專欄5：金融機構運用人工智慧(AI)科技之潛在風險及監理趨勢

近年來金融機構對人工智慧(artificial intelligence, AI)之應用日益增加，AI技術能有效提升營運效率與客戶服務體驗，為金融機構及消費者帶來益處，但亦衍生金融排斥(Financial Exclusion)[1]、侵犯隱私權、黑箱作業、委外集中度高、群聚行為等問題及風險，進而影響金融體系之穩定。因此，在支持AI技術創新的同時，如何適度監管以確保消費者權益與金融體系穩定，成為各國監理機關之重要議題。本專欄簡述AI之效益與潛在風險，並說明國際間對金融業運用AI之監理趨勢，以及我國參酌國際監理趨勢訂定金融業運用AI之政策與指引內容。

一、人工智慧(AI)之效益與潛在風險

目前全球監理機關對AI尚無一致定義，最常被引用者是金融穩定委員會(Financial Stability Board, FSB)[2] 定義之「運用計算工具來解決傳統上需要人力完成之複雜任務可統稱為AI」。另近期蓬勃發展之生成式人工智慧(Generative Artificial Intelligence, GenAI)，係指可以生成模擬人類智慧創造內容的相關AI系統，其內容形式包括但不限於文章、圖像、音訊、影片及程式碼等。

AI技術透過強大的運算力快速處理大量資料，在預測能力、優化營運及客製化服務等方面發揮極大效益，相關應用包括內部流程自動化、分析顧客資訊提供客製化建議，以及藉由人臉辨識、圖像辨識等技術優化顧客流程等。近年來興起的GenAI技術也為人類生活帶來巨大改變，人們只需將需求及相關資料輸入GenAI系統，便能在短時間內取得結果，大幅減少人工作業時間。

儘管運用AI技術有諸多效益，但其應用於金融領域可能衍生下列潛在風險，若未能良好監督管理，可能影響金融消費者權益及衝擊金融穩定：

(一) 對金融消費者：有個人隱私外洩、預測結果存在偏見或歧視等疑慮。

(二) 對金融機構：面臨AI黑箱作業、責任不明及委外集中少數第三方供應商等風險。

(三) 對金融市場：AI運用(如高頻程式交易)可能出現群聚行為或提高市場連動性等。

此外，2024年1月世界經濟論壇(World Economic Forum, WEF)發布之「2024年全球風險報告」[3]，已將「AI產生之不實資訊」列為2024年全球第二大風險及未來兩年(2024-2026年)之首要風險，並指出若AI未經適當管理而被不當運用，有引發仇恨犯罪及恐怖主義、就業機會消失、犯罪與網路攻擊及偏見與歧視等疑慮，甚至有衝擊全球政治體制、經濟市場及國家安全之虞。

二、金融業運用AI之國際監理趨勢

由於金融業運用AI日益普遍，如何適度監管以維護消費者權益及金融穩定，備受國際組織及各國金融主管機關之重視。2019年經濟合作暨發展組織(OECD)先提出「AI原則」(Principles on Artificial Intelligence)[4]，列出「包容性成長、永續發展與福祉」、「以人為本之價值觀與公平」、「透明度與可解釋性」、「穩健性與安全性」及「問責性」等五項重要原則，並被G20沿用。其後，國際金融組織陸續發布金融機構運用AI之監理建議，歐盟並通過「人工智慧法」，主要國家亦參考國際組織建議陸續提出AI之監理原則或指引。

(一) 國際金融組織提出金融機構應用AI之原則或建議

2021年9月國際證券管理機構組織(IOSCO)發布指引[5]，就市場中介機構及資產管理機構運用AI提出6項監理規範，建議金融機構應：(1)建立適當的治理、控制與監督架構；(2)持續監測AI之發展、測試、使用及表現；(3)確保人員有足夠知識技能及經驗，以使用及監督系統產出之結果；(4)瞭解本身對提供AI服務之第三方機構的依賴性，並建立良好管理與監督機制；(5)對投資人、主管機關及利害關係人提供足夠透明度與資訊揭露；(6)建立適當控制機制，確保資料及系統表現之偏見最小化。

此外，2021年8月國際清算銀行(BIS)旗下之金融穩定學院(Financial Stability Institute, FSI)發布AI監理報告[6]，建議金融監理機關應以「透明度」、「可信賴性與穩健度」、「問責」及「公平與倫理」等4項原則採取相關監理措施，並注意比例原則。該報告並建議將金融部門運用AI，依是否面對客戶區分為兩類，其中面對客戶之AI系統，建議就對客戶影響程度高低(例如聊天機器人之影響較低，信用評分之影響較高)調整監理強度，未面對客戶之AI系統若是須監理機關核准(如法定資本適足性評估)者，應加強監理，若不需核准(如內部作業)者僅需適度監理。

(二) 歐盟通過「人工智慧法」，以風險為基礎區分AI系統之監管等級

2023年12月歐盟通過「人工智慧法」[7]，依風險程度將AI系統分為四種監理等級，必要時政府得對特定風險等級之AI系統進行監管，並適度保留技術創新空間。四種監理等級包括：(1)「不可接受風險」等級：明訂禁止透過AI系統從事特定活動，例如使用敏感特徵(例如政治、宗教信仰、性傾向、種族)之生物識別分類系統、依據社會行為或個人特徵進行社會評分，以及透過網路或監視器(CCTV)蒐集不特定對象之臉部畫面，以建立或擴充臉部辨識資料庫等；(2)「高風險」等級：用於生物辨識、關鍵基礎設施管理、教育與職業訓練及執法等行為，以法規進行約束，並要求業者接受合格評定，且遵循風險管理及網路安全規範；(3)「有限風險」等級：如聊天機器人、AI換臉、GenAI及深度偽造等，須遵循資訊透明化規範；及(4)「低風險」等級：不強

制規範。

(三) 主要國家/地區對金融業運用AI發布監理原則或指引

美國、英國、新加坡、香港、法國及荷蘭等多個已開發國家/地區，皆已參考國際組織之建議，針對金融部門之AI應用發布相關原則或指引，涵蓋「可靠性與穩健性」、「問責制」、「透明度」、「公平性」及「道德倫理」等五大共同原則。其中，前三項原則之監理概念與傳統模型相似，監理機關可以傳統模型之標準為基礎微調，後兩項「公平性」及「道德倫理」原則係為防止AI模型產生具歧視性或偏見之資料或結果，須額外制定相關標準。此外，「資料隱私」、「第三方依賴性」及「營運韌性」亦為許多指導文件關注之面向。

此外，為建立金融業使用GenAI之風險管理架構，新加坡金融管理局(MAS)於2023年11月發布「GenAI對銀行之新興風險與機會白皮書」[8]，係全球首份金融業應用GenAI之指導文件，內容涵蓋金融業運用GenAI之七大風險面向：「公平及偏見」、「道德及影響」、「問責及治理」、「透明及解釋」、「法律及監管」、「監控及穩定」及「網路及資安」，期使銀行業以負責態度運用GenAI，未來並逐步將其應用至整個金融體系。

三、我國參酌國際監理原則，發布金融業運用AI之政策與指引

依據金管會2023年5月調查，受訪175家金融機構中約36%(63家)有採用AI技術，應用範疇包括客群經營(如智能客服)、風險管理與法令遵循(如可疑交易分析)、流程優化(如後臺流程自動化)及數據分析等；至於GenAI應用，我國金融機構及周邊單位多處於評估階段，僅少數規劃導入金融業務或內部作業，但未正式使用。

為協助金融機構善用AI科技優勢，並有效管理風險，金管會參酌OECD等國際組織建議，於2023年10月發布「金融業運用人工智慧(AI)之核心原則與相關推動政策」[9]，明定金融業運用AI之6項核心原則，包括「建立治理及問責機制」、「重視公平性及以人為本的價值觀」、「保護隱私及客戶權益」、「確保系統穩健性與安全性」、「落實透明性與可解釋性」及「促進永續發展」(表A5-1)，並訂定8項配套政策，例如訂定指引、法規調適、督促研訂自律規範等。

此外，2023年12月金管會進一步就前述6項核心原則發布「金融業運用人工智慧(AI)指引」草案，依AI生命週期[10]及所評估風險，提出需關注的重點及可行措施，以鼓勵金融業在風險可控下，導入、使用及管理AI系統，規範方向與FSI、IOSCO建議及主要國家做法大致相同。另金管會亦呼籲金融業相關公會參考該指引，納入或訂定AI自律規範。

表 A5-1 金融業運用人工智慧(AI)之核心原則

核心原則	摘要說明
建立治理及問責機制	金融機構應確保其人員具備足夠之AI應用知識，且應建立風險管理機制，並承擔AI運用相關之外部責任(如消費者隱私保護及資訊安全等)。
重視公平性及以人為本的價值觀	AI運用應秉持以人為本及人類可控原則，並盡可能避免演算法偏見所造成之不公平。
保護隱私及客戶權益	金融機構應尊重客戶選擇是否運用AI之權利；運用AI提供服務時，應充分保護客戶隱私，妥善管理客戶資料。
確保系統穩健性與安全性	金融機構應確保AI系統之穩健性與安全性；如有委外辦理AI系統相關作業，應對第三方業者進行風險管理與監督。
落實透明性與可解釋性	金融機構應確保AI系統運作之透明性及可解釋性，並適當對外揭露。
促進永續發展	AI系統之發展策略應結合永續發展原則，且應提供員工培訓協助其適應新技術，並盡可能維護其工作權益。

資料來源：金管會。

四、結語

AI技術在提升金融服務效率、促進普惠金融及深化客戶關係等方面，具有極大潛力，惟須確保其被妥適應用，並充分因應潛在風險，以維護消費者權益及金融穩定。鑑於AI對金融體系之影響力日增，金管會已參酌國際組織建議及主要國家做法，逐步加強對我國金融業運用AI之監理，本行基於總體審慎監理目的，將持續觀察國際間對金融業運用AI之監理發展，並研究分析AI在我國金融業之運用情形及可能影響，確保在AI運用之效益與風險間取得平衡，以維護金融穩健發展。

註：1. 金融排斥係指經濟弱勢族群無法接觸到主流金融產品與服務之現象。以授信為例，AI模型資料集中代表性不足之群體，可能因模型學習到這些申請人過去沒有獲得足夠多的貸款，而難以取得有利之信用評分。

2. FSB (2017), "Artificial Intelligence and Machine Learning in Financial Services," November.
3. WEF (2024), "The Global Risks Report 2024," January.
4. 參見OECD (2019), "Principles on Artificial Intelligence," May。2024年5月OECD提出該報告之更新版本，參見OECD (2024), "Principles for trustworthy AI," May。
5. IOSCO (2021), "The use of artificial intelligence and machine learning by market intermediaries and asset managers," September.
6. Jermy Prenio & Jeffery Yong (2021), "Humans keeping AI in check – emerging regulatory expectations in the financial sector," *FSI Insights on policy implementation No*. 35, BIS, August.
7. Council of the European Union (2024), "Artificial Intelligence Act," January.
8. MAS (2023), "Emerging Risks and Opportunities of Generative AI for Banks – Executive Summary," November.
9. 金管會(2023)，「金融業運用人工智慧(AI)之核心原則與相關推動政策」，10月17日。
10. AI系統之生命週期，可分為「系統規劃及設計」、「資料蒐集及輸入」、「模型建立及驗證」及「系統佈署及監控」四個階段。