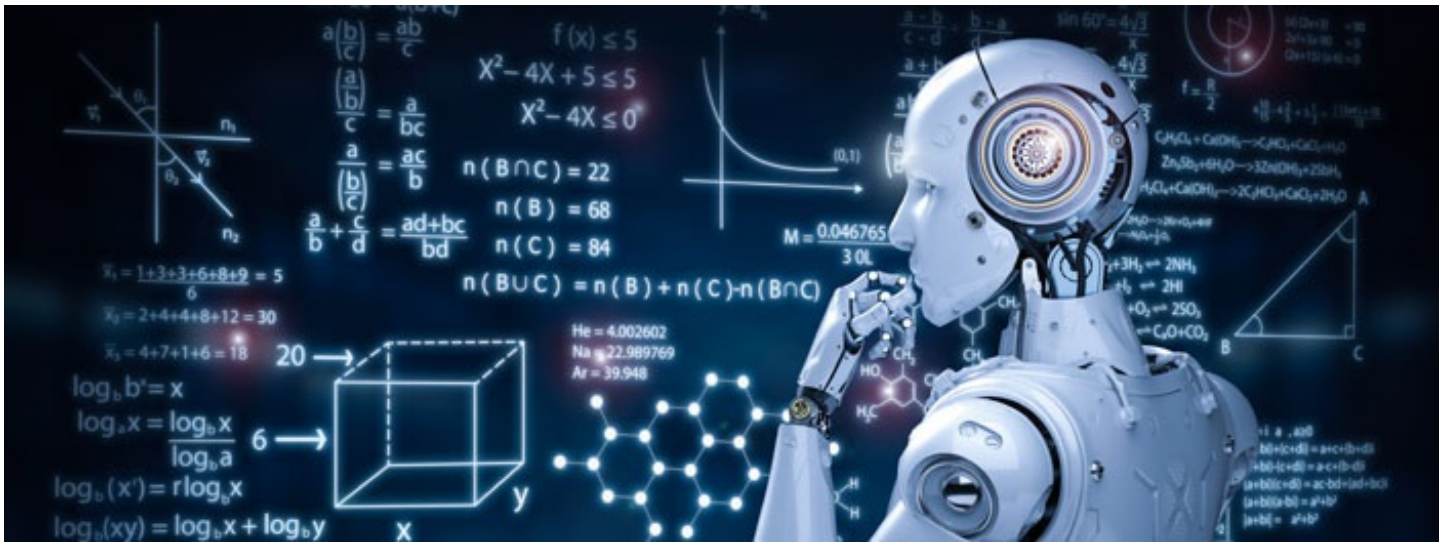
歐盟公布人工智慧法，建立全球首部AI全面監管框架



歐盟公布人工智慧法，建立全球首部AI全面監管框架

資訊工業策進會科技法律研究所

2024年07月12日

歐盟理事會於2024年5月22日正式批准《人工智慧法》（Artificial Intelligence Act，下稱AIA）[1]，該法於2024年7月12日公告於歐盟的官方公報上，將自8月1日起生效，成為全球首部全面性監管AI的法律框架。

壹、事件摘要

人工智慧技術的應用廣泛，隨著使用情境增加，潛在的風險也逐一浮現。歐盟於2018年就提出「可信任的人工智慧」（Trustworthy AI）的概念[2]，認為透過妥善的制度管理人工智慧的研發與使用，即使人工智慧具有多種風險，也可以使民眾享受人工智慧帶來的福祉。因此，歐盟執委會提出全球第一部全面監管人工智慧的法案，為人工智慧的設計、開發、部署、及使用建立適當的規範，希望法律的確定性能促進該技術的創新，並建立各界對於該技術的信心，擴大其採用，使該技術能造福人群。

自從歐盟執委會於2021年4月提出人工智慧法草案以來，其後續發展備受全球矚目，也吸引歐洲的人權組織、學術團體以及大型科技公司的關注。在多方利益關係者的遊說與介入下，該法案一度陷入僵局，其中生成式人工智慧（Generative AI）亦為爭議焦點。歐洲議會和理事會的AIA草案修正版本中，曾經納入生成式AI的定義與監管條款，然最後拍板定案以AI系統與基礎模型為監管對象，並未針對生成式AI。理事會、執委會和歐洲議會經過多次三方會談，終於在2023年12月8日就內容達成協議[3]，草案在2024年3月13日交由歐洲議會大會表決，最終以壓倒性的票數通過該法。[4]

貳、重點說明

AIA全文分為13個章節，總計有113個條文以及13個附件。[5]AIA採分階段實施的方式，該法在生效三年後才可能完全實施。[6]本文擬就該法建立的AI監管框架，包括其適用範圍與規範、管理方式、治理組織、實施和配套措施等規定，擇重點說明如下。

（一）規範對象

AIA的規範對象分為兩類，其一為AI系統；另一為通用人工智慧模型（General Purpose Artificial Intelligence Model, GPAI，下稱通用AI模型）。

1. AI系統

為與國際接軌，歐盟修改AIA有關AI系統的定義，使其與「經濟合作暨發展組織」（Organisation for Economic Cooperation and Development，OECD）的定義一致，令該法更具國際共識基礎。AI系統被定義為「一種機器的系統，它以不同程度的自主性運作，在部署後可能展現適應性，並且對於明確或隱含的目標，從接收到的輸入推斷如何產生預測、內容、建議或可能影響實體或虛擬環境的決策等輸出。」[7]

AIA設有豁免規定，涉及國安和軍事領域、科學研究和開發目的、純粹個人非專業活動使用的AI系統、以及大部分的免費及開源軟體並不適用AIA規範。免費及開源軟體只有屬於高風險或生成式AI系統、或涉及生物特徵和情緒識別目的，才須遵守AIA規範。[8]

2. 通用AI模型

執委會的草案原本不包含通用AI模型，在歐洲議會和理事會的建議下，AIA最後亦將通用AI模型納入監管。所謂通用AI模型，係指具有顯著通用性的AI模型，它可以勝任各種不同任務的執行，並且可以與下游的系統或應用程式整合。[9]

值得注意的是，AIA只約束已經在歐盟上市的通用AI模型，在上市前用於研究、開發和原型設計活動的通用AI模型並不包括在內。

（二）以風險為基礎的分級管理方式

AIA採取風險途徑監管AI系統和通用AI模型，視潛在風險和影響程度決定義務內容，對於兩者建立不同的分類規則，並針對AI系統整個生命週期

進行規劃、建立AI系統和通用AI模型在各階段應符合的要求，由AI價值鏈的參與者分別承擔相應責任，其中以提供者（provider）和部署者（deployer）為主要的責任承擔者。[10]

1. AI系統的分級管理

根據風險程度對系統進行分類，以具有高風險的AI系統為主要規範對象，該類系統在投入市場或使用前必須通過合格評估，並遵守嚴格的上市後規範；而具有不可接受風險的AI系統則禁止使用。另外，AIA還訂有透明性義務，舉凡與人互動、具生成內容能力之AI系統提供者皆應遵守；如果AI產生內容具有深偽（deep fake）效果，其系統部署者還應遵守額外的規定，揭露該內容係人工生成或操縱的結果[11]。

2. 通用AI模型的分級管理

AIA訂有通用AI模型的共通義務[12]，並根據模型的能力判定其是否具有系統性風險（systemic risks）。[13]所有的通用AI模型提供者都須公開模型訓練內容的詳細摘要，並遵守歐盟著作權法的規定[14]；而具有系統性風險的通用AI模型提供者，還須負擔額外的義務。[15]

（三）治理組織

1. AI辦公室

為順利實施AIA，執委會已成立一「人工智慧辦公室」（AI Office，下稱AI辦公室），負責促進、監督AIA落實，它同時也是通用AI模型的監管機構。[16]AIA框架下，會員國市場監管機構僅負責AI系統的監管工作。

2.人工智慧委員會

除了AI辦公室外，還設有一「人工智慧委員會」（AI Board），由歐盟會員國派代表成立，主要負責協調各國的作法、交換資訊、以及提供各國市場監管機構建議。[17]

3.「獨立專家科學小組」與「諮詢論壇」

歐盟層級還有兩個支持性的組織：「獨立專家科學小組」（Scientific Panel of Independent Experts）和「諮詢論壇」（advisory forum），可提供落實AIA規範所需之專業技術知識與實施建議。

獨立專家科學小組的成員係由執委員會指定，執委會將視任務所需的最新科學或技術專業知識進行挑選，該小組最重要的任務在於支援通用AI模型和系統相關規定的實施和執行，包括向AI辦公室通報存在系統性風險的通用AI模型、開發通用AI模型和系統能力評估的工具和方法等。[18]

諮詢論壇成員亦由執委會指定，執委會應顧及商業和非商業利益間的平衡，從AI領域具有公認專業知識的利害關係人當中，尋找適當的人選。諮詢論壇主要任務是應理事會或執委會的要求，準備意見、建議和書面報告，供其參考。[19]

4.會員國內部各自之市場監管機關

在會員國層級，由各國市場監管機關負責督導AIA規定之實施[20]，各國並將成立或指定公告主管機關（notifying authority），負責進行公告合格評估機構（notified bodies）評選與指定事宜，日後將由各公告合格評估機構負責AIA下的第三方合格評估業務。[21]

（四）實施與配套措施

1.分階段實施

AIA的規定將在該法生效24個月後開始實施，然考慮到歐盟和會員國的治理結構尚在討論中，且業界在法遵上也需要時間調適，因此AIA的部分條文將分階段實施。

(1) AIA通則以及不可接受風險的AI系統禁令在該法生效6個月後即實施；

(2) 通用AI模型、第三方認證機構和會員國公告合格評估機構、以及違反AIA的罰則等相關規範，於該法生效12個月後開始實施；

(3) AIA附件III清單之高風險AI系統相關義務，要等該法生效36個月後才開始實施；

(4) 而AIA生效前已上市之通用AI模型提供者，應在該法生效36個月內，採取必要行動使其模型合乎AIA規定。[22]

2.罰則規定

AIA訂有罰則，在AIA措施正式實施後，違規者可能面臨鉅額罰款[23]。

3.配套措施

由於AIA以建立監管框架為主，相關規定之實施細則或標準，這仍待執委會逐步制定。因此，在AIA各配套辦法提出之前，AI辦公室將以「實踐守則」（codes of practice）[24]和「行為守則」（codes of conduct）之訂定與推動為主，另外又提出「人工智慧公約」，希望藉由此些配套措施協助受AIA規範的各方，使其在最短時間內能順利履行其應盡義務。

(1)「實踐守則」

實踐守則（codes of practice）針對的是通用AI模型提供者。AI辦公室將鼓勵所有通用AI模型提供者推動和參與實踐守則的擬定，AI辦公室亦將負責審查和調整守則內容，確保反映最新技術及利害關係各方的觀點。實踐守則應涵蓋通用AI模型和具系統性風險的通用AI模型提供者的義務、系統性風險類型和性質的風險分類法（risk taxonomy）、以及具體的風險評估和緩解措施。[25]

(2)「行為守則」

行為守則（codes of conduct）之目的在於推動AIA的廣泛適用，由AI辦公室和會員國共同推動，鼓勵高風險AI系統以外的AI系統提供者、部署者和使用者等響應，自動遵循AIA關於高風險AI的部分或全部要求。AI系統的提供者或部署者、或任何有興趣的利害關係人，都可參與行為準

則。[26]

(3)「人工智慧公約」

AIA中的高風險AI系統以及其他重要規定需待過渡期結束才開始適用[27]，因此執委會在AIA的框架外，另提出「人工智慧公約」（AI Pact，下稱AI公約）計畫，鼓勵企業承諾在AIA正式實施前，即開始實踐該法規範。

AI公約計畫有兩個行動重點，其一是要提供對AI公約有興趣的企業有關AIA實施流程的實用資訊，並鼓勵這些企業進行交流。AI辦公室將舉行研討會，使企業更了解AIA以及如何做好法遵的準備，而AI辦公室也可藉此收集企業的經驗反饋，供其政策制定參考。

另一個重點是要推動企業承諾儘早開始實踐AIA，承諾內容包括企業滿足AIA要求的具體行動計畫和行動時間表，並且定期向AI辦公室報告其承諾進展；AI辦公室會收集並發布這些報告，此作法不僅有助提高當責性和可信度，亦可增強外界對該些企業所開發技術的信心。[28]

參、事件評析

執委會希望透過AIA提供明確的法律框架，在推動AI創新發展之際，也能確保民眾的安全權利保障，並希望AIA能夠複製GDPR所創造的「布魯塞爾效應」（Brussels Effect），為國際AI立法建立參考標竿，使歐盟成為AI標準的領導者。然AI技術應用的革新發展速度驚人，從AIA草案提出後的兩年內，AI技術應用出現顛覆性的變革，生成式AI的技術突破以及該技術已顯現的社會影響，使得歐盟內部對於AIA的監管格局與力度有了更多的討論，看法莫衷一是。因此，AIA最後定案時，內容有多處大幅調修與新增。

（一）AI系統定義與OECD一致

首先，執委會的原始草案中，強調AI系統的定義方式應根據其關鍵功能特徵，並輔以系統開發所使用之具體技術和方法清單。[29]然AIA最後捨棄詳細列舉技術和方法清單的作法，改採與OECD一致的定義方式，強調AI的技術特徵與運行模式。採用OECD的定義方式固然係因OECD對AI系統的定義更具彈性，更能因應日新月異的AI新技術發展；這樣的作法亦有助AIA與國際接軌、更為國際社會廣泛接受。

（二）規範通用AI模型並課予生成式AI透明性義務

其次，生成式AI衍生的眾多問題和潛藏風險引發全球熱議，在AIA的三方會談過程中，生成式AI的管制也是談判的焦點議題。原本外界以為歐盟應該會在AIA嚴加控管生成式AI的應用，尤其是「深偽」（deep fake）技術的應用。然而「深偽」技術在AIA的分類方式下，卻僅屬於有限風險的系統，雖負有透明性義務，卻僅需揭露若干資訊即可。「深偽」的問題暴露出生成式AI系統的監管難題，最後AIA拍板定案，僅在透明性義務的章節中提及生成式AI，並且以技術描述的方式取代一般慣用的「生成式AI」(Generative AI)一詞。

歐盟另闢途徑管理生成式AI。AIA的原始草案僅針對AI系統，並無管制AI模型的條文[30]，然有鑑於生成式AI模型係以通用AI模型開發而成，因此AIA新增「通用AI模型」專章，從更基礎的層次著手處理生成式AI的問題。在AIA生效後，歐盟境內的通用AI模型將統一由歐盟的AI辦公室負責監管。考慮到生成式AI應用的多樣性，歐盟從通用AI模型切入、而不針對生成式AI進行管理，可能是更務實的作法。

（三）推出多項配套措施強化AI治理與法遵

最後，歐盟在AIA框架外，針對不同的對象，另建多項配套措施，鼓勵非高風險AI系統提供者建立行為守則、推動通用AI模型提供者參與「實踐守則」的制定和落實、並號召AI業者參與「AI公約」提早遵循AIA的規定。這些措施可指導相關參與者採取具體的步驟與作法達到合規目的，俾利AIA之實施獲得最佳成效。

AIA眾多執行細則尚待執委會制定，包括高風險AI清單的更新、通用AI模型的分類方式以及標準制定等，這些細節內容將影響AIA的實際執行。我國應持續關注其後續進展以因應全球AI治理的新格局，並汲取歐盟經驗作為我國AI監管政策與措施的參考。

[1] 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, 2024, OJ L(2024/1689), http://data.europa.eu/eli/reg/2024/1689/oj (last visited July. 12, 2024).

[2] High-Level Expert Group on AI of the European Commission, Ethics Guidelines for Trustworthy Artificial Intelligence, April 8, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last visited June 25, 2024). 該小組在2018年12月提出草案並徵求公眾意見，並於2019年4月正式提出該倫理指引。

[3] 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 (last visited June 25, 2024).

[4] European Parliament, Press Release: Artificial Intelligence Act: MEPs adopt landmark law, March 13, 2024, https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law (last visited June 25, 2024).

[5] European Parliament, *Position of the European Parliament adopted at first reading on 13 March 2024 with a view to the adoption of Artificial Intelligence Act*, https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.html# (last visited June 25, 2024).

[6] AIA, art. 113.

[7] AIA和OECD對AI系統的定義的差異僅在於用字遣詞及語句編排方面，兩者在意涵上其實是一致的。See AIA, art. 3(1).

[8] AIA, art. 2.

[9] AIA, art. 3(63). 執委會原先認為，AI模型無法獨立使用，僅需鎖定AI系統監管即可，然而生成式AI衍生的諸多問題，令人擔憂放任通用AI模型發展可能產生無法預期的後果，因此歐盟最後決定在AIA條文中加入通用AI模型規範。

[10] 但AIA訂有豁免適用的規定，包括國安和軍事領域、科學研究和開發目的、以及純粹個人非專業活動使用的AI皆不受AIA約束。AI價值鏈的其它參與者還包括進口商、授權代表、經銷商等。See AIA, art. 2.

[11] AIA, art. 50.

[12] AIA, art. 53.

[13] AIA, art. 51.「系統性風險」是指通用AI模型特有的高影響力所造成的風險。由於其影響範圍廣大，或由於其對公共健康、安全、公眾的實際或合理可預見的負面影響，進而對歐盟市場產生重大影響。See AIA, art. 3(65).

[14] AIA, art. 53. 在上市前用於研究、開發和原型設計活動的通用AI模型除外。

[15] AIA, art. 55.例如進行模型評估、進行風險評估和採取風險緩解措施、確保適當程度的網路安全保護措施。

[16] Commission Decision On Establishing The European Artificial Intelligence Office, *C(2024) 390 final*, 2024, https://ec.europa.eu/newsroom/dae/redirection/document/101625 (last visited June 25, 2024).

[17] AIA, art. 65.

[18] AIA, art. 68.

[19] AIA, art. 67. 該條款規定，歐盟的基本權利局（The Fundamental Rights Agency）機構、歐盟網路安全局（The European Union Agency for Cybersecurity）、歐洲標準化委員會 (CEN)、歐洲電工標準化委員會 (CENELEC) 和歐洲電信標準協會 (ETSI) 應為諮詢論壇的永久成員。

[20] AIA, art. 70.

[21] AIA, art. 28 & 29.

[22] AIA, art. 113.

[23] AIA, art. 99.

[24] AIA, art. 56.

[25] AIA, recital 116 & art. 56.

[26] AIA, art. 95.

[27] AIA有關治理組織、罰則、通用AI模型的規定於該法生效12個月後才開始實施，屬於附件二範圍的高風險AI系統的相關規定則遲至該法生效36個月後才實施。AIA, art. 113.

[28] European Commission, Shaping Europe's digital future: AI Pact, (last updated May 6, 2024) https://digital-strategy.ec.europa.eu/en/policies/ai-pact (last visited June 25, 2024).

[29] 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, recital (6). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 (last visited June 25, 2024).

[30] 執委會的原始草案中，僅於第四章關於AI系統透明性的條文中提及具有「深偽」（deep fake）能力的系統應負揭露義務。

你可能會想參加

→ 【2023科技法制變革論壇】AI生成時代所帶動的ChatGPT法制與產業新趨勢
→ 「跨域數位協作與管理」講座活動
→ 新創採購-政府新創應用分享會
→ 【線上場】113年「新創採購機制及鼓勵照護機構參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 113年新創採購-照護機構獎勵說明會
→ 【南部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【北部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【中部場】113年「新創採購機制及鼓勵地方政府參與推動」說明會
→ 【臺北場】113年度新創採購-招標作業廠商說明會
→ 【臺中場】113年度新創採購-招標作業廠商說明會
→ 【高雄場】113年度新創採購-招標作業廠商說明會

沈娟娟
法律研究員 編譯整理

上稿時間：2024年07月

文章標籤

推薦文章