doi: 10.29843/JCCIS.202501_(48).0005

人工智慧全面監管時代來臨？—剖析歐盟人工智慧法之光與影

郭戎晉

摘要

全球正處於應否針對人工智慧進行監管以及合適作法的討論浪潮，在風險漸增下，人工智慧的監管思維已由初期的自律為主，逐步向他律靠攏並有轉為以他律為重之趨勢，而 2024 年 8 月正式生效的歐盟「人工智慧法」（AIA），成為人工智慧全面監管的里程碑式立法。歐盟甚早便確定推動監管專法，立於「風險管制」基準將人工智慧系統應用上可能衍生的風險，劃分為：1、無法接受的風險；2、高度風險；3、有限風險；及 4、最小風險或無風險等四個級別，並按風險級別的高低設定其規範密度，同時在法案研議後期，加入現時備受關注的生成式人工智慧／通用人工智慧模型之規範。受歐盟影響，現時已有若干國家刻正推動相近立法。然而全面性監管專法是否為人工智慧之治理良器，並不乏爭論，歐盟「人工智慧法」本身難以褪去受詬病之處，除可能導致「監管失調」情形，亦可能衍生料想之外的負面作用。此外，執政者所採取的監管舉措極可能落入科林格里奇困境，亦未必可憑之解決人工智慧實務應用衍生的所有問題，使得當前仍有國家對人工智慧之全面監管採取保留態度。儘管現時國際上存在著多樣化的人工智慧治理作法，但不同方法之間已可窺見共通之處，無論是受歐盟人工智慧法所產生的布魯塞爾效應之影響、選擇制定全面性專法，抑或擬保持監管彈性而採取軟法機制，異中求同應是人工智慧治理推動上可預見之必然走向。

⊙ 關鍵字：人工智慧、歐盟、人工智慧法、風險管制基準、人工智慧系統、通用人工智慧模型、布魯塞爾效應、監管失調

⊙ 本文作者郭戎晉為南臺科技大學財經法律研究所副教授。

⊙ 聯絡方式：Email ：ronkuo@stust.edu.tw；通訊處：71005 台南市永康區南台街1號。

⊙ 收稿日期：2024/11/05　接受日期：2025/01/24

Is the Era of Comprehensive Regulation of Artificial Intelligence Coming?—Analysis of Light and Shadow of the EU's Artificial Intelligence Act

Jung-Chin Kuo

Abstract

The world is in a wave of discussion on whether Artificial Intelligence (AI) should be regulated, and the appropriate regulatory approach for AI. Under the increasing risk, the regulatory thinking corresponding to AI has gradually changed from the initial soft law mechanism to the hard law mechanism. The European Union's Artificial Intelligence Act (AIA), which came into force in August 2024, has become a landmark legislation for the comprehensive regulation of AI. According to the "risk-based approach", the AIA classifies the risks that may arise from the application of AI Systems into: 1. Unacceptable risks; 2. High risk; 3. Limited risk; and 4, Minimal risk or no risk four levels, and according to the level of risk to set its subject to the norms. The European Union also added regulations on General-purpose Artificial Intelligence in the later stages of the AIA discussion. Influenced by the European Union, several countries are now promoting similar legislation. However, there is debate about whether comprehensive regulatory legislation is a good governance tool for AI. The AIA itself has been criticized for several things, including the possibility of regulatory misalignment and the possibility of negative effects that could not be foreseen in advance. In addition, regulatory legislation toward AI is also likely to fall into the so-called "Collingridge Dilemma", and may not be able to solve all the problems derived from the practical application of AI, so that there are still countries to adopt a reservation about the comprehensive regulation of AI. Although there are a variety of AI governance practices in the world, there are some common points between different approaches, and seeking

common ground in different approaches should be the inevitable trend of AI governance promotion.

⊙ Keywords: Artificial Intelligence, European Union, Artificial Intelligence Act, Risk-based Approach, AI System, General-purpose Artificial Intelligence Model, Brussels Effect, Regulatory Misalignment

⊙ The author, Jung-Chin Kuo, is an Associate Professor in the Institute of Financial & Economic Law at Southern Taiwan University of Science and Technology.

⊙ Corresponding author: Jung-Chin Kuo, email: ronkuo@stust.edu.tw; address: No.1, Nantai St., Yungkang Dist., Tainan City 710301, TAIWAN R.O.C.

⊙ Received: 2024/11/05 Accepted: 2025/01/24

壹、前言

人工智慧（Artificial Intelligence, AI）此一詞彙首見於 1956 年，用以表示所有「非人類（機器）所表現的智慧」。千禧年後機器學習（Machine Learning）此一人工智慧的子技術取得了突破性發展，從而提高了機器根據歷史資料進行預測之能力（OECD, 2019a）；而 2022 年以降包括大型語言模型（Large Language Models）在內的生成式人工智慧（Generative AI）之問世，對許多人而言人工智慧無疑地變得更為真實（OECD, 2024）。

人們持續驚艷於人工智慧的發展潛力，並將之視為改變世界面貌的變革工具，如果運用得當，人工智慧可望開啟一個繁華的新時代；對人工智慧抱持懷疑態度者，則認為人工智慧或如同未為人知的外星生命，應當謹慎地面對並加以限制，以防止人工智慧接管人類甚或扼殺人類（Roose, 2023）。

隨著人工智慧應用領域的拓展與多元化，此一嶄新資通訊技術引發的問題也隨之劇增。面對人工智慧發展伴隨而生的各類爭議之處理，「自律」與「他律」機制各有擁護者，儘管產業自律概念立意良善且對科技發展相對友善，在人工智慧衍生爭端急遽增加且難以援引既有的科技監管經驗之下，國際應對人工智慧的治理思維也開始出現變化。觀察近年發展，國際上針對人工智慧的監管，除了由全然自律開始加入他律，近年更有進一步向他律靠攏、甚至轉以他律為主之趨勢（郭戎晉，2023a，頁 115）。

歐盟無疑是人工智慧「他律」機制的旗手，並率先制定通過全球首部人工智慧全面性監管專法，2024 年 8 月 2 日正式生效的「人工智慧規則」（Artificial Intelligence Act；以下簡稱 AIA），也成為此一發展趨勢下最具代表性之立法，當前並有國家如加拿大及巴西等，亦刻正仿歐盟著手推動其人工智慧監管專法。當前許多國家仍在激烈爭辯人工智慧發展帶來正反效應，各國監管機關也努力應對此一嶄新技術所帶來的挑戰以及如何實現監管平衡。本文以歐盟甫生效的 AIA 為核心，探討人工智慧監管立法可能帶來的正反效益，並就臺灣現況進行檢視及建言，以期作為我國在關聯法制構思與設計上之參考。

貳、人工智慧之概念界定與所生影響

一、人工智慧／人工智慧系統之定義

（一）世界智慧財產權組織

「人工智慧」一詞就其字面意義而言，泛指「非人類（機器）所表現的智慧」，惟此一簡要定義易於使人工智慧被誤解為單一或特定技術，但人工智慧其實是眾多技術的結合運用（郭戎晉，2020，頁 180）。就技術層面而言，世界智慧財產權組織（World Intellectual Property Organization, WIPO）提出的「人工智慧技術趨勢報告」，便明確表示人工智慧並非單一技術概念（WIPO, 2019）。

根據 WIPO 的說明，人工智慧涵括多個細部技術概念，主要包括：1、機器學習（Machine Learning）；2、邏輯程式設計（Logic programming）；3、模糊邏輯（Fuzzy Logic）；4、概率推理（Probabilistic Reasoning）； 5、本體工程（Ontology Engineering）；以及 6、功能應用（Functional Application）關聯技術。而最後的功能應用關聯技術，又可再細分為：1、電腦視覺（Computer Vision）：2、自然語言處理；3、語音處理（Speech Processing）及 4、其他功能應用等子概念（WIPO, 2019）。

（二）經濟合作暨發展組織

除上述 WIPO 技術面的定義，經濟合作暨發展組織（Organisation for Economic Cooperation and Development, OECD）在 2019 年發布的「人工智慧建議書」中，針對「人工智慧系統」（AI System）所作定義亦受到廣泛引用。依 OECD 定義，人工智慧系統係指「一種基於機器的系統，其可以針對由人類所界定的目標，作成影響真實或虛擬環境之預測、建議或決策。人工智慧系統可被設計為具有不同程度之自主性」（OECD, 2019b）。

OECD 在 2023 年 11 月修正上揭建議書中有關「人工智慧系統」之定義，將之重新界定為「一種基於機器的系統，針對明確或隱含目標，根據所接獲的輸入推斷如何產製可能影響真實或虛擬環境之預測、內容、建議或決策等輸出。不同的人工智慧系

統實際部署後之自主性與適應性程度各不相同」。自發布以降，OECD 人工智慧建議書受到諸多國家採納，OECD 將之稱為「跟隨者」（Adherents），除歐盟外，包括了 38 個 OECD 成員國及 8 個非成員國。

（三）歐盟

OECD 提出的「人工智慧系統」概念與定義受到廣泛引用，歐盟於研訂 AIA 時，也直接採納了「人工智慧系統」一詞。自歐盟執委員提出 AIA，截至 2023 年 6 月歐洲議會通過其談判立場版本，歐盟將人工智慧系統定義為「指基於機器之系統，其被設計為以不同程度自主運行，可得針對明確或隱含之目標，產製影響真實或虛擬環境之預測、建議或決策」，此一定義直接反映了 OECD 在 2019 年人工智慧建議書中針對「人工智慧系統」所作定義。

隨著 OECD 在 2023 年 11 月修正「人工智慧系統」定義，歐盟在 2024 年最終通過的 AIA，其人工智慧系統定義也調整為「指基於機器的系統，以不同程度的自主性進行操作並得於部署後展現適應性，針對明確或隱含的目標，根據所接獲的輸入推斷如何產製可能影響真實或虛擬環境之內容、建議或決策等輸出」。

歐盟針對「人工智慧系統」所作定義，受到 OECD 在內的國際組織文件之重大影響，歐盟 AIA 的定義原則上與 OECD「人工智慧建議書」保持一致，但文字呈現上則進一步納入法律實施與遵循之考量，從而未與 OECD 所作定義完全相同，以確保法律適用的明確性與廣泛接受（Freshfields, 2024）。

二、人工智慧實務應用衍生之課題

（一）民眾層面：滋生不確定風險並可能對人類造成傷害

人工智慧技術發展與實務應用可能衍生不確定風險，已成為各界共識，從而如何降低人工智慧可能肇致之危害，也成為國際組織及主要國家熱議之事。諸如歐盟與其他 27 個國家在 2023 年 11 月締結的「布萊切利宣言」（Bletchley Declaration），便表明人工智慧著實潛藏著各種不可預見風險，其中包括肇因虛假訊息（dis-information）與內容操控（manipulated content）所引發之相關風險，蓋人工智慧可能放大（amplify）錯誤及虛假訊息，從而加劇其所生危害。此外，人工智慧也可能造

成處於公開狀態的資訊受到污染，或破壞相關資料的可用性，亦即導致所謂的「資料中毒」（data poisoning）情形，損及資料品質並侵蝕公眾對公開資料之信任（Moss, 2024）。

布萊切利宣言同時強調人工智慧模型可能潛藏的安全風險，在追求前瞻應用之下，其嶄新功能可能導致重大甚至是災難性傷害。特別是人工智慧已愈發頻繁地被嵌入（embedding）「關鍵基礎設施」（critical infrastructure）之中，此舉固然有助於提高效率或降低成本，然而此一情形亦可能使得有心之人得以輕易地利用人工智慧造成物理或虛擬傷害（Zwetsloot & Dafoe, 2019）。

人工智慧衍生的偏見／歧視問題亦備受矚目，當人工智慧的訓練資料包含了偏見與其他有問題的資訊，便可能造成人工智慧在其輸出（outputs）中產生社會偏見（social prejudices）、刻板印象（stereotypes）與歧視等情形，進一步邊緣化（marginalise）或排除特定族群（Schwartz, Vassilev, Greene, Perine, Burt & Hall, 2022: OECD, 2024）。

隨著人們對人工智慧暨其衍生風險的認識加深，部分人工智慧風險亦相對顯而易見，七大工業國組織（G7）便將「錯誤與虛假訊息」、「智慧財產權侵害」與「隱私侵害」視為近期生成式人工智慧快速發展下之主要威脅（OECD, 2024）。

（二）產業層面：衍生限制競爭及不正當競爭等反競爭風險

人工智慧的「數位本質」（digital nature）創造了顯著的「先發者優勢」（first-mover advantages）環境，易於讓此一前瞻技術的早期採用者，拉開與其他事業之間的差距（Liu, Boy, Khurana & Sinha, 2024）。紐約時報在 2024 年 6 月便指出人工智慧持續發展可能造就市場上的獨占巨獸，特別是人工智慧仰賴大量的資料與運算能力，無疑地為佔據市場主導地位的事業帶來巨大優勢。考量人工智慧實務運作可能衍生的限制競爭風險，美國政府擬針對人工智慧技術巨擘，包括 Microsoft、OpenAI 及 Nvidia 等事業展開反壟斷（Antitrust）調查（McCabe, 2024）。

另一方面，有鑑於人工智慧易於形塑限制競爭，現時不乏人工智慧事業針對其所開發的人工智慧模型／系統，採取「開源」（open-source）形式加以提供；而開源有助於愈發快速的人工智慧創新與發展，從而緩解少數企業居市場主導地位所產生的「贏家通吃現象」（winner-take-all dynamics）（Dickson, 2023）。然而，採用開源形

式的人工智慧模型／系統事實上也潛在著若干重大風險，包括未有防制不當使用行為的保障措施，或相關保障措施效果極為薄弱（OECD, 2024）。

除了形成托拉斯（Trust）／限制競爭問題，人工智慧亦可能成為事業遂行其「不正當競爭」行為之工具。隨著人工智慧成為眾人注目的焦點，實務上開始出現所謂的「漂人工智慧」（AI Washing）情形，「漂人工智慧」此一概念衍生自「漂綠」（greenwashing）一詞，蓋隨著環境保護及永續發展成為風潮，事業可能不當誇大了其商品或服務的環保程度，藉吸引富有生態意識的消費者。在各界競相追逐人工智慧之下，事業亦可能誇大或誤導標榜作為「人工智慧」所出售的商品或服務之功能，但事實上相關商品或服務可能僅有少量的人工智慧，甚至完全不存在人工智慧（Marr, 2024）。以美國為例，包括聯邦交易委員會（Federal Trade Commission, FTC）及證券交易委員會（Securities and Exchange Commission, SEC）在內，均關注並警示「漂人工智慧」問題，後者並針對存有虛假與誤導性言論的事業採取執法行動（SEC, 2024）。

各行各業爭先恐後導入人工智慧之下，Kevin Roose（2023）指出人工智慧的創造者對此一嶄新技術既感好奇又憂懷恐懼，試圖通過群體治理限制它的力量。但在人工智慧無法免於「資本家」（capitalist）操控之下，一個不爭的事實便是企業利益的追求，最終仍將戰勝人們對於未來的擔憂。針對人工智慧在產業層面可能衍生的市場競爭問題，現階段主要國家持續思考人工智慧技術快速發展與實務應用推陳出新，對固有競爭法制與競爭執法所帶來的挑戰，特別是在生成式人工智慧掀起前所未見的熱潮之下，競爭層面的討論重心有自人工智慧系統，向生成式人工智慧/通用人工智慧模型靠攏之勢（FTC, 2023; Autorité de la concurrence, 2024）。

以法國競爭委員會在 2024 年 6 月發布的生成式人工智慧產業競爭報告為例，法國競爭委員會指出生成式人工智慧存在著高參進障礙特質，而當前主要業者在生成式人工智慧之相關市場上極可能形成競爭優勢。另值得留意者，法國競爭委員會指出後續將關注生成式人工智慧「模型即服務」（Model as a Service, MaaS）的發展情形，並評估根據歐盟「數位市場法」（Digital Markets Act, DMA），將提供此類服務的事業指定為受 DMA 規範的「守門人」（gatekeeper）之可能性（Autorité de la concurrence, 2024）。

（三）治理層面：應否立即針對人工智慧進行監管饒富爭議

1、難題一：難以避免空白現象

在技術層面持續精進且應用層面日益多元化下，人工智慧所引發的問題也隨之急遽增加，然而著眼人工智慧爭議展開的早期討論，經常出現言人人殊或說者聚訟之現象，而形成此一情形的主要原因，便在於人工智慧概念內涵的複雜，以及說者立論基礎每有不同所產生之差異（郭戎晉，2020，頁 175）。人工智慧規範討論上可能出現的迷思，主要包括：1、規範「技術」，抑或規範「應用」；2、規範「現在」，抑或規範「未來」；以及 3、「事前」規範，抑或「事後」規範等（郭戎晉，2020，頁 186-192）。

人工智慧治理推動被認為難以避免所謂的「空白現象」（void），其可能成因包括，包括：1、人工智慧發展與創新速度之快，使得對應此一技術的監管路徑存在著高度不確定性；2、執政者為了避免背負破壞技術前瞻發展，從而可能選擇漠視人工智慧之監管；3、在影響層面過大之下，協調不同的監管與執法機關，長期存在著困難；以及 4、私部門往往對政府的監管舉措採取抵制立場（Bollier, 2018）。特別是人工智慧技術與應用領域的快速發展與更迭，更讓各國面對人工智慧所肇致的風險時，不免有著捉襟見肘之感（Lothian, 2023）。

2、難題二：難以援引既有的監管經驗

面對資通訊科技應用衍生的治理課題，若國家在所涉領域的監理發展時間較早，多有著顯著的「路徑依循」（path dependence）特質，亦即面對科技應用帶來的監管難題，執政者多數情況下傾向援引過往的科技監管經驗，同時採用既有可用之立法加以處理（Lin & Nestarcova, 2019）。

惟人工智慧所涉技術與應用領域的複雜性，使得人工智慧難以直接援引既有的監管經驗。Braden R. Allenby（2011）甚早即指出當前針對特定議題所採行的監管作法，包括隱私保護及智慧財產權（如著作、專利）保護等，在欠缺合適性之下，實不應輕易地套用於人工智慧等嶄新技術的監管推動。

3、難題三：難以有效衡量及判定衍生風險

人工智慧發展無可避免地出現正反效益併存情形，從而促使主要國家積極思考人

工智慧合宜的監管作法，然而人工智慧監管架構建構上最重要、但也是最為困難的首要步驟，便是確定可能受到人工智慧影響的領域、伴隨而生的風險種類以及相應之風險級別（Giacobbe, 2022）。

在人工智慧側重自律機制階段， Kate Crawford 及 Ryan Calo（2016）曾指出在人工智慧快速滲透各個行業之際，尚無有效方法可得評估人工智慧應用對於人類社會產生的持續性影響。在監管思維由自律逐步轉向他律後，監管手段的採行亦不免面臨相同問題，由於監管舉措通常是為了避免或降低可能對於人類健康、安全、環境甚或道德層面的風險而推動，然而人們對於人工智慧可能誘發的風險尚未全數知悉，甚至是不可知的（Guihot, Matthew & Suzor, 2017）。國內亦有論者指出「在風險難以掌握甚或不可知之下，針對人工智慧所採取的任何監管舉措，便可能成為假定人工智慧風險存在的預先立法。惟倘若因為與人工智慧發展有關的各式焦慮，便率爾主張政策制定者應該以『預先立法』的立場，提早處理與人工智慧發展所帶來的假設性問題，恐有值得商榷之處」（劉靜怡，2018，頁 7）。

歐盟在確定針對人工智慧立法制定監管立法時，便表明歐盟將立於「風險基準」進行規範設計。AIA 將人工智慧系統應用上可能衍生之風險，具體區分四個風險級別：(1)、無法接受的風險（unacceptable risk）；(2)、高度風險（high risk)；(3)、有限風險（limited risk)；及 (4)、最小風險（minimal risk）／無風險（no risk）。惟人工智慧可能衍生的風險往往言人人殊，諸如受歐盟影響擬制定監管專法的巴西，將人工智慧風險區分為：(1)、過度風險（excessive risk）；(2)、高風險（high risk）；及 (3)、其他（非過度風險亦非高風險之系統）；而加拿大所提出之立法草案，則是將人工智慧風險概分為應受規範「高影響性人工智慧系統」（high-impact AI systems），以及不受限制之「非高影響性人工智慧系統」，突顯出各國在風險界定上的不易與差異存在情形。

參、全球首部全面性監管專法：歐盟人工智慧法之分析

一、歐盟人工智慧法推動經緯

（一）立於「風險管制基準」設計專法

歐盟執委會（European Commission）在 2018 年揭櫫歐洲人工智慧的三個主要發展願景，包括：1、增加公、私部門對人工智慧之投資；2、著眼社會發展預先進行準備；及 3、確保對應人工智慧之適當的道德暨法律架構；歐盟並於「適當的道德暨法律架構」此一願景中明確揭示「建構適當法律架構」的重要性（European Commission, 2018）。

為落實上述願景，執委會成立了「人工智慧高級專家小組」（The High-Level Expert Group on Artificial Intelligence, AI HLEG）並在 2019 年 6 月發布「可信賴人工智慧政策暨投資建議書」（Policy and Investment Recommendations for Trustworthy Artificial Intelligence），表明歐盟應當採取「風險管制基準」（risk-based approach）方法，基此評估充分對應人工智慧監管需求的法律規範設計（AI HLEG, 2019）。

充分考量人工智慧高級專家小組建言並整合各界回饋意見，歐盟執委會於 2020 年 2 月發布「人工智慧白皮書」（White Paper On Artificial Intelligence），正式表明歐盟將採取風險管制基準，作為人工智慧監管之核心精神，並基此制定人工智慧監管專法（European Commission, 2020）。經過廣泛討論，2021 年 4 月執委會正式公布「AIA 草案」，期待藉由制定全球首見的人工智慧全面性監管專法，使歐洲真正成為其所揭示的「足資信賴人工智慧之全球樞紐」此一重要目標。

（二）執委會、歐洲議會與理事會之三方角力與底定

AIA 草案在 2021 年 4 月公開後旋即引發各界關注，展開漫長討論。繼歐洲議會（European Parliament）在 2023 年 6 月 14 日通過其談判立場之 AIA 草案版本，歐洲議會與歐盟理事會（Council of the European Union）並於 2023 年 12 月 9 日達成臨時協議（provisional agreement），針對 AIA 草案揭櫫四項修正重點：

（1）擴大禁止清單（亦即無法接受之風險），惟執法機關於遵循保障措施之前

提下，仍可得於公共場所使用遠端生物識別。

（2）明確高風險人工智慧系統之部署人員，其於系爭人工智慧系統投入市場應用前，負有對之進行「基本權利衝擊分析」（fundamental rights impact assessment）之義務，以利更為妥適地保障權利。

（3）建立可能導致系統性風險的高影響力通用人工智慧模型（亦即正式將生成式人工智慧納入專法之中）及高風險人工智慧系統之規則。

（4）調整治理架構，使其於歐盟層級具備特定執行權（Council of the European Union, 2023）。

歐盟執委會根據臨時協議要求，持續修正 AIA 草案細節並由輪值主席國將其最終版本提交予所有成員國代表批准。歐洲議會與歐盟理事會於 2024 年 3 月 13 日及 5 月 21 日正式投票通過 AIA，全文並於同年 7 月 12 日於歐盟公報發布。

二、對應「人工智慧系統風險」並按風險級別進行規範

（一）區分為四個風險級別

歐盟甚早便確立基於「風險管制」概念研訂 AIA，並根據風險設計不同風險級別的人工智慧系統所應受到的規範。設自草案提出之初至最終正式通過，歐盟始終將 AIA 所界定的人工智慧系統風險，劃分為：「無法接受的風險」、「高度風險」、「有限風險」，以及「最小風險或無風險」等四個風險級別。

除設定與劃分各個風險級別，AIA 同時根據風險級別的高低，對應級別設定其相應之規範密度。若擬供使用的人工智慧系統經判定屬於「無法接受風險」，除符合 AIA 臚列的例外情形，新法將完全禁止此等人工智慧系統之實務應用。若被判定屬於「高度風險」，則該等人工智慧系統於實際應用（進入市場）前，必須遵循 AIA 提出的嚴格規範，包括進行人工智慧系統風險評估作業並執行相應的風險調降措施、提供人工智慧系統高品質之資料集資以減少歧視性結果之發生、留存活動紀錄以確保人工智慧系統設計過程得以事後追溯，並提供清楚且足夠之資訊予使用者等（European Parliament, 2023）。

若人工智慧系統被判定為「有限風險」，其應遵循並符合最低限度透明度

（minimal transparency）之要求，以利使用者作出明智決策，同時在實際應用後，使用者亦可得決定是否繼續使用。若人工智慧系統判定後屬於最後之「最小風險」等級，因不存在風險或僅對使用者造成極小之風險，歐盟將不會對此類系統進行干預（European Parliament, 2023）。

（二）規範分析：無法接受之風險

2024 年 7 月 12 日於歐盟公報發布、同年 8 月正式生效的歐盟 AIA，其第 5 條第 1 項明訂了「無法接受風險」（應受到禁止）之人工智慧系統具體態樣：

（a）於市場展示、出於特定目的提供服務或使用之人工智慧系統，該系統採用超出個人意識之潛意識技術，或有目的之操縱或欺騙性技術，其目的或效果係嚴重扭曲個人或群體之行為，明顯損害該人做出明智決定之能力，從而導致該人做出其原先不會作成之決定，導致或極可能導致該人、其他人或群體受到重大傷害。

（b）於市場展示、出於特定目的提供服務或使用之人工智慧系統，該系統利用特定自然人或群體肇因其年齡、殘疾或特定社會或經濟狀況所存在之任何弱點，其目的或效果係嚴重扭曲該人或隸屬該群體之人之行為，導致或極可能導致該人或其他人受到重大傷害。

（c）於市場展示、出於特定目的提供服務或使用之人工智慧系統，根據自然人或其所屬群體之社會行為，或已知或預測之個人或人格特徵，在一定時期內對自然人或其群體進行社會分數評估或分類。基此所產生之社會分數導致下述其一或兩款情形：

（i）在與最初產製或蒐集資料之目的無關的社會背景下，對特定自然人或整個群體造成不利或負面之對待；

（ii）對特定自然人或整個群體之差別對待，導致其社會行為或社會地位受到不公平或不合比例之影響。

（d）於市場展示、出於特定目的提供服務或使用之人工智慧系統，針對自然人進行風險分析，僅根據針對自然人所作分析或評量其人格特質與特徵，藉以評估該自然人從事刑事犯罪之風險。本款規定並不適用於支持針對特定人是否參與犯罪活動進行人類評估之人工智慧系統，而該等評估係基於與犯罪活動直接相關之客觀且可資驗證之事實。

（e）於市場展示、出於特定目的提供服務或使用之人工智慧系統，透過網際網路或監視器錄影畫面，無針對性地獲取人臉畫面，藉以創建或擴展人臉識別資料庫。

（f）於市場展示、出於特定目的提供服務或使用之人工智慧系統，係應用於工作場所與教育機構等領域，藉以推斷自然人之情緒反應。但出於醫療或安全事由從而擬於市場展示或已於市場實際使用之有關人工智慧系統，不在此限。

（g）於市場展示、出於特定目的提供服務或使用之生物識別分類系統，根據生物識別資料針對自然人進行分類，藉以推斷或斷定其種族、政治觀點、工會成員身分、宗教或哲學信仰、性生活或性取向。本款規定不包括針對合法取得的生物識別資料集所進行之標記或過濾，諸如基於生物識別資料之影像或執法領域之生物識別資料分類。

（h）出於執法目的於公共場所使用「即時性」遠端生物特徵識別系統，但出於下述目的之一且絕對必要者，不在此限：

（i）針對性地尋找綁架、人口販運或性剝削之特定受害者，以及尋找失蹤者；

（ii）防止對自然人之生命或人身安全造成具體、重大與急迫之威脅，或防止真實存在或當前可預見之恐怖攻擊威脅。

（iii）對應附件二所臚列之犯罪行為進行刑事調查、起訴或執行刑事處罰，且有關犯罪行為在相關成員國可被判處不少於四年之監禁或拘留令，針對犯罪嫌疑人進行定位或身分識別。

（三）規範分析：高度風險

1、納為高度風險與否之基本判斷要件

根據 AIA 第 6 條第 1 項規定，於符合下述兩項要件時，人工智慧系統應被視為高度風險系統：

（1）該人工智慧系統擬作為產品的安全元件（safety component）加以使用，或人工智慧系統本身即屬於受附件一所臚列之歐盟立法所規範之產品；

（2）根據第（1）款規定人工智慧系統係產品之安全元件，或人工智慧系統本身即作為一項產品，被要求接受第三方合格評估，以利根據附件一所臚列之歐盟立法，使產品進入市場或提供使用。

基於上述規定，當人工智慧系統屬於產品的安全元件，或者人工智慧系統本身即

屬於產品，同時受到 AIA 附件所表列的歐盟產品相關立法之規範時，該等產品在市場銷售或提供使用前，必須通過由第三方進行之合格評估。而根據 AIA 前言第 50 點規定之說明，可能被界定為高風險人工智慧系統的具體實例，包括：潛在爆炸性環境的機械、玩具、升降機、設備與保護系統、無線電設備、壓力設備、娛樂製程設備、索道裝置、燃燒氣體燃料的器具、醫療器材、體外診斷醫療器械、汽車與航空等。

2、另明訂特定領域之人工智慧系統亦應視為高度風險

除根據上述基本原則從而界定的高度風險人工智慧系統，AIA 第 6 條第 2 項另行規定該法附件三所臚列的人工智慧系統，亦應當視為高度風險人工智慧系統。觀察 AIA 附件三內容，其提出下述八項關鍵應用領域，從而相關領域下之特定人工智慧系統，將視為高度風險人工智慧系統：

（1）生物辨識，受歐盟或成員國立法允許使用；

（2）關鍵基礎設施；

（3）教育與職業培訓；

（4）就業、員工管理與自僱型職業；

（5）獲取及享受基本應有之私人與公共服務及福祉；

（6）執法部門，受歐盟或成員國立法允許使用；

（7）移民、難民庇護與邊境管制，受歐盟或成員國立法允許使用；

（8）司法執法與民主推動之用。

3、高度風險人工智慧系統所受之監管要求

AIA 要求高度風險人工智慧系統應遵循針對此一風險級別所制定的專門規範，並應慮及其預期目的以及人工智慧暨其關聯技術受到公認之技術水平。高度風險人工智慧系統所受到的具體要求，主要包括：

（1）風險管理系統（AIA 第 9 條）；

（2）資料運用暨資料治理規範（AIA 第 10 條）；

（3）編撰必要之「技術文件」並保持最新狀態（AIA 第 11 條）；

（4）記錄留存（AIA 第 12 條）；

（5）透明度與資訊提供（AIA 第 13 條）；

（6）人類監督（AIA 第 14 條）；

（7）正確性、穩定性與網路安全（AIA 第 15 條）。

除針對「高度風險人工智慧系統」所提出之要求，AIA 另規定了高度風險人工智慧系統的「提供者」、「部署者」及「其他利害關係人」所承擔之義務。

另值得留意的是考量人工智慧對健康、安全、基本權利、環境、民主法治潛藏的的重大危害，歐洲議會與歐盟理事會於 2023 年 12 月 9 日通過臨時協議時，在歐洲議會議員要求下，新增納入強制性的「基本權利衝擊分析」（fundamental rights impact assessment）規範要求（Council of the European Union, 2023）。

根據 AIA 第 27 條第 1 項規定，在部署同法第 6 條第 2 項規定所界定的高度風險人工智慧系統之前，除「關鍵基礎設施」領域之高度風險人工智慧系統受到豁免之外，公務機構、提供公共服務之非公務機構，以及附件三第 5 項第（b）款與第（c）款高風度險人工智慧系統之部署者，應評估相關人工智慧系統應用上對於基本權利之所生影響。

AIA 同時明訂基本權利衝擊分析所應納入的事項，其包括：

（1）描述部署者依照其預期目的使用高度風險人工智慧系統之流程；

（2）描述個別高度風險人工智慧系統的使用期限與使用頻率；

（3）在特定情況下可能受到其使用影響的自然人與群體之類別；

（4）考量提供者根據 AIA 第 13 條規定所提供之資訊，可能對上述的自然人或群體產生影響之具體損害風險；

（5）根據使用說明，描述人類監督措施之實施情形；

（6）於相關風險實際出現時所應採取的措施，包括內部治理與申訴機制之設置。

（四）規範分析：有限風險與最小風險／無風險

除明文加以禁止的人工智慧系統（不可接受風險），AIA 的規範重心無疑是「高度風險人工智慧系統」，針對高度風險以外的風險等級，包括「有限風險」與「最小風險／無風險」之人工智慧系統，AIA 第 95 條則是規定依據該法所成立的歐盟人工智慧辦公室（AI Office）及成員國，應當鼓勵與促進制定「行為守則」（codes of conduct），使高度風險以外的其他風險等級之人工智慧系統，可得自願應用並遵循 AIA 第三篇第二章之規範。

AIA 明訂人工智慧辦公室與成員國應當制定可對應所有人工智慧系統的具體要求之自願性行為準則，並以明確的目標與關鍵績效指標（key performance indicators）作為基礎，藉以衡量有關目標之實現情形。AIA 同時例示了行為準則之可能內容，包括：

1、歐盟足資信賴之人工智慧道德準則（Ethics Guidelines for Trustworthy AI）所規定之適用要件；

2、評估並盡量減少人工智慧系統對環境永續（environmental sustainability）之所生影響，包括節能編程（programming）與有效設計、訓練與使用人工智慧之技術；

3、提升人工智慧素養，特別是從事人工智慧開發、操作與使用人員的素養之提升；

4、促進人工智慧系統的包容性與多元化設計，包括透過建立包容性與多元化的開發團隊，並促進利害關係人參與此一過程；

5、評估並防止人工智慧系統對弱勢族群或弱勢族群所屬團體產生的負面影響，包括身心障礙者之無障礙環境以及性別平等。

三、著眼「生成式人工智慧」發展趨勢提出基本要求

（一）後期甫決定對生成式／通用人工智慧進行監管

歐盟執委會於 2021 年提出 AIA 草案時，生成式人工智慧（Generative Artificial Intelligence）或所謂的「通用人工智慧」（General-purpose Artificial Intelligence, GPAI）尚未興未艾，因此最初的 AIA 草案版本中並未見有關生成式人工智慧之規定。隨著 ChatGPT 等生成式人工智慧工具在 2023 年廣受注目，歐洲議會與歐盟理事會於同年 12 月就 AIA 草案達成臨時協議時，也表明在生成式人工智慧／通用人工智慧」快速發展下，通用人工智慧系統暨其仰賴的通用人工智慧模型（GPAI Model），應當受到必要規範（Council of the European Union, 2023）。

歐盟於 AIA 前言中指出大型生成式人工智慧模型是通用人工智慧的典型使用案例，其允許靈活地生成內容，諸如文字、音訊、圖像或視訊等形式，輕易地滿足各種任務。AIA 前言同時強調應當就「通用人工智慧模型」予以明確定義，並將其與「人

工智慧系統」概念進行區分，甫利於實現法律確定性。

考量通用人工智慧系統亦存在導致系統性風險（systemic risks）之可能，且相關風險將隨著模式的範疇與能力之增長而增加，歐盟爰決定於 AIA 中針對「具系統性風險之通用人工智慧模型」進行必要規範。

（二）嚴格規範具「系統性風險」之通用人工智慧模型

為判斷通用人工智慧模型存在系統性風險與否，AIA 第 51 條第 1 項明訂符合下述兩款條件之一者，視為「具系統性風險之通用人工智慧模型」：

1、具有根據適當技術工具與方法（包括指標與基準等）加以評價之「高影響能力」（high impact capabilities）；

2、根據執委會的決定，依職權或遵循科學專家小組（scientific panel）之有效警示，按 AIA 附件十三所列標準進行評估，具有前款規定所稱之能力或影響。

通用人工智慧模型是否該當 AIA 第 51 條第 1 項規定所稱之能力或影響，歐盟於 AIA 附件十三提出了執委會評估上應納為考量的七款標準，包括：

1、模型參數之數量；

2、資料集的品質或規模，諸如是否透過 Token 進行衡量；

3、用於訓練模型的計算量，以浮點運算測量或由其他變數之組合表示，諸如所估算的訓練成本、訓練所需時間或訓練所產生之能源損耗；

4、模型的輸入和輸出態樣（諸如文字至文字（大型語言模型）、文字至圖像、多態樣，以及確定每種態樣的高影響力之最新閾值），以及特定的輸入與輸出態樣（諸如生物序列）；

5、模型能力的基準與評估，包括考慮無需額外訓練的任務數量、學習新的、不同任務之適應性、自主性與可擴展性之水平、以及其可得近用之工具；

6、是否因其覆蓋範圍從而對內部市場產生重大影響，當系爭產品已於歐盟擁有逾一萬個已註冊的企業用戶時，便應推定該產品符合重大影響；

7、已註冊的終端用戶之數量。

AIA 對「具系統性風險之通用人工智慧模型」施加嚴格的義務要求，於該當適用條件時，提供者必須進行模型評估（model evaluations）、衡量並減輕系統性風險、進行對抗測試（adversarial testing）、向執委會通報重大事件並確保網路安全。

四、治理體系、處罰規定與後續實施時程

（一）治理體系設計

為落實此一全球首見的人工智慧監管專法，AIA 要求歐盟應於聯盟層級設置「人工智慧辦公室」（AI Office）。在 AIA 正式底定前，歐盟執委會於 2024 年 1 月揭櫫將於其「資通訊網路暨科技總署」（Directorate-General for Communications Networks, Content and Technology, DG CNECT）轄下設置人工智慧辦公室，並於同年 5 月 29 日正式運作（European Commission, 2024）。

人工智慧辦公室由五個部門組成，包括：

1、法規暨遵循部門（Regulation and Compliance Unit）；

2、人工智慧安全部門（AI Safety Unit）；

3、卓越人工智慧暨機器人部門（Excellence in AI and Robotics Unit）；

4、造福社會人工智慧部門（AI for Societal Good Unit）；

5、人工智慧創新暨政策協調部門（AI Innovation and Policy Coordination Unit）。

根據 AIA 及人工智慧辦公室公告資訊，人工智慧辦公室肩負下述之核心任務：

1、確保新法（AIA）執行上的一致性

人工智慧辦公室將直接負責 AIA 的執行與通用人工智慧有關規範之落實，同時支持成員國主管機關之執法。人工智慧辦公室並將與人工智慧開發商、學界與其他利害關係人，共同研擬與時俱進之行為準則及通用人工智慧之評測標準。

2、確保人工智慧決策係基於充分資訊作成：

人工智慧辦公室將與成員國代表所組成的歐盟「人工智慧委員會」（European Artificial Intelligence Board）、由獨立專家所組成之「科學小組」（scientific panel），以及產業代表、中小企業、新創企業、學者專家等利害關係人之「諮詢論壇」（advisory forum）密切合作，確保相關決策在獲取充分資訊之基礎上作成。

3、推廣足資信賴人工智慧之創新生態體系

人工智慧辦公室將提供實務最佳操作建議，並打造人工智慧監理沙盒（AI regulatory sandboxes）與真實世界測試計畫（real-world testing plan）等資源。此外，人工智慧辦公室將支持人工智慧及機器人技術之研發，並確保在歐洲開發與利用超級

電腦所訓練的通用人工智慧模型可與創新應用加以整合，進而帶動相關投資。

（二）過往未見之高昂處罰金額規定

受規範對象若有違反 AIA 之情形，將面臨包括金錢與非金錢措施在內的處罰。其中，違反第 5 條規定使用了「無法接受風險」之人工智慧系統，將可處以 3,500 萬歐元之罰鍰，若違反者屬於事業，亦得以其全球年度營收之 7% 計算處罰金額，兩者取其高者。此一設計遠遠高出「一般資料保護規則」（GDPR）中的 2,000 萬歐元／全球年度營收之 4% 處罰規定。

針對第 5 條規定以外的其他義務要求，受規範者若有違反時，將可處以 1,500 萬歐元之罰鍰，若違反者屬於事業，亦得以其全球年度營收之 3% 計算處罰金額。此外，受規範者若有向 AIA 所訂公告機關或成員國主管機關提供不正確、不完整或誤導性資料之行為，則可處以 750 萬歐元之罰鍰，若違反者屬於事業，亦得以其全球年度營收之 1% 計算處罰金額。

（三）後續實施進程

AIA 全文在 2024 年 7 月 12 日正式於歐盟公報發布，並於同年 8 月 2 日生效。惟 AIA 並非於 2024 年 8 月 2 日全面實施，其重要施行時程如下：

1、生效後 6 個月（2025 年 2 月 2 日）：正式禁止「無法接受之風險」之人工智慧系統。

2、生效後 9 個月（2025 年 5 月 2 日）：確定通用人工智慧之實務操作行為準則。

3、生效後 12 個月（2025 年 8 月 2 日）：

（1）實施通用人工智慧之治理規則；

（2）成員國確定其主管機關；

（3）執委會進行年度審查並評估是否修正禁止項目。

4、生效後 18 個月（2026 年 2 月 2 日）：委員會發布實施法，針對高風險人工智慧提供者提出上市後監控計畫範本。

5、生效後 24 個月（2026 年 8 月 2 日）：

（1）實施附件三臚列之高風險人工智慧系統之所負義務；

（2）成員國實施罰則；

（3）成員國主管機關建立人工智慧監理沙盒；

（4）執委會進行審查並評估應否修正高風險人工智慧系統清單。

6、生效後 36 個月（2027 年 8 月 2 日）：實施附件二臚列之高風險人工智慧系統之所負義務。

肆、歐盟人工智慧法可望催生之正面效益

一、促使國際正視監管專法之重要性

（一）相較於軟法機制更富執行力

現時國際上應對人工智慧風險的監理模式，殊值關注者包括：1、歐盟模式：推動人工智慧監管專法；2、美國模式：部門立法搭配人工智慧風險管理標準；3、英國模式：根基部會固有權責推動客製化監管等不同作法（郭戎晉，2023b）。此外，另有論者將全球人工智慧治理推動，概分為「集中化」（Centralization）及「分散化」（Fragmentation）兩大陣營（Chapman, 2023），前者當以歐盟為代表，而後者則如美國，並未制定全面性監管專法/集中事權，其所採取的部門立法搭配產業管理標準併進作法，展現了靈活的市場導向思維以及根深蒂固之技術自由主義（techno-libertarian）觀點。

論者指出無論是集中化抑或分散化推動，兩種作法事實上各有優缺，而無絕對的優劣之分（Chapman, 2023）。惟軟法機制長期受到詬病之處，便在於易於出現執行力不足或法律約束力有限等情形；相比之下，歐盟立於「風險管制基準」之全面性監管專法推動作法，雖難以消除可能限制人工智慧技術發展潛力等批評意見，仍較軟法機制更富執行力。

（二）人工智慧法或將帶動新一波布魯塞爾效應

1、歐盟法制之全球影響力

歐盟向來不被視為侵略性的經濟霸主，但其在特定領域的法制規範則顯著地引領全球關聯法制之發展。美國哥倫比亞大學法學院教授 Anu Bradford 指出歷來包括環境保護、消費者安全、市場競爭到當前受矚目的個人資料保護等議題，均可窺見歐盟法

制「外部化」（externalize）至歐盟所轄區域之外並形塑為全球標準的現象，Bradford 並將此一現象稱呼為「布魯塞爾效應」（Brussels Effect）（Bradford, 2020）。

2、歐盟人工智慧法發揮全球影響力之可能

隨著數位主權（Digital Sovereignty）廣受重視下，歐盟近年亦根基「歐洲數位十年」（Europe's Digital Decade）此一發展目標，著眼五大領域展開策略性立法。除「人工智慧戰略」（AI Strategy）立法面向下的 AIA，其他重要立法包括：

（1）資料戰略（Strategy for Data）立法：資料法（Data Act）與資料治理法（Data Governance Act）。

（2）網路安全戰略（Cybersecurity Strategy）立法：網路暨資訊系統安全指令（NIS 2 Directive）。

（3）數位服務包裹（Cybersecurity Strategy）立法：數位服務法（Digital Services Act, DSA）與數位市場法（Digital Markets Act, DMA）。

（4）數位隱私（Digital Privacy）立法：GDPR 與電子隱私規則（e-Privacy Regulation）草案。

布魯塞爾效應概念提出者指出歐盟 AIA 是全球首個人工智慧全面性監管專法，其融入包括道德、信任、基本權利與尊嚴在內之歐洲價值觀，同時反映了歐盟將相關價值觀紮根於法治並接受民主機構監督之做法；儘管眾多科技巨擘業已採用各種道德準則減輕人工智慧衍生之風險，惟歐盟認為自我監管機制仍未臻充分（Bradford, 2023）。更為重要者，歐盟透過針對人工智慧制定具有約束力的嚴峻規則，進一步確認了法治與民主的首要地位，並作為其數位憲政（digital constitution）之重要建構基礎（Bradford, 2023）。

歐盟推動上述的新興數位監管法制，不僅進一步增強了其內部地位，同時也強化了其外部地位，包括 AIA 在內的相關新興立法，被視為可望形塑新一波的布魯塞爾效應（Bendiek & Stuerzer, 2023）。另有論者指出 AIA 讓歐盟將其「監管軟實力」（regulatory soft power）擴展至全球技術領域，此一方法也寓含了有關「數位主權」（digital sovereignty）之意味與論述（Gregorio, 2023）。近年在布魯塞爾效應愈發顯著之下，當前已有國家如加拿大、巴西及中國大陸等，刻正推動自身的人工智慧全面性監管專法（Rabacov, 2023）。

二、主要國家人工智慧監管立法觀察

（一）可見之例剖析：加拿大

加拿大在全球人工智慧治理推動上持續扮演重要角色，包括在 2019 年倡議仿「氣候變化專門委員會」（Intergovernmental Panel on Climate Change, IPCC）成立「人工智慧專門委員會」（International Panel on Artificial Intelligence, IPAI），並帶動後續的全球性人工智慧監管討論。

考量數位經濟環境下人工智慧應用已無處不在，加諸人工智慧能力與部署規模持續擴大，加拿大政府體認如果沒有明確的標準，消費者恐難信任此一技術，而企業亦難以證明其係負責任地使用人工智慧系統，繼歐盟於 2021 年提出 AIA 草案後，加拿大亦於 2022 年提出其人工智慧全面性監管專法草案。

加拿大創新、科學與工業部（Ministry of Industry Innovation Science and Technology）在 2022 年 6 月提出 C-27 法案，C-27 法案又被稱作「數位憲章實施法案」（Digital Charter Implementation Act）。C-27 法案承襲曾在 2020 年提出的 C-11 法案主要內容，但除了原即提出的「消費者隱私保護法」（Consumer Privacy Protection Act, CPPA）與「個人資料與數據保護法庭法」（Personal Information and Data Protection Tribunal Act, PIDPTA）兩部預計更新加拿大現行個人資料保護法之草案，C-27 法案還導入新提議的立法，亦即「人工智慧與資料法」（Artificial Intelligence and Data Act；以下簡稱 AIDA）草案。

惟歐盟 AIA 草案最受到質疑的風險監管機制未臻完善，事實上也出現於針對加拿大 AIDA 草案所作討論。AIDA 草案的規範重心，係避免偏見與傷害等「高影響性人工智慧系統」（high-impact AI systems）實務應用可能衍生的相關風險，然而 AIDA 草案本身並未就何謂「高影響性風險」進行定義，擬留待後續進行補充，不免使得各界對於何種人工智慧系統將受到加拿大新法之規範感到困惑（Thompson, 2022）。

考量草案本身並未明確定義何謂「高影響力」，加拿大創新、科學與工業部在 2023 年 4 月發布「AIDA 配套文件」（AIDA Companion Document），除提出一系列助益公、私部門判斷哪些人工智慧系統將視為「高影響力人工智慧系統」的判斷因素，AIDA 配套文件也進一步揭櫫各界所關注的人工智慧系統具體監管作法。依據加

拿大政府的規劃，AIDA 最快將於 2025 年正式生效。

（二）可見之例剖析：巴西

受歐盟 AIA 影響之例還包括巴西，巴西參議院人工智慧臨時委員會（Senate Internal Temporary Commission on Artificial Intelligence）於 2024 年 2 月提出人工智慧法草案（Bill No. 210／2024），擬針對人工智慧創建廣泛而詳盡的監管架構，並擬於巴西聯邦政府層級建立跨部會的監管體系，而主責部會將由現有行政部門中挑選合適者擔任（Bonomo, 2024）。

草案規定在人工智慧系統實際進入市場之前，必須由開發者與部署者針對人工智慧系統進行風險評估與分級。相較於歐盟 AIA 的四個風險級別，巴西人工智慧法草案則將人工智慧系統實務應用可能衍生的風險，具體區分為三個級別，包括：

1、過度風險（excessive risk）；

2、高風險（high risk）；

3、其他（非過度風險亦非高風險之系統）。

若人工智慧系統被判定為「過度風險」，系爭人工智慧系統將遭到禁用。對此，草案指出下述技術將納入過度風險之列：

1、操縱行為與發現弱點之技術；

2、用以建立社會評分政策；

3、助長有關兒童與青少年之性虐待與性探索；

4、創建具犯罪傾向之人格分類；

5、自主性武器系統；

6、於公共場所使用之遠端生物特徵識別系統。

當人工智慧系統係判定為「高風險」級別，此等人工智慧系統雖然未受到禁止，惟實務應用上必須刻遵草案所設定之相關義務要求。草案同時指出下述領域之人工智慧技術運用將納入高風險之列：

1、關鍵基礎設施；

2、教育；

3、人事招募；

4、公共暨私人之基本服務；

5、司法；

6、自動駕駛車輛；

7、醫療照護；

8、犯罪打擊；

9、行政調查；

10、情緒識別；

11、移民與邊境管理；

12、運用於內容創作與流通藉以促進參與。

凡有違反人工智慧法草案者，可得處以 5,000 萬巴西雷亞爾的罰鍰或以事業全球年度營收 2% 所計算之罰鍰，兩者取其高者。

三、併同帶動「全球性治理機制」之討論

包括人工智慧在內，創新科技帶來的監管難理，其影響範疇往往不以特定國家為限，亦非依憑個別國家立法便能妥適解決，從內國立法至國際規範層次，實有必要發展妥適應對人工智慧衍生挑戰的合宜模式。2023 年以降國際社會在雙邊（bilateral）、區際（regional）與多邊（multilateral）等層面提出了多項與人工智慧有關之國際協議，諸如七大工業國組織在 2023 年 10 月發布的「廣島人工智慧進程」（Hiroshima AI Process），以及歐盟與其他 27 個國家在同年 11 月共同締結之「布萊切利宣言」（Bletchley Declaration）。

其他殊值關注者，還包括聯合國於 2023 年 12 月發布的「著眼人性治理人工智慧報告」（Governing AI for Humanity），揭示了建構全球性人工智慧治理機構的重要性，並具體指出此等機構推動上的基本原則暨其所應具備的七大職能：

1、定期評估人工智慧的未來方向與所生影響；

2、藉助在全球受到廣泛認可的人工智慧治理架構，加強全球性治理工作之相互操作性，以及將其納為國際規範之基礎；

3、制定與協調標準、安全及風險管理架構；

4、藉由國際合作促進人工智慧之開發與使用，以利實現經濟與社會效益；

5、促進人才發展、獲取運算基礎設施、建構多樣化且高品質之資料集，以及受惠於人工智慧之公共財（public goods）之國際合作，藉以實現永續發展目標（Sustainable Development Goals, SDGs）；

6、監控風險、通報事故、並針對緊急應變措施（emergency response）進行協調；

7、透過具約束力之規範，實現法律遵循及問責要求（United Nations, 2023）。

無獨有偶，世界經濟論壇（World Economic Forum, WEF）也在 2024 年 1 月成立「人工智慧治理聯盟」（AI Governance Alliance, AIGA），匯聚來自政府、產業界、學術界與其他利害關係部門之領導人，支持全球負責任地開發與使用透明、包容的人工智慧系統。2024 年 1 月發布的「公平人工智慧策略報告」，指出 AIGA 將致力於：1、負責任的應用與轉型；2、彈性治理與監管；以及 3、建構安全的系統與技術等三項核心工作，實現負責任人工智慧之開發、應用與治理，以期最大限度地降低人工智慧快速發展可能出現之相關風險（WEF, 2024）。

伍、歐盟人工智慧法應予關注之負面影響

一、監管立法或未能發揮其應有作用

（一）AIA 本身存在受詬病之處

作為全球首部全面性人工智慧監管專法，歐盟 AIA 自執委會提出草案以降，便受到各界放大檢視。AIA 草案研議階段所受到的主要批評意見，計有：1、關鍵概念的定義過於模糊；2、與既有法規存在疊床架屋情形；3、未能有效保障公民之基本權利；及 4、無法有效掌握及對應人工智慧實務應用衍生之相關風險（郭戎晉，2023a）。

上揭批評意見最受關注者無疑是 AIA 針對人工智慧衍生風險之監管或未臻完善，而造成風險監管機制設計失當之可能成因，包括了：1、風險分類過於僵化且調整不易；2、難以全盤掌握實務應用之整體風險；以及 3、未賦予個人充分了解風險等必要權利（郭戎晉，2023a）。儘管相關缺失隨著歐盟執委會、歐洲議會及歐盟理事會持

續進行談判及反覆修正草案內容而獲得一定程度之解決，但在 AIA 全文正式底定後，仍可窺見針對條文內容之質疑觀點。

（二）難以避免「監管失調」情形

當前許多國家仍在激烈爭辯人工智慧發展帶來正反效應，監管機關也努力應對其所帶來的挑戰以及如何實現監管平衡（regulatory balance）。Neel Guha、Christie M. Lawrence、Lindsey A. Gailmard、Kit T. Rodolfa、Faiz Surani、Rishi Bommasani、Inioluwa Deborah Raji、Mariano-Florentino Cuéllar、Colleen Honigsberg、Percy Liang 及 Daniel E. Ho 指出若監管制度的推動目標，以及擬藉由監管推動所矯正之損害，若兩者並無法匹配（mismatched）或出現難以控制的失衡情形時，便將出現所謂的「監管失調」（regulatory misalignment）現象，Guha 等人另將之稱呼為「價值衝突」（value conflict）現象（Guha et al., 2023）。

（三）可能阻礙產業發展並扼殺技術創新

歐盟立法展現的數位影響力實不容小覷，諸如 GDPR 已充分證明布魯塞爾效應如何影響全球隱私保護實務運作，而近期甫問世的數位平臺競爭監管、資料治理與人工智慧等新興立法，勢將改變全球關聯公、私部門的運作方式（Rzeszucinski, 2022）。

全球最富名望的科技期刊：自然（Nature）雜誌指出人工智慧應受到必要治理已蔚為全球共識，惟實務操作上人們對於人工智慧衍生風險究竟多大，以及與人工智慧有關的哪些內容實際需要限制，仍存在歧見；而主要國家在人工智慧監管思維與具體採取措施上的差異，突顯現時人工智慧風險監管機制的推動，某種意義而言正處於一場盛大的「監管實驗」（regulatory experiment）階段（Hutson, 2023）。

維護基本權利、保護消費者與企業、確保可競爭的市場以及維護開放與民主之社會，無疑是主要國家著眼數位經濟所推動的監管立法與政策之常見宣稱目標，然而過度監管（overregulation）問題，以及所宣稱的目標與實際達成狀況無法匹配，也讓相關立法無可避免地存在著副作用（side effect）（Codagnone & Weigl, 2023）。諸如英國金融時報便指出儘管歐盟 AIA 立法宗旨之一是擬藉由明確的指導方針，促進人工智慧技術持續成長，但 AIA 嚴峻規範所造成的高昂法律遵循成本，可能反向地扼殺中小企業並損害事業之創新投入（Espinoza, 2024）。

值得留意者，歷來多數論者咸認為歐盟缺乏具全球影響力的科技巨擘，並造成歐

盟長期在全球科技競賽處於下風，與其抱持嚴峻的科技監管立場習習相關。布魯塞爾效應概念提出者 Anu Bradford 在近期提出悖論，其認為數位監管（或缺乏數位監管）決定了一個國家科技產業的命運，係不盡正確的單向思維，歐盟無法培養成功的科技產業事實上存在著多重因素，還包括過於分散的數位單一市場拘限了歐盟內部創新之規模、未臻發達的資本市場對科技事業在歐盟之發展能力造成限制，以及無法海納利用全球人才導致數位技能短缺等。Bradford 強調美國與歐盟之間的技術差距，不應全然歸因於美國的寬鬆與歐洲數位監管之嚴格，應更加公允地看待歐盟所秉持的「權利驅動」之監管模式（rights-driven regulatory model），其與科技產業發展與創新之間的互動關係（Bradford, 2024）。

二、人工智慧監管可能落入科林格里奇困境

在主要國家均傾全力發展人工智慧之下，人工智慧的創新步伐，事實上遠遠超過可能應用於人工智慧的監管工具之創新速度（Abbott, 2014）。美國聯邦交易委員會（Federal Trade Commission, FTC）主席 Lina Khan 也表示面對人工智慧發展趨勢，FTC 嘗試在其初期發展階段便發現「潛在問題」，而非在年復一年之後，當問題根深蒂固並且難以矯正時，甫介入進行處理（McCabe, 2024）。

另有論者如 Ryan Khurana 指出監管機構應當具備充足能力了解監管對象，惟監管機關可能因資源有限而難以吸引充足人才，導致對擬監管的人工智慧領域欠缺充分之認識，造成難以在不阻礙創新的前提下規管人工智慧之技術發展與產業應用（Khurana, 2020）。

針對蔚為國際風潮的人工智慧治理推動，若監管機關的應對能力事實上落後於人工智慧技術之創新與更迭，則各國在人工智慧治理推動上便可能落入所謂的「科林格里奇困境」（Collingridge Dilemma）。David Collingridge 在上一世紀 80 年代提出「科林格里奇困境理論」，其指出前瞻技術發展上可能衍生的負面影響，在技術發展前期著實難以預測，導致在所生影響資訊闕如下，執政者有機會進行控制時，卻不知該控制什麼。當創新技術已深入商業與人類生活各個層面並擁有穩固地位，縱使其所生影響與風險，隨著技術的發展已逐漸明朗，執政者此時知道該控制什麼，卻已陷入

難以控制之因境（Leenes, Palmerini, Koops, Bertolini, Salvini & Lucivero, 2017）。

由於人工智慧並非單一技術概念，在細部技術包羅萬象下，不同的子技術可能滋生的風險或危害，事實上並不盡相同。人們在無法完全掌握人工智慧發展走向及所生風險之下，過早或失當的監管舉措便可能出現適得其反之情形。另一方面，囿於人工智慧技術發展與產業應用的不確定性，若對於應否監管人工智慧躊躇不前，最終便可能出現此一嶄新技術暨其實務應用，已在社會中根深蒂固而愈發不易監管（郭戎晉，2020）。

三、監管舉措未必能解決所有問題

當人工智慧的治理思維由自律逐步轉向他律，人工智慧監管設計也成為各國熱切討論的議題。史丹佛大學「以人為本 AI 研究院」（Stanford Institute for Human-Centered Artificial Intelligence, Stanford HAI）指出無論是何種監管機制，人工智慧監管推動似無可避免地面臨所謂的「監管協調問題」（regulatory alignment problem），亦即可能陷入，監管機制或可能無法有效解決人工智慧實務應用衍生的相關風險，抑或可能與其他的社會價值產生衝突（Guha et al,, 2023）。

HAI 同時指出人工智慧監管推動四種常見建議作法，包括：1、強制揭露（Disclosure）；2、註冊要求（Registration）；3、獲取許可（Licensing）；及 4、稽核（Audit），並不一定能同時解決人工智慧發展所帶來的所有問題，諸如透明性、公平性、隱私保護、正確性與可解釋性等。析言之，人工智慧監管推動可能產生兩難局面，不是出現「監管落差」（regulatory mismatch）現象，就是面臨著「價值衝突」（value conflict）情形，兩者均非執政者所樂見（Guha et al., 2023）。

陸、異中求同之全球治理走向？（代結論）

全球正處於人工智慧監管浪潮之中，而可見的監管作法涵括了各種意義的監管舉措，其中包括「硬法」（hard law）方法，藉由法律與命令施加正式的法律義務，亦有採用「軟法」（soft law）方法，諸如推動行業自我承諾、發布指引、制定標準、原則與行業最佳實作參考等（Burns & Bradley, 2024）。

就硬法方式而言，現時最受矚目者無疑是甫生效的歐盟人工智慧法（AIA），著眼人工智慧系統實務應用可能衍生的風險，將之劃分為四個風險等級並制定相應之規範要求；受到歐盟啟發，包括加拿大及巴西等國，亦刻正推動全面性人工智監管立法。

另一方面，在採用「硬法」方式因應人工智慧衍生挑戰的國家中，並非全數均如同歐盟般制定「全面性監管專法」，而是選擇針對人工智慧概念下的「特定技術」或「特定應用場景」進行規範。前者諸如針對生成式人工智慧技術（Generative AI ）或深度偽造技術（Deepfakes）之立法，而後者則如人臉識別應用（Facial Recognition）之規範立法（Burns & Bradley, 2024）。以美國為例，已有多個州立法規範「人臉識別技術」的應用，當前美國在聯邦層級尚未有一體適用的人工智慧監管專法，州法層級則呈現活躍情形，2023 年至 2024 年全美各州合計提出逾 200 項人工智慧立法草案，除特定領域的州立法，近期科羅拉多（Colorado）州更參酌歐盟 AIA，制定了全美首部全面性人工智慧監管州法（Mendelson, 2024）。

相較於實質性監管立法，亦有國家選擇以制定「基本法」方式規範人工智慧，諸如甫於 2024 年 12 月通過人工智慧基本法（AI Basic Act）的韓國（Kim, 2024）。我國國家科學及技術委員會（以下簡稱國科會）亦於 2024 年 7 月預告「人工智慧基本法」（草案），共計 18 條規定。觀察「人工智慧基本法」（草案）內容，儘管屬於「全面性」專門立法，然相較於歐盟 AIA 係實質監管導向的作用法，我國「人工智慧基本法」（草案）如同其名稱般，該法揭櫫目標的落實，事實上仍有賴各該部會進一步制定實質規範（作用法）或積極採取具體舉措，甫能有效落實人工智慧基本法草案所要求之各該事項。另面言之，基本法揭櫫目標的實現，是否真有必要必須透過立法方式加以規定，抑或事實上不待立法、由權責部會主動推動亦可達成相同效果，實有商榷之必要。本文認為若我國決意推動人工智慧專法，合適之道應如同歐盟般、選擇制定全面性且實質監管人工智慧風險的作用法，而不宜僅是偏向政策宣示意味之基本規範。

觀察國科會提出之草案版本，本文認為其在人工智慧規範上亦有若干主要議題值得審思。課題之一為立法目標係規範「技術」，抑或規範「應用」，從國際立法趨勢而言大抵已聚焦於「人工智慧系統」之規範，並開始納入「生成式人工智慧／通用人

工智慧模型」，我國實有加以明確之必要。課題之二則是基本法所揭示政策目標的落實，將採取「集中化」模式，抑或採取「分散化」治理設計，由於草案各該條文咸使用「政府」一詞，正面視之其可海納各部會共同推動，但亦將直接產生權責不分或不明之弊。歐盟 AIA 被視為集中化治理之代表立法，本文認為我國後續實際落實基本法揭櫫目標時，無論是採取集中抑或分散治理，推動重點當在於事權明確，以避免部會因權責不明從而出現疊床架屋或多頭馬車等受人詬病之情形。此外，「人工智慧基本法」（草案）在規範面向遼闊之下，各該目標的達成，是以「事前監管」（ex-ante）為主，抑或宜採取「事後監管」（ex-post）機制，亦無法於草案中直接窺見，從而有進一步討論與確定之必要。

國科會提出之「人工智慧基本法」（草案）在預告結束之後，國科會於其第 13 次委員會議中表示將修改部分草案內容，其中最受矚目者為擬將原本的「風險分級」，調整為「風險分類」。國科會指出進行調整的主因，在於現階段國際對於人工智慧風險分級尚無共識，只有歐盟以法律方式進行風險分級，其餘國家則是以政策文件方式，鼓勵業者自主採取因應風險措施，爰後續規劃由數位部參考國際規範標準制定風險分類架構，再由各目的事業主管機關就主管業務訂定各自的風險分類與管理規範（王若樸，2025）。本文認為國際對於人工智慧風險分級尚無共識，並不代表我國不能提出合適臺灣的風險分級級別，或不必然須如同歐盟區分為四個級別，但不應逕調整為風險分類此一相對模糊的概念。另一方面，此一調整亦無疑地加深了本文前所述及、相關政策目標是否必然須透過立法方式甫能達成之疑問，

韓國甫通過的新法足勘作為我國借鏡，儘管名稱與我國相同，均命名為「人工智慧基本法」，然而韓國「人工智慧基本法」並非僅作基本規範，部分條文係與歐盟 AIA 相仿之實質要求。觀察韓國人工智慧基本法規範架構，其重點有三：1、人工智慧良好發展的制度架構與信任基礎；2、人工智慧技術發展與產業推廣；及 3、人工智慧倫理與可靠性之確保。其中，在人工智慧良好發展的制度架構與信任基礎，韓國將根據新法成立「「國家人工智慧委員會」與「人工智慧安全研究院」等必要機構，並明確各部會之職掌分工；而在人工智慧倫理與可靠性之確保部分，新法提出「高影響力人工智慧」此一概念，凡提供高影響力人工智慧產品或服務的企業，應採取新法明文要求、用以確保安全性與可靠性之相關措施，同時業者應於實際提供相關產品或

服務前，進行人工智慧影響評估，在在顯見韓國「人工智慧基本法」並非單純之基本法，我國實可評估於現有草案版本中，納入合適之實質規範。

在爭端漸增之下，儘管人工智慧的監管思維已由初期的自律為主，逐步向他律靠攏，截至今日仍有國家對於應否立法全面監管人工智慧監管採取保留態度。以脫歐後展現異於歐盟思維的英國為例，2023 年 3 月英國政府發布的「支持創新之人工智慧監管作法白皮書」，表明英國並無制定全面性監管專法的規劃，英國政府同時強調應避免可能扼殺創新的嚴峻立法，將授權各個部會按其所轄行業別，量身設計合適的客製化監管作法（UK Government, 2023）。

從本文分析可知儘管歐盟人工智慧法成為人工智慧監管推動的里程碑式立法，但現時國際針對人工智慧的治理尚未達成一致性共識。在各國監管作法有別之下，不乏論者警示「監管碎片化」（regulatory fragmentation）現象恐將阻礙人工智慧的創新發展，造成執法落差與貿易障礙，並可能導致出現監管套利（regulatory arbitrage）情形，從而降低監管舉措之有效性（Liu et al., 2024）。布魯塞爾效應概念提出者Anu Bradford 則提醒是否存在嚴峻的監管立法，抑或係監管寬鬆甚或闕如狀態，數位監管/科技監管不必然左右產業的興衰與創新發展；析言之，監管機制僅係可能成因，但並非絕對，應進一步審視其他可能造成影響的相關因素（Bradford, 2024）。

另有研究則指出儘管現時國際上存在著多樣化的人工智慧治理作法，但不同方法之間仍可窺見共通之處，包括： 1、採納 OECD 提出的「人工智慧基本原則」並使用 OECD 之「人工智慧系統定義」；2、採用「風險基準」作法，根據風險等級監管人工智慧實務應用；3、鼓勵公、私部門投入人工智慧治理並採用「事前措施」；4、開始關注「通用人工智慧」之管理；5、關聯專業知識及協調機制之需求（Burns & Bradley, 2024）。本文認為在人工智慧實務應用衍生風險持續加劇下，無論各國是受歐盟人工智慧法所產生的布魯塞爾效應之影響，選擇制定全面性監管專法，抑或擬保持監管彈性而採取軟法機制，異中求同應是可預見之必然走向。

參考文獻

王若樸（2025）。國科會揭 2026 年度科技發展布局、AI 基本法草案進展及影響。iThome。https://www.ithome.com.tw/news/167005

郭戎晉（2020）。〈論人工智慧技術應用、法律問題定位及監管立法趨勢—以美國實務發展為核心〉，《成大法學》，39：173-235。https://doi.org/10.53106/168067192020060039004

郭戎晉（2023a）。〈人工智慧風險治理與監管機制建構之研究—以歐盟監管專法（AIA）與美國風險管理標準為核心〉，《世新法學》，17(1)：109-221。

郭戎晉（2023b）。〈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》，27(6)：18-36。

劉靜怡（2018）。〈人工智慧潛在倫理與法律議題鳥瞰與初步分析〉，《人工智慧相關法律議題芻議》，元照出版，3-49

Abbott, K. W. (2014). The Challenges of Oversight for Emerging Technologies. In Marchant, G. E., Abbott, K. W. & Allenby B. R. (Ed.). *Innovative Governance Models for Emerging Technologies* (pp. 1-16). Edward Elgar. https://doi.org/10.4337/9781782545644.00006

Ahzar, A. (2020, Jan. 26), *The Real Reason Tech Companies Want Regulation*. Exponential View. https://www.exponentialview.co/p/-the-real-reason-tech-companies-want

AI HLEG (2019). *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*. AI HLEG Publishing.

Allenby, B. R. (2011). Governance and Technology Systems: The Challenge of Emerging Technologies. In Marchant, G. E., Allenby, B. R. & Herkert J. R. (Ed.). *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight* (pp. 3-18). Springer. https://doi.org/10.1007/978-94-007-1356-7_1

Autorité de la concurrence (2024, June 28). *Generative Artificial Intelligence: The Autorité Issues Its Opinion on the Competitive Functioning of the Sector*. Autorité de la concurrence. https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/

generative-ai-raises-competition-concerns

Bendiek, A. & Stuerzer, I. (2023). The Brussels Effect, European Regulatory Power and Political Capital: Evidence for Mutually Reinforcing Internal and External Dimensions of the Brussels Effect from the European Digital Policy Debate. *Digital Society, 2*(5), 1-25. https://doi.org/10.1007/s44206-022-00031-1

Bollier, D. (2018). *Artificial Intelligence, The Great Disruptor: Coming to Terms with AI-Driven Markets, Governance and Life*. The Aspen Institute Publishing.

Bonomo, D. (2024, June 11), Key Vote Expected on Brazil's Artificial Intelligence Legal Framework. Dentons. https://www.insideglobaltech.com/2024/06/11/key-vote-expected-on-brazils-artificial-intelligence-legal-framework/#page=1

Bradford, A. (2020), *The Brussels Effect: How the European Union Rules the World*. Oxford University Press Publishing.

Bradford, A. (2023), Europe's Digital Constitution, *Virginia Journal of International Law, 64*(1), 1-68.

Bradford, A. (2024), The False Choice Between Digital Regulation and Innovation, *Northwestern University Law Review, 119*, 377-452. https://doi.org/10.2139/ssrn.4753107

Chapman D. (2023), The Ideal Approach to Artificial Intelligence Legislation: A Combination of the United States and European Union. *University of Miami Law Review, 78*, 265-296.

Codagnone, C. & Weigl, Lin. (2023), Leading the Charge on Digital Regulation: The More, the Better, or Policy Bubble?. *Digital Society, 2*(1), 1-19. https://doi.org/10.1007/s44206-023-00033-7

Council of the European Union (2023, Dec. 9), *Artificial Intelligence Act: Council and Parliament Strike a Deal on the First Rules for AI in the World*. Council of the European Union. 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/

Crawford, K. & Calo, R. (2016, Oct. 13), *There is a Blind Spot in AI Research, Nature*. https://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805

Dickson, B. (2023, May 8), *How Open-source LLMs Are Challenging OpenAI, Google, and Microsoft*. TechTalks. https://bdtechtalks.com/2023/05/08/open-source-llms-moats/

Espinoza, J. (2024, July 16), *Europe's Rushed Attempt to Set the Rules for AI*. Financial Times. https://www.ft.com/content/6cc7847a-2fc5-4df0-b113-a435d6426c81

European Commission (2018). 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 on Artificial Intelligence for Europe, COM (2018) 237 final.

European Commission (2020). White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, COM(2020) 65 final.

European Commission (2024, May 29), *Commission Establishes AI Office to Strengthen EU Leadership in Safe and Trustworthy Artificial Intelligence*. European Commission. https://digital-strategy.ec.europa.eu/en/news/commission-establishes-ai-office-strengthen-eu-leadership-safe-and-trustworthy-artificial

European Parliament (2023, June 14). *EU AI Act: First Regulation on Artificial Intelligence*. European Parliament. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Federal Trade Commission (2023, June 29). *Generative AI Raises Competition Concerns*. Federal Trade Commission. https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/06/generative-ai-raises-competition-concerns

Gregorio, G. (2023). The Normative Power of Artificial Intelligence. *Indiana Journal of Global Legal Studies, 30*, 55-79.

Guha, N., Lawrence, C.M., Gailmard, L.A., Rodolfa, K.T., Surani, F., Bommasani, R., Raji, I.D., Cuéllar, M., Honigsberg, C., Liang, P. & Ho, D.E. (2023), *The AI Regulatory Alignment Problem*. Policy Brief of HAI & Stanford RegLab Publishing.

Guihot, M., Matthew, A. F. & Suzor, N. P. (2017). Nudging Robots: Innovative Solutions

to Regulate Artificial Intelligence. *Vanderbilt Journal of Entertainment & Technology Law, 20*(2), 385-456.

Giacobbe, T. (2022). Adapting to Challenges Posed by The Fourth Industrial Revolution: A Regulatory Call to Action Concerning Cybernetic Technology. *Washington University Jurisprudence Review, 15*(1), 141-169.

Hutson, M. (2023, Aug. 8), *Rules to Keep AI in Check: Nations Carve Different Paths for Tech Regulation*. Nature. https://www.nature.com/articles/d41586-023-02491-y

Khurana, R. (2020, Jan. 23). *Artificial Intelligence Needs Private Markets for Regulation — Here's Why. Observer*. https://observer.com/2020/01/artificialintelligence-regulation-private-markets/

Kim, E. (2024, Dec. 27). *AI Basic Act Passes National Assembly, Aiming for Enhanced AI Reliability and Regulation*. https://www.businesskorea.co.kr/news/articleView.html?idxno=232661

Leenes, R., Palmerini, E., Koops, B., Bertolini, A., Salvini, P. & Lucivero, F. (2017). Regulatory Challenges of Robotics: Some Guidelines for Addressing Legal and Ethical Issues. *Law, Innovation and Technology, 9*(1), 1-44. https://doi.org/10.1080/17579961.2017.1304921

Lin, L. & Nestarcova, D. (2019). Venture Capital in the Rise of Crypto Economy: Problems and Prospects. *Berkeley Business Law Journal, 16*(2), 533-571.

Liu, Y., Boy, H.C., Khurana, S. & Sinha, A. (2024), Artificial Intelligence: Revolutionary Potential and Huge Uncertainties, in The World Bank (Ed.), *Digital Progress and Trends Report 2023* (pp. 85-106). The World Bank Publishing.

Lothian, J. (2023, Dec. 6). *How Nations Are Losing a Global Race to Tackle A.I.'s Harms.* The New York Times. https://www.nytimes.com/2023/12/06/technology/ai-regulation-policies.html

Marr, B. (2024, Apr. 25), *Spotting AI Washing: How Companies Overhype Artificial Intelligence*. Forbes. https://www.forbes.com/sites/bernardmarr/2024/04/25/spotting-ai-washing-how-companies-overhype-artificial-intelligence/

McCabe, D. (2024, June 5), *U.S. Clears Way for Antitrust Inquiries of Nvidia, Microsoft and OpenAI*. The New York Times. https://www.nytimes.com/2024/06/05/technology/nvidia-microsoft-openai-antitrust-doj-ftc.html

Mendelson, K. (2024, June 12), *Artificial Intelligence Governance – First, Build on What You Have*. Guidepost. https://guidepostsolutions.com/insights/blog/artificial-intelligence-governance-first-build-on-what-you-have/#_ftn1

Moss, G. (2024, Apr. 3), *How Attackers Weaponize Generative AI through Data Poisoning and Manipulation*. Barracuda. https://blog.barracuda.com/2024/04/03/generative-ai-data-poisoning-manipulation

OECD (2019a), *Artificial Intelligence in Society*, OECD Publishing.

OECD (2019b), *Recommendation of the Council on Artificial Intelligence*, OECD Publishing.

OECD (2024), *Digital Economy Outlook 2024 (Volume 1): Embracing the Technology Frontier*. OECD Publishing.

Rabacov, P. (2023, May 3), *Brazil's New AI Bill: A Comprehensive Framework for Ethical and Responsible Use of AI Systems*. Access Partnership. https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/

Roose, K. (2023, Nov. 22), *A.I. Belongs to the Capitalists Now*. The New York Times. https://www.nytimes.com/2023/11/22/technology/openai-board-capitalists.html

Rzeszucinski, P. (2022, May 26), *Be Ready for The Brussels Effect — It's Coming to Data And AI*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2022/05/26/be-ready-for-the-brussels-effect---its-coming-to-data-and-ai/?sh=88121a330366

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022), *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. NIST Special Publication 1270.

Securities and Exchange Commission (2024, Mar. 18), *SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial*

Intelligence. SEC. https://www.sec.gov/news/press-release/2024-36

Thompson, K. (2022, June 20), *Canada's New Federal Privacy Bill C-27 – Summary of Significant Impacts and New Proposals*. Dentons. https://www.dentons.com/en/insights/articles/2022/june/20/canadas-new-federal-privacy-bill-c27-summary-of-significant-impacts-and-new-proposals

UK Government (2023). *A Pro-innovation Approach to AI Regulation*. UK Government Publishing.

United Nations (2023), *Interim Report: Governing AI for Humanity*. United Nations Publishing.

WEF (2024, Jan. 18), *AI Governance Alliance Unveils Inaugural Report on Equitable AI Strategies*. WEF. https://www.weforum.org/agenda/2024/01/ai-governance-alliance-debut-report-equitable-ai-advancement/

Wheeler, T., Verveer, P. & Kimmelman, G. (2020), *New Digital Realities; New Oversight Solutions in the U.S.*. The Shorenstein Center on Media, Politics and Public Policy Publishing.

WIPO (2019). *WIPO Technology Trends 2019: Artificial Intelligence*. WIPO Publishing.

Zwetsloot, R. & Dafoe, A. (2019, Feb 8), *Thinking About Risks From AI: Accidents, Misuse and Structure*. Lawfare. https://www.lawfaremedia.org/article/thinking-about-risks-ai-accidents-misuse-and-structure