草案，指出草案推動方向係依我國產業與社會需求，並參考美國鼓勵創新發展及歐盟維護人民權利之取徑[367]。

整體而言，我國目前除甫提出基本法草案含納若干一般性原則外，並未如歐盟一般以全面性專法的硬法取徑建構AI治理。此一作法在政策上偏重經濟利益、鼓勵AI創新與產業發展，而暫時迴避直球對決法律挑戰的誘因（特別是對高風險以及不可接受風險之AI系統的立場），推動實質立法之政治動能不高，導致目前管制遲滯的現況。與歐盟《人工智慧法》相比，我國目前對於AI法制的進度至多僅及於《人工智慧法》第95條的「行為準則」，又個別部門基於實際管制需求、機關間競爭、政策績效或私部門壓力，時有特設性的政策作為，但又因無具體管制方案或明確授權（mandate）而以軟法為主，實際規範意義有限[368]。尤其如警政署推行之M-Police[369]、

書，院總第20號委員提案第10043021號，2023年12月6日，https://ppg.ly.gov.tw/ppg/SittingRelatedDocumentReportMatter/download/agenda1/02/pdf/10/08/11/LCEWA01_100811_00042.pdf。

365 徐子苓，人工智慧發展太快 鄭文燦：AI基本法延後推出，自由時報，2023年8月1日，https://ec.ltn.com.tw/article/breakingnews/4382694（最後瀏覽日：2024年12月13日）。

366 張瓊，吳政忠：衝酌產業發展研議AI基本法 公部門指引先行，中央社，2023年7月18日，https://www.cna.com.tw/news/aipl/202307180300.aspx（最後瀏覽日：2024年12月13日）。

367 公共政策網路參與平臺，國家科學及技術委員會公告：預告制定「人工智慧基本法」草案，2024年7月15日，https://join.gov.tw/policies/detail/4c714d85-ab9f-4b17-8335-f13b31148dc4（最後瀏覽日：2024年9月27日）。

368 Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 NATURE MACHINE INTELLIGENCE 501, 503 (2019).

369 胡欣男，M-Police人臉辨識暫時下架 警政署：依國發會建議補強法規，中時新聞網，2021年12月26日，https://www.chinatimes.com/realtimenews/20211226002896-260402?chdtv（最後瀏覽日：2024年12月13日）。

或是司法院啟用評價型與事實型量刑資訊系統[370]、使用生成式AI撰寫刑事裁判草稿[371]等作為，在歐盟《人工智慧法》下可能屬於禁止使用的AI使用情狀或是高風險AI系統，但在臺灣卻得以於缺乏法律授權之情況上路，是否妥適？不無疑問。即便是要促進創新產業發展，有時也會需要部門立法才能有確定性，例如生成式AI大量運用受著作權保護之文字圖像作為訓練資料，是否違反著作權相關規定，經濟部智慧財產局亦宜提出政策方針、函釋或行為準則，避免法規空間之長期模糊。

本文建議，在目前AI發展與應用迅速且複雜多變之時，我國並非必須立即制定全面性專法，而可參酌採納軟法與硬法混合（hybrid）之方式。目前以軟法為主軸的管制取徑（即若後續通過《人工智慧基本法》，雖形式上為硬法，實質規範內容上仍偏向軟法取徑），雖能最大程度免除管制障礙、鼓勵產業創新，但部分議題實有以硬法取徑制定明確規範的急迫性[372]。詳言之，雖因考量產業發展以及科技不確定性，而不欲採取積極硬法取徑以全面性立法介入AI發展與應用[373]，但仍有其他方式可強化AI治理，例如鼓勵供

370 司法院於2023年2月宣布啟用「AI量刑資訊系統」，將基於「事實型模式」與「評價型模式」整理過往判決資料，並基此「提供職業法官與國民法官查詢類似案件的量刑趨勢，讓量刑更公正透明」。見司法院，因應國民法官新制，司法院啟用AI量刑資訊系統——具備二種模式、擁有四大優點，2023年2月6日，https://www.judicial.gov.tw/tw/cp-1887-806741-d6471-1.html（最後瀏覽日：2023年10月30日）。

371 司法院於2023年8月就生成式AI之使用進行說明，指出司法院在毒品罪、民事交通損害賠償、消債更生與清算等案件中，規劃使用生成式AI撰寫裁判書草稿。見司法院，司法院審慎發展生成式AI應用，以撰寫刑事裁判草稿初試啼聲；期望減輕法官工作負荷，審判核心仍由法官自行決定，2023年8月27日，https://www.judicial.gov.tw/tw/cp-1887-929494-8a9fb-1.html（最後瀏覽日：2023年10月30日）。

372 *See* Gary E. Marchant & Carlos Ignacio Gutierrez, *Indirect Enforcement of Artificial Intelligence 'Soft Law'*, TPRC48 RES. CONF. ON COMM. INFO. & INTERNET POL'Y (2020), https://ssrn.com/abstract=3749776.

373 因為過度嚴格的監管作為，確實可能阻礙AI產業的發展而不利於整體經濟發展。*See Transatlantic Relations in AI Policy*, REGULAITE, https://www.regulaite.eu/

應商開發及部署者運用AI系統時，符合行為準則、倫理原則、標準化或技術指引等，或提供不同管制誘因以促使建立第三方驗證系統。軟法規範並非用於取代硬法管制[374]，而是透過實務上踐行、累積經驗與資訊，有助於在具科技不確定性的議題領域中，由下而上發現（identify）並建構必要的硬法管制。此外，同樣作為硬法管制之基礎，應可透過AI相關技術的標準化以及AI系統之安全標準，使AI的管制得更符合技術領域之現況。

與此同時，有些AI發展與應用情狀，仍有積極以硬法管制的急迫性。我國應基於對基本權利的保障、法治原則、選舉公正性或民主制度的維護等重要價值，並參酌歐盟《人工智慧法》之制度設計，考量是否立法禁止對基本權利侵害過鉅而有不可接受的風險之AI系統或應用，如基於執法目的而於公開場所使用即時遠端生物辨識系統等，或至少訂定嚴格的使用條件與許可制度。此外，作為未來進一步強化AI治理的基礎，我國可考慮透過要求登記與影響評估，確保使用AI系統的最低透明度，例如要求公部門就AI應用情形進行登記（registration）並建立資料庫（inventory），據以盤點並揭露運用現況，以及要求AI系統開發者或部署者進行基本權利影響評估等，雖無法完全解決AI造成之風險，但當開發者與部署者需要執行評估並公布評估結果時，不僅有助於加強透明度、掌握AI風險，更讓開發者有機會從社會影響的角度檢視其產品並做出適當調整[375]。

---

ii-the-transatlantic-axis (last visited Oct. 18, 2023); *see also* Jorge Buenfil et al., *Artificial Intelligence Ethics: Governance through Social Media*, 2019 IEEE INT'L SYMP. TECH. HOMELAND SEC. 1, 1-6 (2019), http://www.hstabstractbook.org/index_htm_files/c-IEEE_HST___AI_Ethics__Governance_through_Social_Media%20FINAL.pdf.

374 *See* Nicholas Kluge Corrêa, *Blind Spots in AI Ethics and Biases in AI Governance*, 13 (2020), https://philarchive.org/archive/CORASv3; *see also* Axel Walz & Kay Firth-Butterfield, *Implementing Ethics into Artificial Intelligence: A Contribution, from a Legal Perspective, to the Development of an AI Governance Regime*, 18 DUKE L. & TECH. REV. 176, 215-16 (2019).

375 *See* Soojin Jeong, *Assessing Algorithms for Public Good*, THE REGULATORY REVIEW

無論如何，在國際規範與標準快速發展的環境下，觀望國際動態而未有具體動作，將可能使臺灣未來僅能被動接受其他國家或組織之既定AI規範，因此縱使管制的強度與時間仍未能有明確解方，儘早建立基於核心基礎價值的管制框架、依此促進多方利害參與者之對話，並隨國際共同趨勢逐步調整，似為上策[376]。

二、欠缺整體管制策略與部門立法

相比前述歐盟之AI治理相關作為，我國基於種種在地考量，現階段對於AI治理較傾向原則、指引與芻議類型的軟法，至於未來如何循序建構完整規範與制度，究欲效法歐盟以風險為基礎並管制AI使用情狀，亦或是仿效美國透過行政命令以及責成NIST整合公私部門資源以規範AI發展，目前仍待更充分的管制策略與討論，即使是《人工智慧基本法》草案也係以概括方法指出其同時參考美國鼓勵創新發展及歐盟維護人民權利之作法。而個別部門雖因實際管制需求與政策競爭等考量，設有零星特設性的管制作為，但仍欠缺系統性、跨部會協調後的整體管制方向。中央立法部分，我國對於AI監管之作為偏向謹慎保守。雖然早在2019年便已有立委提出《人工智慧發展基本法》草案，但到了2023年，行政院在4月核定之《臺灣AI行動計畫2.0（2023-2026年）》中，即使三度提及對「AI倫理法制」之重視，就實質作為部分，仍僅著眼於制定AI基本法[377]。雖有論者認為，基本法之制定有助於彰顯我國重要價值與政策取向，

(Oct. 10, 2023), https://www.theregreview.org/2023/10/10/jeong-assessing-algorithms-for-public-good/.

376 Gregory N. Mandel, *Regulating Emerging Technologies*, 1 LAW INNOVATION & TECH. 75, 88-91 (2009).

377 行政院，臺灣AI行動計畫2.0（2023-2026年），頁27，2023年4月7日，https://digi.nstc.gov.tw/File/7C71629D702E2D89（最後瀏覽日：2024年12月14日）。公共政策網路參與平臺，國家科學及技術委員會公告：預告制定「人工智慧基本法」草案，2024年7月15日，https://join.gov.tw/policies/detail/4c714d85-ab9f-4b17-8335-f13b31148dc4（最後瀏覽日：2024年9月27日）。

並確認既有法規因應AI發展之修訂方向[378]，惟相對於歐盟已通過全面性AI專法，韓國、加拿大等國已經著手制定類似法規，而對於AI監管採取保守不干預態度的美國似也將有所調整[379]，我國現階段仍以宣示意義大於實質意義之基本法為立法重點，似已落後國際趨勢，或可參考前述軟法與硬法混合之方式積極研擬管制策略。

以國民黨黨團於2024年4月所提出之《人工智慧基本法草案》為例，其內容參照歐盟《人工智慧法》等各國軟硬法例，涵蓋層面包括勞工權益保障、公平交易維護、救濟補償與保險制度、風險評估及監管機制，以及品質管理機制等[380]，但整體仍僅就前述面向概括納入原則性、寬泛之規範方向，具體規定仍待各機關另行制定，對於AI治理恐難造成實質改變，亦無法著眼於AI在不同使用情境或工具的情況，制定更為貼近監管需求之規範[381]。相對於此，民眾黨黨團過去所提之《人工智慧發展法草案》與《人工智慧發展及管理條例草案》，其內容納入了資料治理與個資保護、風險分級、環境保護相關條文，且明文禁止政府以維護社會秩序為名利用大數據進行大規模監控、或應避免對特定族群產生偏見等，並就部分條文之違反訂有罰則[382]，相較於基本法似有若干更為具體、且更能有效掌控AI衍生之風險。但我國現階段仍以一般性原則、實質操作性不高

---

378 張麗卿等，臺灣人工智慧基本法制定之必要與倡議，月旦法學雜誌，340期，頁87（2023年）。

379 *See* Bradford, *supra* note 43.

380 《人工智慧基本法草案》，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，2024年4月10日，第11、12、13、15與16條，https://ppg.ly.gov.tw/ppg/download/agenda1/02/pdf/11/01/09/LCEWA01_110109_00007.pdf。

381 有關AI部門立法之討論，亦可參見郭戎晉（註104），頁33-34。

382 以2022發展法草案為例，資料治理與個資保護之規定可見於2022發展法草案第8、9、21與23條；風險分級與環境保護部分見同草案第10條與第12條；禁止政府大規模監控，以及罰則規定，則分別訂於同草案第13條與第25條至第27條。2023條例草案則刪除禁止政府大規模監控之條文，並新增強化透明度、避免偏見、不當蒐集資料等規定。

的基本法為AI治理之政策重點，是否足以回應各種類型及程度之AI管制需求，不無疑問。更有甚者，政府未來若面對AI法制建構之公眾壓力，迅速制定「基本法」或可暫時緩解公眾壓力，但可能反而會因政治動能（political momentum）與誘因減少，忽略真正最需要的部門（sectoral）作用法立法，未能針對具體的AI議題、並適時考量個別領域AI發展與應用不同程度之效益與風險而予以適當規範。

即令未有專門的AI作用法立法規畫，我國在中央層級亦未訂有如同美國《人工智慧應用監管指引》與《推動安全可信任的AI發展與運用》，以及美國人工智慧安全研究院等能引導各機關的規範訂定方針，使得各單位使用AI或監管AI之作為缺乏一致標準，特設性的作為仍聚焦於鼓勵產業發展，而整體管制策略，仍欠缺系統性規畫。另一方面，在缺乏中央立法授權的情況下，各主管機關也無以制定明定權利義務的作用法[383]，導致現有規範均屬程序性或原則性之規定或指引，欠缺具體規定，至多僅有衛福部得要求業者揭露相關資訊，對於特定高風險AI系統的監管力度恐有不足[384]。相對地，隨歐盟《人工智慧法》所成立的AI Office，得針對AI相關規範事項發布指引或採取其他行動，或針對監理沙盒之運作提供技術建議與協調、維護資料庫、促進主管機關間的資訊交流、跨機關合作及國際合作等。而在美、英等國亦有類似的跨部會機制與辦公室。我國或可考量透過更制度性的安排，以更有系統性、跨部門的方式，盤點最應優先施力的AI規範治理問題，並促成資訊交流與合作，以利未來立法作為。

---

383 蓋基本法僅有宣示重要原則、指示主管機關訂定相關規範之效果，但若要落實對於AI系統之管制作為，仍須訂有作用法規定各方之權利義務以及違反效果，始能滿足法律保留原則。見李震山，行政組織法與行政作用法之區別及其實益，月旦法學雜誌，85期，頁16-17（2002年）。

384 翁逸泓（註37），頁304-305；陳陽升，從法治原則探索人工智慧之應用界限，台灣法律人，26期，頁106（2023年）。

三、聚焦個人資料保護並無法取代AI法制

我國目前對於AI治理與法制之討論，似多偏向AI衍生問題的課責問題，對於事前管制則較少著墨[385]，即使是後者，多也聚焦於資料與隱私治理之面向[386]。從歷年相關政策或基本法草案均可發現類似情況。例如2018年公布的《臺灣AI行動計畫（2018-2021年）》，將資訊安全、資料開放與個資保護，列為法規鬆綁與調適之首要問題[387]。同年通過的無人載具條例，也在其中實驗場域管理及安全乙節，特別強調資安措施與個資法遵[388]。而在2019基本法草案中，除

385 如李建良，智慧駕駛的法學思辨——自動駕駛法學的試車之旅，月旦法學雜誌，327期，頁101（2022年）。事實上，對於我國個資保護法制，已有類此對於事前管制力道不足批評，例如張容涵，從比較法觀點探討我國個資保護制度之轉型，國立臺灣大學法律學研究所碩士論文，頁172-173（2019年）。另此狀況可能起源於國家對於產業法的管制思維，原則上傾向避免採取事前監管之手段，僅有在市場競爭機制無法發揮功能時，才會介入市場運作，見黃銘傑，徘徊於事後管制v.事前管制之數位時代中的競爭法主管機關，台灣經濟研究月刊，46卷7期，頁96-97（2023年）。不過，在AI治理方面，不乏有學者認為應與歐盟看齊，以預防原則之精神來應對AI可能帶來之風險，見林昱梅，預防原則於人工智慧之適用——歐盟人工智慧規則草案及其基本原理之觀察，台灣法律人，12期，頁14-15（2022年）。

386 在探討AI治理或規範的文獻中，常見以資料治理作為探討重點之情形。例如李建良（註385），頁102；張麗卿，AI倫理準則及其對臺灣法制的影響，月旦法學雜誌，301期，頁104-105（2020年）。林昱梅（註385），頁16。此外，科技部所推動之「回應國家重要挑戰之人工智慧主題研究專案」，亦將「資料治理」與「核心技術」、「領域應用」及「社會倫理」並列為重點議題，臺灣AI卓越中心基此設立的17件專案中，涉及法律議題之「聰明生產與智慧精準製造之數位決策、AI模型、大數據治理與核心技術之整合製造平台及其科技法律與產業生態系統」計畫，亦特別凸顯資料治理之議題。見科技部（110）科部前字第1100014912號函，以及臺灣AI卓越中心，AICoE專案：聰明生產與智慧精準製造之數位決策、AI模型、大數據治理與核心技術之整合製造平台及其科技法律與產業生態系統，https://www.twaicoe.org/聰明生產與智慧精準製造之數位決策-ai模型-大數據治理與核心技術之整合製造平台及其科技法律與產業生態系統（最後瀏覽日：2024年12月14日）。

387 行政院，臺灣AI行動計畫（2018-2021年），頁38，2018年6月29日，https://digi.nstc.gov.tw/File/4C622B6A10053DAD（最後瀏覽日：2020年9月30日）。而在2023年提出的《臺灣AI行動計畫2.0（2023-2026年）》中，也將資料治理列為未來推動計畫的總體目標之一，藉以「完善運作環境」，但對於AI風險則隻字未提。行政院（註377），頁17-18。

388 無人載具科技創新實驗條例第16、17條。

了要求開發者遵循透明性與使用者中心等四原則、要求朝可信任的方向發展AI外，亦納入資料之蒐集與隱私保護相關規定，並要求在AI開發涉及「高敏感性個資」或「弱勢族群之資料」時採用特別審核機制[389]。相較於AI運用可能會對廣泛不同權利所產生之風險，我國立法上之討論似過度偏重AI所涉及之個資保護面向，多過於AI運用可能造成之風險[390]。直到2022年國科會出版之《2022臺灣AI國力調查》，仍可看到將個資保護與資通安全作為控管AI風險之主要切入點[391]，顯示我國對於AI治理之思維，仍多聚焦在個資與資安風險之控管。

誠然，由於AI的訓練仰賴大量資料，使AI與資料治理密不可分，對於資料之管制作為，勢必也將對AI系統產生影響[392]。對於AI容易在開發過程中衍生各種層次的個資保護問題，不僅GDPR設有若干規範，《人工智慧法》亦對此有所著墨。不過，AI之技術特性或利用方式，不僅只會對個人資料產生危害，更可能會違反正當法律程序、加深偏見，或侵害其他類型的基本權利，已如前述。是以，個資治理與AI治理雖均密切涉及資料治理，但兩者不能混為一談，前者之目的在於保障公眾對於個資之自主控制權，並掌握其個資之使用方式[393]，也可限縮AI系統濫用個資所造成危害[394]；而後者則是確保AI之發展與應用不會造成負面衝擊，且不限於對於個資之危害[395]。

---

389 詳見2019基本法草案第7、8、9條與第12、13條。

390 隔年由鄭麗文委員所提出之草案，內容大多類似，不再另作討論。

391 國家科學及技術委員會，2022臺灣AI國力調查，頁29-30，2022年12月13日，https://digi.nstc.gov.tw/File/EF7772C7B6116E10/6bce6d0f-11b5-48eb-959d-fc02786b0e5f?A=C（最後瀏覽日：2024年12月14日）。

392 John Hillman, *Smart Regulation: Lessons from the Artificial Intelligence Act*, 37 EMORY INT'L L. REV. 775, 801 (2023).

393 *Id*. at 808.

394 *See* William Davidow & Michael S. Malone, *Don't Regulate Artificial Intelligence: Starve It*, SCI. AM. (May 4, 2020), https://blogs.scientificamerican.com/observations/dont-regulate-artificial-intelligence-starve-it/.

395 *See* Storm & van der Wolk, *supra* note 120.

過去適用GDPR規定之AI開發公司，在《人工智慧法》通過之後將面臨更多新類型的法遵要求，也凸顯了兩法制本質上之不同，蓋GDPR並不直接限制AI系統使用資料的方式[396]。

另一方面，運用資料之敏感程度與AI風險高低程度實屬二事。如同GDPR對敏感性個資給予高度保護[397]，除保障更多核心法益外，亦避免資料之使用產生歧視性等負面結果[398]，我國個人資料保護法亦將病歷、醫療、基因、性生活、健康檢查及犯罪前科列為敏感性個資並區分保障程度，原則上不得蒐集、處理或利用[399]。以醫療診斷AI為例，其用以訓練之資料多屬資料法制給予高度保護之敏感性個資，使其在訓練資料之蒐集上須滿足較高的法規要求，惟於應用上可能並不落入高風險AI範疇[400]。此外，在近年「推論經濟」（inference economy）[401]爆炸性發展的脈絡下，敏感資料仍可能透過演算法與大數據從非敏感資料中推導出來，進而突破原先個人資料保護法之舊管制框架與規範設計[402]。是以資料治理固為必要基礎規範，且面對AI發展與應用的各種管制挑戰，我國目前的個人資料

396 *Id.*

397 李惠宗，健保資料庫的合理利用與個資保護——從歐盟一般資料保護規則的「資料治理」探討憲法法庭111年憲判字第13號判決未解決問題，台灣法律人，24期，頁8（2023年）。

398 *See* Michele E. Gilman, *Five Privacy Principles (from the GDPR) the United States Should Adopt to Advance Economic Justice*, 52 ARIZ. ST. L.J. 368, 420-21 (2020).

399 個人資料保護法第6條。

400 相關風險除了AI運作不良而對人體或醫師判斷造成影響外，尚包括因演算法之不透明性，使得主管機關難以確實理解、審查AI產品之運作方式等。*See generally* Douglas McNair & W. Nicholson Price II, *Health Care Artificial Intelligence: Law, Regulation, and Policy*, *in* ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 197, 207-212 (Michael Matheny et al. eds., 2022); *see also EU AI Diplomacy towards China*, REGULAITE, https://www.regulaite.eu/iii-eu-china-relations-in-ai-regulation (last visited Oct. 18, 2023).

401 *See generally* Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357 (2022).

402 Daniel J. Solove, *Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data*, 118 NW. U. L. REV. 1081, 1085-86 (2024).

保護法制已顯不足而過度軟弱、符合AI時代的全面性資料治理法制更屬欠缺。此等資料治理基礎規範亟待重新建構，然仍須留意，單純聚焦於狹義的個資保護問題，雖為必要基礎而有支持、補充之效，但無法取代AI法制[403]。就以歐盟法為例，由於GDPR並不直接限制AI系統使用資料的方式而可能產生規範罅隙，《人工智慧法》之通過便能補充GDPR、填補歐盟AI監管之不足[404]。我國未來若僅以個資保護之思維推動AI法制與治理，可能導致監管焦點與資源之錯置，無法確實應對AI可能帶來之問題[405]。

四、多方利害關係人參與制度不足

我國現階段在AI法制與治理部分缺乏整體完善策略等種種問題，可能是因為在相關政策與規範之形塑過程中，多方利害關係人之參與不足，此點也是各國在制定AI政策與法律時共通之問題[406]。以行政院之《行政院及所屬機關（構）使用生成式AI參考指引》為例，雖然行政院表示該指引之研擬係經「與AI技術、法制專家學者及12個相關部會協作」並「於公共政策網路參與平臺徵詢民眾意見」[407]，惟實際參與討論之管道並未公開，而於該公共政策網路參與平臺之意見收集期間僅有10日，且表達意見之民眾僅有4位[408]，相較歐盟《人工智慧法》制定時進行為期4個月的公眾諮詢，與美國NIST訂定《AI風險管理框架》時開放為期1個月公眾討論與回

403 Hillman, *supra* note 392, at 807-13.

404 *See* Storm & van der Wolk, *supra* note 120.

405 以AI偏見為例，雖然係已受到普遍承認之問題，但有學者指出我國對此的規範仍然不足，個人資料保護法並無法提供完善規範。參見翁逸泓（註37），頁304。

406 Susan Ariel Aaronson & Adam Zable, *Missing Persons: The Case of National AI Strategies* (CIGI PAPERS No. 283, Sept. 21, 2023), https://www.cigionline.org/publications/missing-persons-the-case-of-national-ai-strategies/.

407 行政院（註357）。

408 公共政策網路參與平臺（註358）。

饋[409]，以及NIST為回應拜登總統的行政命令《推動安全可信任的AI發展與運用》與隨後成立的美國人工智慧安全研究院，公開徵詢公眾討論與意見[410]，我國之作法似難確實反映專家、業者、公民社會或大眾之意見。過去衛福部於訂定AI醫療器材相關指引時，立法院之研究員亦認為須加強專家參與機制，以確保醫療器材分類之有效性[411]，顯見在AI法制與治理方面，利害關係人參與不足乃為跨部門的共通議題。

考量AI技術發展日新月異且AI規範不僅涉及法律，更涉及技術標準之制定[412]，歐盟《人工智慧法》即指出標準制定過程中業界利害關係人參與的重要性[413]。我國因為國際現勢而不若他國有充分參與國際規範與標準制定及討論之機會[414]，更應強化各方利害關係人參與之管道與制度，才能確保政府在制定相關規範時能納入足夠專業能量，確保AI法制能兼顧風險管理與鼓勵創新[415]。此外，考量AI之風險複雜性與不定性，政府於制定AI法律與政策時擴大公眾參

409 U.S. National Institute of Standards and Technology, *AI Risk Management Framework: Second Draft, Notes for Reviewers: Call for Comments and Contributions* (Aug. 18, 2022), https://www.nist.gov/system/files/documents/2022/08/18/AI_RMF_2nd_draft.pdf; U.S. National Institute of Standards and Technology, *Comments on the 2nd Draft of AI Risk Management Framework* (Nov. 8, 2022), https://www.nist.gov/itl/ai-risk-management-framework/comments-2nd-draft-ai-risk-management-framework.

410 U.S. National Institute of Standards and Technology, *NIST Calls for Information to Support Safe, Secure and Trustworthy Development and Use of Artificial Intelligence* (Dec. 19, 2023), https://www.nist.gov/news-events/news/2023/12/nist-calls-information-support-safe-secure-and-trustworthy-development-and.

411 李郁強，人工智慧醫療器材之法制初探，2022年10月17日，https://www.ly.gov.tw/Pages/Detail.aspx?nodeid=6590&pid=223217（最後瀏覽日：2023年9月30日）。

412 *Non-Governmental AI Standards and EU Policy*, REGULAITE, https://www.regulaite.eu/v-private-standards-for-ai (last visited Oct. 18, 2023).

413 Artificial Intelligence Act recital 121.

414 例如日本積極參與OECD於AI議題之討論。詳見第壹部分有關日本之介紹。

415 *See generally* Cosmina Crăciunescu Timofte, *The Role of Artificial Intelligence in Collaborative Governance, Trust Building and Community Development Analysis*, *in* COLLABORATIVE GOVERNANCE, TRUST BUILDING AND COMMUNITY DEVELOPMENT 99 (Cristina Hinţea et al. eds., 2020).

與及溝通，亦有助於強化大眾對於AI使用之風險認知與信心[416]。亦有論者認為，公部門使用科技設備，更加需要建立與社會成員之間的信任關係，因為科技手段對人民基本權利干預可能性更高[417]。在我國就AI整體治理路線與部會權責劃分仍不夠明確之情況下，本文建議我國可在前述軟法與硬法混合的管制取徑上，斟酌考量歐盟《人工智慧法》相關設計（利害關係人得參與諮詢論壇以提供意見）[418]，指定專責單位強化利害關係人參與及跨部會溝通，相關作為包括在制定AI相關政策時積極徵詢公民團體與公眾意見、鼓勵外界主動提供政府AI管制作為相關資訊，以及拓展公眾參與之管道與制度多元性等，俾使AI軟硬法制定過程納入多元意見[419]，蓋AI政策不僅是產業政策，更會對於所有公民的日常生活產生重要影響，長期而言更需要多元教育與資訊流通以培力（empower）各方利害關係人，特別是對AI議題的識讀能力，方能促進有效的參與及民主

416 例如有英國智庫發現，即使民眾未必會閱讀政府使用AI情形之資訊，但此類資訊的存在本身可能就能增進公部門使用演算法的透明度，進而增進公眾信心，且這些資訊的存在也讓專家能代表民眾監督政府。BritainThinks, *Complete Transparency, Complete Simplicity: How Can the Public Sector Be Meaningfully Transparent about Algorithmic Decision Making?*, GOV.UK 45 (June 17, 2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/995014/Complete_transparency__complete_simplicity_-_Accessible.pdf.

417 翁逸泓（註37），頁251-252。作者亦建議可參照歐盟AI HLEG發布之《可信任AI之倫理準則》，對於AI應用建立信任之架構。參見翁逸泓（註37），頁305-306。

418 Artificial Intelligence Act art. 67; Artificial Intelligence Act recital 150. 除了利害關係人參與外，私部門的參與亦值得注意，在AI的發展與法規制定上私部門皆具備不小的影響力，以《人工智慧法》的私部門參與情況為例，即有學者指出《人工智慧法》的規範明顯聚焦於公部門管制，忽視對私部門的管制。Ruschemeier & Bareis, *supra* note 253, at 10。法規制定上，私部門的遊說亦很大程度反映於《人工智慧法》的討論與最終條文，削弱《人工智慧法》的實質管制力道。*See* Jascha Bareis, *BigTech's Efforts to Derail the AI Act*, VERFBLOG (Dec. 5, 2023), https://verfassungsblog.de/bigtechs-efforts-to-derail-the-ai-act/; *see also* Alan Dignam, *Artificial Intelligence, Tech Corporate Governance and the Public Interest Regulatory Response*, 13 CAMBRIDGE J. REGIONS ECON. & SOC'Y 37, 44-49 (2020).

419 Aaronson & Zable, *supra* note 406, at 6-7.

審議。相關部會亦應積極揭露參與AI政策之制定者組成[420]，使外界能檢驗政策制定之代表性與正當性；在徵詢公眾意見時，也應於事後說明其對公眾意見之個別回應與採納程度[421]，確保利害關係人參與之機制並非流於形式，而是確實有助於政策制定、確實反映我國產業發展與監管需求之AI法制。

伍、結論

在全球AI科技發展、產業標準與規範管制均高度競爭的時代，歐盟《人工智慧法》為目前針對AI風險與挑戰的最新立法，在2024年8月正式通過後已成為全球第一部全面性AI專法。在制度設計上，《人工智慧法》有諸多創新管制取徑與規範內涵，本文將主軸集中於近年來討論AI所帶來的管制挑戰與治理應對中最受重視者，亦即風險分級體系、人類監督與透明度等三大面向，而其另分別牽動其他細節性法律規定。首先，風險區分之制度設計衍生出不同之對應管制措施，包含對存在不可接受的風險之AI系統全面禁用、對非高風險AI系統則原則上僅要求企業自律，而針對高風險AI系統，則賦予相關行為者各項義務以掌握系統生命週期所有可能的風險。此外，為有效掌握潛在的風險，《人工智慧法》亦將通用目的AI模型與通用目的AI系統納入管制範圍，並設立AI Office以調和歐盟會員國間於AI之管制、建立獨立專家科學小組以應對通用目的AI系統可能衍生的風險，以及強化監理沙盒鼓勵創新與輔助法遵之功能。又如人類監督，在GDPR的基礎上，要求涉及AI系統用於輔助或取代人類決策之應用場域強化人類監督機制，更將歐盟《人工智慧法》適用範圍具體延伸到政府公權力執行乃至於數位平臺監管，不

420 *Id.*
421 *Id.*

僅要人類參與AI系統的決策中，更要求其「有意義的參與」、能有權利變更系統決定。而透明度之要求，更為各國政策、原則，或是國際組織採納的AI準則等，均視為重要治理原則者，在《人工智慧法》中，則是要求相關人等積極揭露與系統有關之資訊，並確保使用者知悉其正與AI系統互動。本文認為此三大面向為歐盟《人工智慧法》最重要的制度設計特色，更是其他具體細節規定所依附之規範主軸。另一個平行於此三大面向的重要制度設計，則是監理沙盒，雖平行於其他管制措施，但監理沙盒將直接影響歐盟會員國在地管制成效的重要面向。如歐盟對於監理沙盒若欠缺適當調和，可能在各會員國間產生管制碎裂化問題，造成管制逐底競爭或企業挑選管轄區域等問題，影響歐盟AI治理的一致性與有效性。

如本文所論，歐盟《人工智慧法》部分制度與條文雖可能存有爭議以及治理侷限，如風險分級的認定疑義、AI系統透明度相關要求之規範罅隙、與人類監督機制是否足夠、監理沙盒可能衍生之管制碎裂化等問題，有待未來進一步調整與改善。然其作為全球第一部針對AI的全面性專法，除直接對歐盟會員國發生效力外，歐盟法的布魯塞爾效應，更將透過跨國界市場驅動力、大型企業規範擴散效果，以及對各國管制措施之間接影響力，甚至形塑國際組織對於AI管制、標準、政策倡議與技術發展之方向，而我國勢必也將受到不同程度之影響。

然而相對於歐盟完成最終立法，我國目前對於AI法制與治理之回應腳步似乎過於躊躇，迄今仍未有針對AI之全面性規範，雖各部會陸續就AI之使用制定相關軟法指引，以及目前的《人工智慧基本法》草案，或暫時能緩解若干管制壓力與挑戰，然我國對系統性與具體AI問題之管制需求亟待重視。本文建議，考量AI發展與應用迅速且複雜多變，我國並非必須立即制定全面性專法，或可參酌採納軟法與硬法混合之方式，透過行為準則、倫理原則、標準化、或技術指引等方式鼓勵相關行為者遵循一定的規範要求，以避免不必要

的管制障礙、鼓勵產業創新，並經由踐行、累積經驗與資訊，由下而上建構必要的硬法管制。另一方面，為維護基本權利、法治原則、民主制度等重要價值，特定AI發展與應用情狀，亦有以硬法管制的急迫性，因此我國宜積極考量立法禁止對公民社會有不可接受的風險之AI系統或應用，如基於執法目的而於公開場所使用即時遠端生物辨識系統等，或至少訂定嚴格的使用條件與許可制度，並要求公部門盤點並揭露運用現況，作為未來強化AI治理的基礎。綜上，未來我國欲採取何種管制取徑，除應於適當時參考國際管制趨勢外，更應系統性審酌AI快速發展與應用下的基本人權保障、利害關係人之參與、相關規範與標準制定及其與既有法規框架之調和、乃至於關鍵基礎科技及國家安全，以及整體AI產業的創新與發展，借力軟法與硬法取徑的不同功能，擬定一適合我國實務需求、民主法治制度、確保基本權利與鼓勵創新發展之AI規範。

參考文獻

1. 中文部分

李建良（2022），智慧駕駛的法學思辨——自動駕駛法學的試車之旅，月旦法學雜誌，327期，頁94-114。

李惠宗（2023），健保資料庫的合理利用與個資保護——從歐盟一般資料保護規則的「資料治理」探討憲法法庭111年憲判字第13號判決未解決問題，台灣法律人，24期，頁1-25。

李榮耕（2022），刑事程序中人工智慧於風險評估上的應用，政大法學評論，168期，頁117-186。

李震山（2002），行政組織法與行政作用法之區別及其實益，月旦法學雜誌，85期，頁16-17。

林昱梅（2022），預防原則於人工智慧之適用——歐盟人工智慧規則草案及其基本原理之觀察，台灣法律人，12期，頁1-17。

林勤富（2022），智慧法院之發展與界限（下）——演算法、科技治理與司法韌性，月旦法學雜誌，324期，頁109-131。

———（2024），人工智慧自動決策系統的透明度難題，收於：李建良編，人工智慧的法制基礎：破壞式創新的建構式法制，頁131-183，臺北：元照。

林勤富、李怡俐（2020），人工智慧時代下的國際人權法——規範與制度的韌性探索與再建構，收於：李建良編，法律思維與制度的智慧轉型，頁416-468，臺北：元照。

翁逸泓（2022），社群媒體不實訊息之治理——以個資保護模式為選項，中研院法學期刊，30期，頁171-230。

———（2022），面部自動辨識監控系統與個資保護問題研究——Bridges案之啟示，政大法學評論，168期，頁249-324。

財團法人台灣經濟研究院（2022），2022產業技術白皮書，臺北：

經濟部。
郭戎晉（2023），國際趨勢下之人工智慧監管可能模式與臺灣推動課題，全國律師月刊，27卷6期，頁18-36。
張容涵（2019），從比較法觀點探討我國個資保護制度之轉型，國立臺灣大學法律學研究所碩士論文。
張麗卿（2020），AI倫理準則及其對臺灣法制的影響，月旦法學雜誌，301期，頁97-117。
張麗卿、周伯翰、林勤富、吳振吉、王紀軒、洪兆承、陳俊榕、韓政道、朱宸佐（2023），臺灣人工智慧基本法制定之必要與倡議，月旦法學雜誌，340期，頁79-101。
陳弘儒（2019），論法律人工智慧系統與法治理念的緊張關係——初探技術開發的道德許可性，台灣法理學會，https://ivrtaiwan.wordpress.com/2019ailaw/。
陳陽升（2023），從法治原則探索人工智慧之應用界限，台灣法律人，26期，頁97-106。
黃銘傑（2023），徘徊於事後管制v.事前管制之數位時代中的競爭法主管機關，台灣經濟研究月刊，46卷7期，頁95-101。
劉靜怡（2019），淺談GDPR的國際衝擊及其可能因應之道，月旦法學雜誌，286期，頁5-31。
———（2020），人工智慧時代的法學研究路徑初探，收於：李建良編，法律思維與制度的智慧轉型，頁91-134，臺北：元照。
薛智仁（2023），初探人工智慧對刑法學的挑戰——以自動駕駛為例，台灣法律人，27期，頁104-134。

2. 外文部分

Aaronson, Susan Ariel, and Adam Zable. 2023. Missing Persons: The Case of National AI Strategies. CIGI Papers No. 283. Available at https://www.cigionline.org/publications/missing-persons-the-case-of-national-ai-strategies/.

Abbott, Kenneth W. 2013. Introduction: The Challenges of Oversight for Emerging Technologies. Pp. 1-16 in *Innovative Governance Models for Emerging Technologies*, edited by Gary E. Marchant, Kenneth W. Abbott and Braden Allenby. Northampton, MA: Edward Elgar Publishing.

Ananny, Mike, and Kate Crawford. 2018. Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. *New Media and Society* 20:973-989.

Bianchi, Andrea. 2013. On Power and Illusion: The Concept of Transparency in International Law. Pp. 1-20 in *Transparency in International Law*, edited by Andrea Bianchi and Anne Peters. Cambridge: Cambridge University Press.

Black, Julia, and Andrew Douglas Murray. 2019. Regulating AI and Machine Learning: Setting the Regulatory Agenda. *European Journal of Law and Technology* 10(3):1-21.

Bradford, Anu. 2012. The Brussels Effect. *Northwestern University Law Review* 107:1-68.

———. 2020. *The Brussels Effect: How the European Union Rules the World.* New York, NY: Oxford University Press.

———. 2023. The race to regulate artificial intelligence: Why Europe has an edge over America and China. In *Foreign Affairs*, https://www.foreignaffairs.com/united-states/race-regulate-artificial-intelligence?check_logged_in=1.

———. 2023. Whose AI revolution?. In *Project Syndicate*, https://www.project-syndicate.org/onpoint/ai-regulation-us-eu-china-challenges-opportunities-by-anu-bradford-2023-09.

BritainThinks. 2021. Complete Transparency, Complete Simplicity: How Can the Public Sector Be Meaningfully Transparent about Algorithmic Decision Making?. Available at https://assets.

publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/995014/Complete_transparency__complete_simplicity_-_Accessible.pdf.

Brkan, Maja, and Grégory Bonnet. 2020. Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas. *European Journal of Risk Regulation* 11:18-50.

Brownsword, Roger, and Karen Yeung, eds. 2008. *Regulating Technologies: Legal Futures, Regulatory Frames and Technological Fixes.* Portland, OR: Hart Publishing.

Buenfil, Jorge, Ross Arnold, Benjamin Abruzzo, and Christopher Korpela. 2019. Artificial Intelligence Ethics: Governance through Social Media. *2019 IEEE International Symposium on Technologies for Homeland Security*. Available at http://www.hstabstractbook.org/index_htm_files/c-IEEE_HST___AI_Ethics__Governance_through_Social_Media%20FINAL.pdf.

Buijze, Anoeska. 2013. The Six Faces of Transparency. *Utrecht Law Review* 9(3):3-25.

Burrell, Jenna. 2016. How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data and Society* 3(1):1-12.

Casey, Bryan, Ashkon Farhangi, and Roland Vogl. 2019. Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise. *Berkeley Technology Law Journal* 34:143-188.

Chander, Anupam. 2017. The Racist Algorithm?. *Michigan Law Review* 115:1023-1045.

Citron, Danielle K., and Frank Pasquale. 2014. The Scored Society: Due Process for Automated Predictions. *Washington Law Review* 89:1-33.

Clermont, Kevin M., and Theodore Eisenberg. 1995. Exorcising the Evil of Forum-Shopping. *Cornell Law Review* 80:1507-1535.

Coglianese, Cary, and David Lehr. 2019. Transparency and Algorithmic Governance. *Administrative Law Review* 71:1-56.

Corrêa, Nicholas Kluge. 2020. Blind Spots in AI Ethics and Biases in AI Governance. Available at https://philarchive.org/archive/CORASv3.

Crăciunescu Timofte, Cosmina. 2020. The Role of Artificial Intelligence in Collaborative Governance, Trust Building and Community Development Analysis. Pp. 99-110 in *Collaborative Governance, Trust Building and Community Development*, edited by Cristina Hinţea, Bianca Radu and Raluca Suciu. Cluj-Napoca: Babeş-Bolyai University.

Datasphere Initiative. 2022. Sandboxes for Data: Creating Spaces for Agile Solutions Across Borders. Available at https://www.thedatasphere.org/wp-content/uploads/2022/05/Sandboxes-for-data-2022-Datasphere-Initiative.pdf.

Department for Science, Innovation and Technology. 2023. A Pro-Innovation Approach to AI Regulation. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1176103/a-pro-innovation-approach-to-ai-regulation-amended-web-ready.pdf.

Dignam, Alan. 2020. Artificial Intelligence, Tech Corporate Governance and the Public Interest Regulatory Response. *Cambridge Journal of Regions, Economy and Society* 13:37-54.

Fenster, Mark. 2006. The Opacity of Transparency. *Iowa Law Review* 91:885-949.

Fernández, José Vida. 2023. Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State. *Indiana Journal of Global Legal Studies* 30:65-95.

Folberth, Anja, Jutta Jahnel, Jascha Bareis, Carsten Orwat, and Christian Wadephul. 2022. Tackling Problems, Harvesting Benefits - A Systematic Review of the Regulatory Debate Around AI. KIT Scientific Working Papers No. 197. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4212447.

Germany Federal Ministry for Economic Affairs and Energy. 2019. Making Space for Innovation: The Handbook for Regulatory Sandboxes. In *German Federal Ministry for Economic Affairs and Climate Action*, https://www.bmwk.de/Redaktion/EN/Publikationen/Digitale-Welt/handbook-regulatory-sandboxes.pdf?__blob=publicationFile&v=1.

Gilman, Michele E. 2020. Five Privacy Principles (from the GDPR) the United States Should Adopt to Advance Economic Justice. *Arizona State Law Journal* 52:368-444.

Government Communications Headquarters. 2021. Pioneering a New National Security: The Ethics of Artificial Intelligence. Available at https://www.gchq.gov.uk/files/GCHQAIPaper.pdf.

De Gregorio, Giovanni. 2021. The Rise of Digital Constitutionalism in the European Union. *International Journal of Constitutional Law* 19:41-70.

Grochowski, Mateusz, Agnieszka Jabłonowska, Francesca Lagioia, and Giovanni Sartor. 2021. Algorithmic Transparency and Explainability for EU Consumer Protection: Unwrapping the Regulatory Premises. *Critical Analysis of Law* 8:43-63.

Hacker, Philipp. 2023. The European AI Liability Directives—Critique of a Half-Hearted Approach and Lessons for the Future. Working Paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4279796.

Hemrajani, Asha. 2023. China's new legislation on deepfakes: Should

the rest of asia follow suit?. In *The Diplomat*, https://thediplomat.com/2023/03/chinas-new-legislation-on-deepfakes-should-the-rest-of-asia-follow-suit/.

Hillman, John. 2023. Smart Regulation: Lessons from the Artificial Intelligence Act. *Emory International Law Review* 37:775-826.

Houston, Joel F., Chen Lin, and Yue Ma. 2012. Regulatory Arbitrage and International Bank Flows. *Journal of Finance* 67:1845-1895.

Humerick, Matthew. 2018. Taking AI Personally: How the E.U. Must Learn to Balance the Interests of Personal Data Privacy & Artificial Intelligence. *Santa Clara High Technology Law Journal* 34:393-418.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1: 389-399.

Kingsman, Nigel, Emre Kazim, Ali Chaudhry, Airlie Hilliard, Adriano Koshiyama, Roseline Polle, Giles Pavey, and Umar Mohammed. 2022. Public Sector AI Transparency Standard: UK Government Seeks to Lead by Example. *Discover Artificial Intelligence* 2(1)2:1-9.

Kwan, Stephen K., and Jim Spohrer. 2021. Reducing Industry Complexity with International Standards: Current Efforts for Services, E-commerce, Artificial Intelligence. Pp. 67-76 in *Advances in the Human Side of Service Engineering*, edited by Christine Leitner, Walter Ganz, Debra Satterfield and Clara Bassano. Cham: Springer.

Lazcoz, Guillermo, and Paul de Hert. 2023. Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Essential Pre-Requisites against Abdicating Responsibilities. *Computer Law and Security Review* 50(105833):1-20.

Li, Wenlong, and Jill Toh. 2023. Data Subject Rights as a Tool for Platform Worker Resistance: Lessons from the Uber/Ola Judgments.

Pp. 119-156 in *Data Protection and Privacy, Volume 15: In Transitional Times*, edited by Hideyuki Matsumi, Dara Hallinan, Diana Dimitrova, Eleni Kosta and Paul de Hert. London: Bloomsbury Publishing.

Lin, Ching-Fu. 2023. Sorting Teachers Out: Automated Performance Scoring and the Limit of Algorithmic Governance in the Education Sector. Pp. 189-204 in *Money, Power, and AI: Automated Banks and Automated States*, edited by Zofia Bednarz and Monika Zalnieriute. Cambridge: Cambridge University Press.

Liu, Han-Wei, Ching-Fu Lin, and Yu-Jie Chen. 2019. Beyond *State v. Loomis*: Artificial Intelligence, Government Algorithmization, and Accountability. *International Journal of Law and Information Technology* 27:122-141.

Lu, Sylvia. 2020. Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence. *Vanderbilt Journal of Entertainment and Technology Law* 23:99-159.

———. 2022. Data Privacy, Human Rights, and Algorithmic Opacity. *California Law Review* 110:2087-2147.

Madiega, Tambiama. 2024. Artificial Intelligence Act. European Parliamentary Research Service, PE 698.792. Available at https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

Madiega, Tambiama, and Anne Louise Van De Pol. 2022. Artificial Intelligence Act and Regulatory Sandboxes. European Parliamentary Research Service, PE 733.544. Available at https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf.

Mandel, Gregory N. 2009. Regulating Emerging Technologies. *Law,*

Innovation and Technology 1:75-92.

Manheim, Karl, and Lyric Kaplan. 2019. Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law and Technology* 21:106-188.

Marchant, Gary E., and Carlos Ignacio Gutierrez. 2020. Indirect Enforcement of Artificial Intelligence 'Soft Law.' *TPRC48: The 48th Research Conference on Communication, Information and Internet Policy.* Available at https://ssrn.com/abstract=3749776.

McNair, Douglas, and W. Nicholson Price II. 2022. Health Care Artificial Intelligence: Law, Regulation, and Policy. Pp. 197-234 in *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*, edited by Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed and Danielle Whicher. Whashington, DC: National Academy of Medicine.

Mendoza, Isak, and Lee A. Bygrave. 2017. The Right Not to Be Subject to Automated Decisions Based on Profiling. Pp. 77-97 in *EU Internet Law: Regulation and Enforcement*, edited by Tatiana-Eleni Synodinou, Philippe Jougleux, Christiana Markou and Thalia Prastitou. Cham: Springer.

Mittelstadt, Brent. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence* 1:501-507.

Monterossi, Michael W. 2019. Algorithmic Decisions and Transparency: Designing Remedies in View of the Principle of Accountability. *The Italian Law Journal* 5:711-730.

De Nederlandsche Bank, and Dutch Authority for the Financial Markets. 2019. Continuing Dialogue—Innovation Hub and Regulatory Sandbox: Lessons Learned after Three Years. Available at https://www.afm.nl/~/profmedia/files/rapporten/2019/innovationhub-rapport-eng.pdf.

Norwegian Ministry of Local Government and Modernisation. 2020. National Strategy for Artificial Intelligence. Available at https://www.regjeringen.no/contentassets/1febbbb2c4fd4b7d92c67ddd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf.

Organisation for Economic Cooperation and Development. 2023. Regulatory Sandboxes in Artificial Intelligence. OECD Digital Economic Papers No. 356. In *OECD*, https://doi.org/10.1787/8f80a0e6-en.

Oswald, Marion, Luke Chambers, Ellen P. Goodman, Pam Ugwudike, and Miri Zilka. 2022. The UK Algorithmic Transparency Standard: A Qualitative Analysis of Police Perspectives. Available at https://ssrn.com/abstract=4155549.

Pagallo, Ugo. Forthcoming. Why the AI Act Won't Trigger a Brussels Effect. In *AI Approaches to the Complexity of Legal Systems*. Berlin: Springer.

Paige, Mark A., and Audrey Amrein-Beardsley. 2020. "Houston, We Have a Lawsuit": A Cautionary Tale for the Implementation of Value-Added Models for High-Stakes Employment Decisions. *Educational Researcher* 49:350-359.

Partnoy, Frank. 1997. Financial Derivatives and the Costs of Regulatory Arbitrage. *The Journal of Corporation Law* 22:211-256.

Pasquale, Frank. 2016. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Cambridge, MA: Harvard University Press.

Plesner, Louis L., Felix C. Müller, Janus D. Nybing, Lene C. Laustrup, Finn Rasmussen, Olav W. Nielsen, Mikael Boesen, and Michael B. Andersen. 2023. Autonomous Chest Radiograph Reporting Using AI: Estimation of Clinical Impact. *Radiology* 307(3)e222268:1-10.

Riles, Annelise. 2014. Managing Regulatory Arbitrage: A Conflict of

Laws Approach. *Cornell International Law Journal* 47:63-119.

Ruschemeier, Hannah, and Jascha Bareis. 2024. Searching for Harmonised Rules: Understanding the Paradigms, Provisions and Pressing Issues in the Final EU AI Act. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4876206.

Schwab, Klaus. 2016. *The Fourth Industrial Revolution*. New York, NY: World Economic Forum.

Shen, Weixing, and Yun Liu. 2022. China's Normative Systems for Responsible AI: From Soft Law to Hard Law. Pp. 150-166 in *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives*, edited by Silja Voeneky, Philipp Kellmeyer, Oliver Mueller and Wolfram Burgard. Cambridge: Cambridge University Press.

Smuha, Nathalie A. 2021. From a "Race to AI" to a "Race to AI Regulation": Regulatory Competition for Artificial Intelligence. *Law, Innovation and Technology* 13:57-84.

Solove, Daniel J. 2001. Privacy and Power: Computer Databases and Metaphors for Information Privacy. *Stanford Law Review* 53:1393-1462.

———. 2024. Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data. *Northwestern University Law Review* 118:1081-1138.

Solow-Niederman, Alicia. 2022. Information Privacy and the Inference Economy. *Northwestern University Law Review* 117:357-423.

Storm, Marijn and Alex van der Wolk. 2021. Privacy and the EU's Regulation on AI: What's New and What's Not?. *The Journal of Robotics, Artificial Intelligence & Law* 4:451-458.

Stoycheff, Elizabeth. 2016. Under Surveillance: Examining Facebook's Spiral of Silence Effects in the Wake of NSA Internet Monitoring.

Journalism and Mass Communication Quarterly 93:296-311.

Truby, Jon, Rafael Dean Brown, Imad Antoine Ibrahim, and Oriol Caudevilla Parellada. 2022. A Sandbox Approach to Regulating High-Risk Artificial Intelligence Applications. *European Journal of Risk Regulation* 13:270-294.

Tsai, Chang-Hsien, Ching-Fu Lin, and Han-Wei Liu. 2020. The Diffusion of the Sandbox Approach to Disruptive Innovation and Its Limitations. *Cornell International Law Journal* 53:261-296.

Vale, Sebastião Barros, and Gabriela Zanfir-Fortuna. 2022. Automated Decision-Making Under the GDPR: Practical Cases from Courts and Data Protection Authorities. *Future of Privacy Forum*. Available at https://fpf.org/wp-content/uploads/2022/05/FPF-ADM-Report-R2-singles.pdf.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law* 7:76-99.

Walz, Axel, and Kay Firth-Butterfield. 2019. Implementing Ethics into Artificial Intelligence: A Contribution, from a Legal Perspective, to the Development of an AI Governance Regime. *Duke Law and Technology Review* 18:176-231.

Whytock, Christopher A. 2011. The Evolving Forum Shopping System. *Cornell Law Review* 96:481-534.

Wulf, Alexander J., and Ognyan Seizov. 2024. "Please Understand We Cannot Provide Further Information": Evaluating Content and Transparency of GDPR-Mandated AI Disclosures. *AI & Society* 39:235-256.

Yeung, Karen. 2018. Algorithmic Regulation: A Critical Interrogation. *Regulation and Governance* 12:505-523.

Zalnieriute, Monika, Lyria Bennett Moses, and George Williams. 2019. The Rule of Law and Automation of Government Decision-Making. *Modern Law Review* 82:425-455.

Zetzsche, Dirk A., Ross P. Buckley, Janos N. Barberis, and Douglas W. Arner. 2017. Regulating a Revolution: From Regulatory Sandboxes to Smart Regulation. *Fordham Journal of Corporate and Financial Law* 23:31-103.

Institutional Design, Normative Substance, and Governance Limits of the EU Artificial Intelligence Act

*Ching-Fu Lin**

Abstract

The European Union Artificial Intelligence Act, as the world's first comprehensive legislation focusing on artificial intelligence regulation, poses significant implications and reference power in terms of its institutional design and substantive provisions. The Artificial Intelligence Act may also extend beyond the EU borders, having a direct or indirect impact on other countries or organizations regarding their regulation and governance, standards-setting, and technological development of artificial intelligence. This warrants an in-depth exploration of its institutional framework, normative substance, and potential governance limitations. This article reviews the various challenges posed by artificial intelligence, along with the trajectories of regulatory responses of countries such as the United States, the United Kingdom, Japan, China, and South Korea. Building on this foundation, this Article analyzes the structure and key provisions of the Artificial Intelligence Act—in particular those related to risk-based approach, human oversight, transparency, and regulatory sandboxes—and examines possible normative ramifications and problems. The article concludes by drawing a comparative lens and shedding light on Taiwan's AI governance, identifying potential issues and avenues for refinement, so as to inform policymakers of global regulatory trends including the

* Professor and Director, Institute of Law for Science and Technology, National Tsing Hua University.

EU regulatory model and in turn, reinforce Taiwan's AI governance.

KEYWORDS: artificial intelligence (AI), European Union, AI regulation, risk-based approach, transparency, accountability.