

國際AI沙盒的類型與推動模式

◎黃仁志／中華經濟研究院第二（國際經濟）研究所 分析師

隨著AI的快速發展，AI沙盒被視為支持創新、監管風險的重要機制，並因經營者的需求和資源而有所異。學術機構的AI沙盒以支持研究和教學為主，科技公司利用AI沙盒測試新服務功能，國家層級的AI沙盒強調兼顧創新應用和監管風險。未來，我國可以在既有的評測機制上設立AI沙盒，強化我國AI創新能力。

關鍵詞：人工智慧、AI沙盒、AI創新

Keywords: Artificial Intelligence, AI Sandbox, AI Innovation

人工智慧（Artificial Intelligence，簡稱 AI）的快速發展，既讓人充滿期待，但也引發諸多擔憂。一方面期待 AI 協助解決專業而複雜的問題，提升工作效率和品質；另一方面也擔心 AI 濫用造成權益侵犯，帶來未預期的衝突與威脅。因此，如何針對不同類型和影響程度的 AI 技術建立創新支持和應用監管制度，成為政策上必須關注的議題。近期歐盟理事會通過的《人工智慧法案》（Artificial Intelligence Act，簡稱 AIA），即強調以技術應用的風險分級監管策略，兼顧 AI 應用的創新與安全。然而，面對 AI 技術的快速演化與擴散，仍有許多潛在應用會帶來尚未可知的衝擊和影響，包括個人資訊隱私安全、產出成果的準確性、是否存在特定偏見或歧視等，同樣需要

在合適的監管機制下讓其 AI 應用有充分測試的機會，才能真正掌握與評估其應用對社會的可能效益和風險威脅。AI 沙盒（AI Sandbox）的機制，因而成為一個可以在隔離的測試環境中，驗證 AI 應用風險和制度變革需求的推展策略。

沙盒機制的核心是建立一個專屬的隔離環境，用於開發、測試和驗證特定新興技術的運作影響，並在驗證期間獲得特定法律規定的豁免，進而以驗證結果確立開放應用時所需的配套措施和制度調整。在這個受控環境中，開發者和研究人員可以安全地進行各種實驗，而不會對沙盒外實際運行系統和數據造成風險。而 AI 沙盒的特殊性，即在於以驗證 AI 技術於特定應用的衝擊影響為主軸，並具有數據驅動、模型訓練與演算法測