金融業運用人工智慧(AI)之核心原則與相關推動政策

一、前言

近來 AI[1]在金融服務領域的應用日益增加，為金融產業提供客戶服務帶來效益，亦同時衍生一些新的風險問題及監理挑戰。為協助金融機構善用 AI 科技優勢，並能有效管理風險、確保公平、保護消費者權益、維護系統安全及實現永續發展，本會依據行政院「臺灣 AI 行動計畫 2.0」政策規劃及「數位政策法制協調專案會議」之推動策略，並參考全球主要國家監理機關及國際組織之相關指導原則，及結合我國金融市場發展狀況及本會監理政策方向，擬定適合我國金融業的 6 項 AI 應用核心原則，以期引導金融業在兼顧消費者權益、金融市場秩序及社會責任下，積極投入科技創新，促進金融服務升級。

以下就 AI 之影響及國際組織或主要國家對運用 AI 之立場與規定、我國金融業運用 AI 現況、訂定 AI 原則及政策之必要性、我國金融業運用 AI 之 6 項核心原則，以及本會因應 AI 發展推動之配套政策等事項進行說明。

[1] 由於 AI 技術與日俱進，因此各國際組織或各國並未定義 AI，反而聚焦「AI 系統」，本文交替使用 AI 與 AI 系統。依據經濟合作暨發展組織(OECD)對「AI 系統」之定義，係指一種以機器為基礎的系統，在給予設定之一組目的下，能透過產製輸出品(例如預測、建議或決策)來影響環境。它使用以機器或人類為基礎的數據與輸入品來(1)感知真實或虛擬環境；(2)萃取這些感知，並透過自動分析(例如，使用機器學習)或人工分析，再轉化為模型；以及(3)使用模型推理來形成結果的選項。AI 系統係設計以不同的自主程度來運作。

二、 AI 之影響及國際組織或主要國家對運用 AI 之立場與規定

(一)AI 技術的優點及可能衍生之問題

AI 技術的發展，預期將為各領域帶來廣大的經濟與社會效益。這些效益會表現在環境、健康、公共部門、金融、交通、內政及農業等。尤其 AI 運用大量數據資料及運算力，因此在預測能力、優化營運、資源分配，及客製化服務上，能對人類提供相當大的助益。

然而，AI 系統亦引發各種新型態倫理問題，可能影響人類的價值與生活，因此在運用 AI 時須謹慎以對。這些問題可能對決策、勞工就業、社交互動、醫療保健、教育、媒體、資訊獲取、數位落差、個人資料或消費者保護、環境、民主、法治、安全與執法，及人權與基本自由(包括言論自由、隱私和非歧視)等都有影響。此外，AI 可能加重這些倫理的挑戰，因為 AI 演算法可以複製並強化現有偏見，進而加劇現有形式的歧視、偏誤及刻板印象。以往某些問題由人類處理時，情況較容易控制，但 AI 系統具有快速執行任務能力，因此若無妥善的設計與控制，當 AI 代替人類執行而出現事先無法預期情況時，可能帶來極大風險。

此外，近年來生成式 AI[2]快速發展，且影響層面廣泛，已被視為 AI 之一項重大突破。生成式 AI 之導入雖有助於提升生產效率及提供多元之功能與服務，但亦可能涉及個人資料及隱私洩漏、資訊安全風險及其他法律風險等問題，其快速生成新內容之能力，可能創造大量真偽難辨或不存在的資訊，引發虛假訊息傳播之疑慮。

從長遠來看，人類運用 AI 系統後，尤其是生成式 AI 能依據個人偏好生成客製化之文字、影像內容，可能對特殊經驗的感受與代理的感受將與以往不同，同時也將引發更多人類對自我理解、社會、文化與環境互動、自主、代理制度、價值及尊嚴等方面的擔憂。

(二)重要國際組織及主要國家對運用 AI 的看法

由於 AI 近年來快速發展，重要國際組織對 AI 規範也逐漸受到重視，該等組織並呼籲其成員能將其建議納入政策制定、法規及措施中。聯合國教科文組織(UNESCO)於 2021 年發布「對 AI 倫理的建議(Recommendations on the Ethics of AI)[3]」，將「尊重、保護及推廣人權、基本自由及人類尊嚴」列為核心價值，並期望各個政策領域都參考「比例原則與無危害」、「安全

[2] 依歐盟(EU)人工智慧法草案第 28b(4)條文內容，生成式 AI 係指特別用以生成具有不同自主程度之文章、圖像、音訊或影片等內容的 AI 系統。

[3] 詳參 https://unesdoc.unesco.org/ark:/48223/pf0000380455

與防護」、「公平與無歧視」、「永續發展」、「隱私權與資料保護」、「人類的監督及最終決定」、「透明與可解釋性」、「責任及問責」、「覺察與能力素養」、「多元參與、適應性治理及協力」等 10 項原則。

經濟合作暨發展組織(OECD)於 2019 年發布「AI 原則(AI Principles)[4]」，推廣運用「創新與值得信賴，且尊重人權與民主價值」的 AI。其並以價值為基礎，列出「包容性成長、永續發展及福祉」、「以人為本的價值與公平」、「透明與可解釋性」、「穩健及安全性」及「問責」等 5 項重要原則，建議各國政策制定者採用。

G7 數位及科技部長會議於 2023 年 4 月對外表示[5]，同意以 5 項原則發展新興科技(包含 AI)，該 5 項原則為「法治」、「適當程序」、「民主」、「尊重人權」，以及「利用創新機會」。

美國於 2020 年公布並於 2021 年施行「國家 AI 倡議法(National Artificial Intelligence Initiative Act of 2020)[6]」，目的在確保美國在 AI 研發上的領先地位，並在公私部門運用值得信賴的 AI、為現在及未來工作人力與 AI 的整合予以準備，並協調聯邦政府各部門持續進行的 AI 工

[4] 詳參 https://oecd.ai/en/ai-principles
[5] 詳參 https://asia.nikkei.com/Business/Technology/G-7-ministers-agree-to-five-principles-for-assessing-AI-risks
[6] 詳參 https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210

作。白宮於2022年10月發布「AI 權利法案藍圖(Blueprint for an AI Bill of Rights)[7]」，列出「安全與有效的系統」、「運算歧視保護」、「資料隱私權」、「通知與解釋」、「人類替代方案、考量及應變」等5項原則，據以引導AI自動化系統的設計、運用及佈署，以保護美國大眾。

歐盟議會於 2023 年 6 月通過「AI 法案(Artificial Intelligence Act)[8]」的草案，以確保在歐洲開發及使用 AI 可符合歐盟的權利與價值觀。該法案將 AI 系統的風險分為「不可接受風險」、「高風險」、「有限風險」及「低風險」等4類，並依該等風險程度採行「禁用 AI 措施」、「法規管制」、「揭露」及「無法定義務」等對應方式。舉例而言，對民眾生計或權利有清楚威脅的AI措施，例如社會評分為「不可接受風險」，因此即為禁用的措施。對民眾基本權利或安全可能造成不利影響者，例如取得民間基礎服務為「高風險」，此類服務提供者需先進行自我遵法測試，並向歐盟註冊後始得在市場推出服務(既有者按既有規範辦理)，同時必須遵循風險管理、測試、技術穩健度、資料訓練與治理、透明度、人類監督及資安等規範。「有限風險」包含與人互動的系統(如聊天機器人)、情緒辨識系統、生物特徵分類系統、生成或操縱影像聲

[7] 詳參 https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[8] 詳參 https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

音等系統(如深偽)等，須依透明度規範辦理。至於「低風險」者，如電子郵件篩檢程式，則可逕行開發與使用，並無須遵循額外的法規規範，但歐盟將發布行為準則供企業參考。

我國科技部於 2019 年 9 月訂定「人工智慧科研發展指引」[9]，期以「以人為本」、「永續發展」及「多元包容」為核心價值，並兼顧 AI 科研人員學術自由、鼓勵 AI 研究發展創新及維護人權與普世價值的理念，完善我國 AI 科研發展環境。該指引訂有 8 項原則，包括「共榮共利」、「公平性與非歧視性」、「自主權與控制權」、「安全性」、「個人隱私與數據治理」、「透明性與可追溯性」、「可解釋性」及「問責與溝通」等。

行政院 2023 年 4 月核定發布「臺灣 AI 行動計畫 2.0」，揭櫫「以 AI 帶動產業轉型升級、以 AI 協助增進社會福祉、讓臺灣成為全球 AI 新銳」之願景，並從產業端出發，帶動我國整體產業轉型升級，並建構兼顧科技創新及風險治理的可信任 AI 發展環境，及因應 AI 衍生的各項社會衝擊，以 AI 科技發展具包容性的數位經濟，協助解決社會面臨重大挑戰，增進全民福祉。

(三)國際組織及重要國家對「金融業」運用 AI 之監理建議

[9] 詳參 https://www.nstc.gov.tw/nstc/attachments/53491881-eb0d-443f-9169-1f434f7d33c7

近來 AI 在金融服務領域的應用日益增加，其導入雖為金融產業帶來效益，包括提升服務效率、降低服務成本及增加客戶體驗等，但同時也衍生一些新的風險問題或監管挑戰，例如個資及隱私外洩之資安威脅、道德風險、人員轉型等。因此，如何確保金融業在運用 AI 的同時，能夠有效管理風險、確保公平、保護消費者權益、維護系統安全，並實現永續發展等，係當前金融監理機關及金融機構必須正視的課題。

金融穩定委員會(Financial Stability Board, FSB)於 2017 年發布「金融服務業的 AI 與機器學習—市場發展與對金融穩定的涵義(Artificial intelligence and machine learning in financial services—Market developments and financial stability implications)[10]」，說明 AI 與機器學習(ML)可協助金融機構更有效率地處理訊息，亦可透過法遵科技(RegTech)及監理科技(SupTech)，以確保金融機構的遵法性並增強監管效能。同時，FSB 亦提醒相關風險如下：

1.新技術的網路效應及規模擴張性可能導致金融機構對第三方機構的依賴，這些新的系統重要參與者可能不受監管範圍約束。

2.因採用許多以往不曾用過的機構或不相關的數據來源，

[10] 詳參 https://www.fsb.org/wp-content/uploads/P011117.pdf

因此金融市場與機構間可能出現過往未曾意想到的相互聯繫。

3.方法如不具解釋性或可稽核性，可能成為一個宏觀風險。同樣的，廣泛使用不透明的模型亦可能導致意想不到的後果。

FSB 表示，評估使用 AI 及 ML 的風險很重要，包括金融機構應遵守數據隱私、行為風險與網路安全的協議、進行充分的測試與訓練工具，使用無偏差的數據與反饋機制，確保應用程序能夠實現預期的功能。

國際清算銀行(BIS)旗下的金融穩定學院(FSI)於 2021 年發布「人類控制 AI—對金融業新興的監理期待(Human Keeping AI in check—emerging regulatory expectations in the financial sector)[11]」，將金融業運用 AI 系統的方式依是否面對客戶分作二類—面對客戶者，需視 AI 服務對客戶的影響再分二類：低影響(如聊天機器人)及高影響(如信用評分)；未面對客戶者，則視是否需要監理機關核准分作二類：不需要核准者(如內部運作程序)及需要核准者(如法定資本適足性評估)。並依照此等分類，希望金融監理機關依循「透明度」、「可信賴性與穩健度」、「問責」、「公平與倫理」等 4 個原則給予相關

[11] 詳參 https://www.bis.org/fsi/publ/insights35.pdf

監理措施，同時亦希望在推出政策措施時，依照比例原則來處理可能的挑戰。

國際證券管理機構組織(IOSCO)於 2021 年發布「市場中介機構及資產管理機構運用 AI 及 ML 指引(The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers)[12]」，提出 6 項金融監理機關可以採取的措施，包含要求金融機構(1)有適當的治理、控制與監督架構；(2)對 AI 與 ML 的發展、測試、使用及表現持續監測；(3)人員需有足夠知識技能及經驗，以實施、監督及挑戰 AI 與 ML 產出的結果；(4)瞭解本身對提供 AI 與 ML 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、主管機關及相關利害關係人提供妥適的透明度與揭露資訊；(6)有妥適控制機制，以確保資料及 AI 與 ML 的表現能將偏見最小化。

美國[13]、新加坡[14]、南韓[15]等國金融監理機關針對金融業運用 AI，亦提出相關指引及原則，重點多集中在「公平與道德」、「透明」、「問責」、「消費者權益與隱私權保護」、「安全與穩健」等。

[12] 詳參 https://www.iosco.org/library/pubdocs/pdf/IOSCOPD684.pdf
[13] 詳參 https://content.naic.org/sites/default/files/inline-files/AI%20principles%20as%20Adopted%20by%20the%20TF_0807.pdf
[14] 詳參 https://www.mas.gov.sg/-/media/mas/resource/news_room/press_releases/2018/annex-a-summary-of-the-feat-principles.pdf
[15] 詳參 https://www.fsc.go.kr/eng/pr010101/76209

三、我國金融業運用 AI 現況及訂定相關原則與指引之必要性

(一)我國金融業運用 AI 或生成式 AI 之情形

依據本會 112 年 5 月份調查我國金融機構共 175 家，其中有 63 家採行 AI 技術(約 36%)，應用之範疇包含(1)客群經營，如智能客服、機器人理財等；(2)風險管理及法令遵循，包含洗錢防制、分析可疑交易、開戶案件審查等；(3)流程精進，包含影像辨識(OCR)及後臺流程自動化等；(4)數據分析，如客戶屬性及消費等行為數據分析、市場趨勢分析等；(5)其他，如利用威脅情資分析資安情境等。

另為瞭解我國金融機構應用 ChatGPT 或其他生成式 AI 之情形，本會業於 112 年 4 月以問卷形式完成相關調查，結果顯示目前尚無金融機構導入生成式 AI 應用於金融業務或內部作業。惟有 2 家金融機構規劃導入應用，並有 80 家金融機構及周邊單位評估中。

另有 95 家金融機構及周邊單位針對內部員工使用生成式 AI 進行管制，包括設置防火牆以阻擋公司內網連結至 ChatGPT 網站、依機構內部規範控管以防止客戶個人資料外洩等。

未來本會將持續針對生成式 AI 之使用與金融機構進行溝通對話，並視業界發展情況及應用情形適時評估

調整相關規範，以利金融機構遵循。

(二)本會針對金融機構導入新興科技之相關規範：

1.銀行業：

(1)為因應新興科技技術之發展，本會已於 109 年 4 月核備銀行公會修訂金融機構運用新興科技作業規範，以協助銀行管理運用新興科技之風險，並促進銀行業務健全經營，其中包含雲端服務安全控管、社群媒體控管程序、自攜裝置安全控管、生物特徵資料安全控管等新興科技運用相關規範。

(2)另為防範 AI 深度偽造(deepfake)技術，本會已責成銀行公會納入所定「金融機構辦理電子銀行業務安全控管作業基準」，規範銀行辦理電子銀行業務使用視訊會議時，應確認為真人、本人辦理，以防止透過科技預先錄製影片、製作面具、模擬影像或深度偽造等機制偽冒身分，並應留存驗證紀錄及交易軌跡以利查證。並已請銀行公會研議將 AI 納入自律規範。

2.證券期貨業：

(1)防範 AI 透過深度偽造技術進行身分偽造：本會已於 111 年 9 月核備證券期貨業公會修訂新興科技資通安全自律規範納入有關深度偽造技術之防範，包括證券期貨業使用影像視訊方式進行身分驗證時，另應使用

一次性密碼(OTP)、專人電訪或查驗本人並核對證件照片等方式強化驗證，及宜定期辦理涵蓋深度偽造認知及防範之資訊安全教育訓練，以提升業者對新興科技之風險控管，並將視未來 AI 之發展，適時檢討現行相關規範。

(2)強化自動化投資顧問服務之監理，保障投資人權益：現行規範已要求投顧事業執行自動化投顧服務，對於所設計之演算法程式是否符合邏輯、運算結果是否正確等，應進行有效的監督與管理，並應辦理期初評估作業與定期審核，以及指派人員監管該系統，以確保投資人權益；另要求投顧事業內部或集團應組成專責委員會，負責演算法之開發與調整之監督管理，或參與外部軟體開發供應商之審核與實地調查，以評估系統設計之允當性，並應確保該事業對於網路安全已建構完善之預防、偵測及處理措施。

3.保險業：本會因應金融科技之發展，推動保險業結合創新科技(如 AI、區塊鏈、生物辨識等)之運用於保險商品及服務，並制定相關監理規範，如訂定「保險業辦理電子商務應注意事項」、「保險業辦理遠距投保及保險服務業務應注意事項」及各項業務之應遵循事項、自律規範等事項，明定要求保險業須落實資訊安全防護(包括取得資訊安全管理系統國際標準(ISO 27001)、個人資料管理

系統(PIMS)之驗證)、防制洗錢控管、個人資料保護、客戶權益保障及納入內部控制制度等規範，以利保險業者遵循。

(三)本會訂定 AI 原則及指引之必要性

站在科技中立的立場，金融機構運用 AI 仍應遵循現行金融法令及市場自律規範，惟 AI 技術之應用可能衍生如倫理、公平性、隱私權與透明性等問題，為完整現有規範所提醒處理之風險，是以依循國際組織之建議訂定一套適當之原則及政策有其必要性，以便在享有 AI 優勢的同時，亦能確保其對社會經濟及消費者利益的正面效益。

本會參考國際組織及主要國家對 AI 與 ML 之監理原則及指引後，依據 AI 在金融領域的應用實例、可能引發的問題，及在我國金融市場發展狀況，擬定我國金融業應用 AI 的 6 項核心原則，期望透過該等 6 項原則，引導金融機構在 AI 應用上的創新與實踐，同時確保 AI 能在維護金融體系穩定、保障消費者權益等方面發揮其應有效用，本會後續亦將依此 6 項核心原則訂定指引供金融業參考遵循，未來亦將依科技發展及我國金融市場對 AI 系統運用之狀況予以滾動式演進與調整，期能與時俱進，符合實務發展需要。

四、我國金融業運用 AI 之 6 項核心原則

本會參考國際組織或主要國家監理機關針對金融業運用 AI 之相關指導原則，並結合本會「負責任創新」、「強化法遵」、「公平待客」、「普惠金融」、「資通安全」、「資訊揭露」、「永續金融」及「關懷員工」等監理政策理念，提出以下 6 項金融業運用 AI 之核心原則：

(一)原則一：建立治理及問責機制(對應之監理理念：負責任創新)

1.金融機構應對其使用之 AI 系統承擔相應之內、外部責任。內部責任包含指定高階主管負責 AI 相關監督管理並建立內部治理架構；外部責任則涉及對消費者與社會之責任，包括保護消費者之隱私及資訊安全等。

2.金融機構應建立全面且有效的 AI 相關風險管理機制，並整合至現行風險管理及內部控制作業或流程中，且應進行定期的評估及測試。

3.金融機構應確保其人員對 AI 有足夠之知識及能力，並應以風險為基礎做出適當之決策及監督。

說明：

本會長期以來秉持推動「負責任創新」的策略，一方面鼓勵金融業者運用科技發展金融創新商品或服務，一方面亦應重視金融市場秩序及消費者權益保護，以確保金融創新的

風險可控且對金融市場穩定具有正面效益。在此背景下，本會提出「建立治理及問責機制」的原則，強調金融機構在使用 AI 系統時應對內部治理與對消費者的權益保護負責，並對 AI 系統的風險管理與使用進行適當的監督，以確保金融機構在實踐創新的同時，亦實現其對社會責任的承諾，進一步營造出一個穩健且公平的金融市場環境。

「建立治理及問責機制」的原則對於金融機構運用 AI 系統極為重要。此一原則的精神在於，無論在內部或外部，金融機構都需要對其運用的 AI 系統負責。為此，機構必須建立完善的內部治理架構並書面化，指定高階主管進行專責的監督及管理，並明確界定不同業務線的責任及確保紀錄的保存。這些措施有助於保證 AI 系統的正常運作，並能即時發現及解決可能出現的問題。

同時，金融機構的責任也涵蓋了對外部的消費者及社會責任。這涉及到在使用 AI 系統時，必須充分尊重並保護消費者的隱私及資訊安全。此不僅是對消費者權益的基本保障，亦是維護金融機構商業信譽及營運穩定的重要條件。

金融機構須以風險基礎為導向，建立有效的 AI 相關風險管理機制，並且將其整合至現行整體之風險管理及內部控制作業或流程中。此外，AI 系統上線後，金融機構應進行定期的評估及測試，其中包含針對開發時未預料到情況，進行

記錄及監控，必要時進行系統之修正，以確保其 AI 系統的安全性及有效性。

最後，此一原則也強調金融機構需要培養及增進人員對 AI 的知識、風險辨識及管理能力，並應以風險為基礎做出適當之決策及監督。隨 AI 技術之重要性與日俱增，員工(包含開發、測試、監督、法遵、風控及內部稽核等人員)的知識及能力將直接影響到金融機構能否做出恰當的決策，並有效監督 AI 系統的運作。因此，機構需要進行持續的教育及培訓，以確保員工能夠適應該技術的快速發展與變化，且能妥適因應相關風險及挑戰。

(二)原則二：重視公平性及以人為本的價值觀(對應之監理理念：公平待客及普惠金融)

1.金融機構在使用 AI 系統之過程中，應儘可能避免演算法之偏見所造成的不公平。

2.AI 系統之運用應符合以人為本及人類可控之原則，並尊重法治及民主價值觀。

3.生成式 AI 產出之資訊，仍需由金融機構人員就其風險進行客觀且專業的管控[16]。

說明：

「重視公平性及以人為本的價值觀」原則在金融業引入

[16] 參考國科會「行政院及所屬機關(構)使用生成式 AI 參考指引」草案第 2 點。

AI 系統的過程中至關重要。金融機構在應用 AI 系統時，必須充分認知並儘可能避免潛在的演算法偏見，尤其對於面對客戶所提出之 AI 服務，必須確認資料來源的合宜性及數據資料之品質，在正式推出前於獨立環境中測試及驗證演算法，以避免產出市場不樂見之結果，儘可能讓每個客戶都能獲得公平、非歧視性的金融服務，以實現普惠金融之目標。另使用 AI 系統的數據、資料庫及模型，應進行定期審查及驗證準確性，以減少偏差。

此外，以人為本、人類可控的價值觀及應用也是金融機構運用 AI 系統時必須納入考量之層面。除了尊重消費者的隱私權，也要確保金融機構運用 AI 系統時遵循法治及民主等普世價值，並有機制確保 AI 系統之演進仍能維持原本創建該系統之初衷係為協助人類、對人類無危害及確保人類之自主權與可控制權。生成式 AI 產出之資訊，仍需由金融機構人員就其風險進行客觀且專業的管控。

透過強調「重視公平性及以人為本的價值觀」原則，本會期望金融機構在設計 AI 系統之初，即重視多元、包容的價值，並於運用 AI 系統時，提供公平、普惠的金融服務，並且在整個過程中始終符合社會價值觀及民眾期待。

(三)原則三：保護隱私及客戶權益(對應之監理理念：金融消費者保護)

1.金融機構應充分尊重及保護消費者之隱私，並妥善管理及運用客戶資料。

2.金融機構如運用 AI 系統向客戶提供金融服務，應尊重客戶選擇的權利，並提醒客戶是否有替代方案。

說明：

「保護隱私及客戶權益」原則是金融機構在運用 AI 系統時，必須考量到的一個重要環節。在大數據及 AI 技術發展下，客戶的個人資訊常常會被大量蒐集及利用，用於訓練 AI 以提升其準確度，但可能對客戶的隱私權產生影響，也將影響到民眾對金融機構的信任度及服務滿意度。

金融機構於使用客戶資料時，必須充分尊重及保護客戶的隱私權，妥善管理及運用相關資訊，避免任何可能導致個人資料外洩之風險，例如無適當管控機制(所稱適當管控機制，如封閉式地端部署之生成式 AI 模型，且確認系統環境安全性)下，不得向生成式 AI 提供未經客戶同意公開之資訊。金融機構應尊重客戶選擇是否使用 AI 服務的權利，並提醒客戶是否有替代方案，這不僅可以保障客戶之選擇權，亦是維護其權益的一種表現。

透過「保護隱私及客戶權益」原則之落實，金融機構除可符合法規要求，亦能提升消費者信心及滿意度，推動其業務的健全發展。

(四)原則四：確保系統穩健性與安全性(對應之監理理念：強化資通安全)

1.金融機構在運用 AI 系統時，必須確保其系統之穩健性(robustness)與安全性，以避免對消費者或金融體系造成損害。

2.若金融機構運用第三方業者開發或營運之 AI 系統提供金融服務，應對第三方業者進行適當之風險管理及監督。

說明：

在現今日新月異、數位科技快速變革的年代，AI 的運用已成為金融業推動創新與提升服務品質的重要工具。然而，隨著 AI 的普及與應用，其可能帶來的風險也不容忽視，尤其是金融產業特殊的性質，使得任何的系統錯誤或資料洩露，都可能對消費者甚至整個金融體系產生巨大的傷害。這些風險可能源自於技術故障、惡意攻擊，或是由於 AI 系統的運作模式導致的錯誤決策。

因此，金融機構在運用 AI 系統的過程中，須以「確保系統穩健性與安全性」為核心。金融機構不僅需要有足夠的能力來開發及維護安全的 AI 系統，還需要對其 AI 系統的運作結果進行持續監控，並在必要時採取適當的管控措施。

為達到「確保系統穩健性與安全性」之目的，在金融機構運用第三方業者開發或營運的 AI 系統提供金融服務的情

況下，風險管理及監督的要求不可或缺。金融機構在委託第三方處理前，必須對該第三方業者進行盡職調查，評估其是否具備相關知識、專業及經驗，並充分了解與評估其 AI 系統的運作方式及潛在風險(包含資料隱私權之處理、營運風險、資安及集中度風險)，並根據評估結果實施適當的監督策略、管理作為，以防止可能的風險或問題，同時亦須針對第三方之責任範疇予以明定，及預先對於可能引發問題或有不良事件發生時規劃解決方式。另業者應有 AI 相關運算規則並留存軌跡紀錄，俾利後續驗證與管理。

透過「確保系統穩健性與安全性」這項原則，本會期望能在維護金融穩定與保護消費者權益的同時，讓金融機構能夠有信心地、穩健地運用 AI 系統進行創新，為消費者提供更好的金融服務。

(五)原則五：落實透明性與可解釋性(對應之監理理念：資訊揭露)

1.金融機構在運用 AI 系統時，應確保其運作之透明性及可解釋性。

2.金融機構使用 AI 與消費者直接互動時，應適當揭露[17]。

說明：

在人工智慧領域，「落實透明性與可解釋性」原則的重要

[17] 參考國科會「行政院及所屬機關(構)使用生成式 AI 參考指引」草案第 6 點。

性已成各界共識，代表 AI 系統的運作方式應該是可以被理解及解釋的，而將 AI 引入金融服務時，此一原則之必要性更加顯著。

對於金融機構而言，可解釋的 AI 系統讓他們能夠有效評估及管理導入 AI 的風險，並對 AI 系統的效能進行判斷。金融機構須理解 AI 如何做出決策，才能找出潛在問題，並進行必要的調整，以確保對 AI 的運作之有效管理。

對於利害關係人來說，「落實透明性與可解釋性」原則意味著金融機構應評估如何向利害關係人進行適當之揭露，例如金融機構使用 AI 與消費者直接互動時，亦應適當揭露；至於與防制洗錢、資訊安全及詐騙預防有關或涉及企業營業秘密之相關 AI 系統，因揭露過度資訊可能衍生相關風險，則應注意揭露內容或選擇不予揭露。另應確保可解釋性的程度與其 AI 系統應用之重要性相稱，尤其對於受到 AI 系統不利影響的人，能夠基於簡單易懂的因素訊息以及作為預測、建議或決策基礎的邏輯，來挑戰其結果。

(六)原則六：促進永續發展(對應之監理理念：永續金融及關懷員工)

1.金融機構在運用 AI 系統時，應確保其發展策略及執行與永續發展之原則相結合，包括減少經濟、社會等不平等現象，保護自然環境，從而促進包容性成長、永續發展

及社會福祉。

2.金融機構在 AI 系統運用過程中，宜對一般員工提供適當之教育及培訓，使員工能適應 AI 帶來之變革，並盡可能維護其工作權益。

說明：

訂定「促進永續發展」原則之目的，係為提醒金融機構，AI 的應用不僅是為了提高自身的效率及獲利，更應該著眼於如何促進整體社會的包容性成長及永續發展。

金融機構應該確保其 AI 的運用策略與實施方式，均符合永續發展的原則。這意味著在創新金融服務的同時，也要考慮到 AI 如何提升弱勢族群的金融參與，如何降低經濟、社會、性別、年齡等族群之不平等現象，以及如何透過金融科技推動環境保護。在永續發展上，宜注意 AI 系統所需之大量能源可能對環境生態產生之負面影響。

此外，金融機構必須尊重並保護一般受僱員工的工作權益，包括在 AI 轉型過程中，提供適當的教育及培訓以助其適應新的工作環境。這不僅能保障員工的工作權益，還可以培育出能充分利用 AI 系統的專業人才，進一步提升機構的競爭力。

整體而言，此一原則強調金融機構在應用 AI 系統追求經濟效益的同時，也須兼顧永續發展的理念，關懷弱勢，消

除不平等及保障員工權益，促進整體社會的福祉與進步。

五、本會因應 AI 發展推動之配套政策

(一)於上述原則之基礎上，本會將發布金融業運用 AI 指引，考量金融機構運用 AI 系統之程度、對數據之依賴、AI 之功能、自主行為及是否面對客戶等，依比例原則及以風險為基礎訂定應遵循之事項，以促進金融業的健全發展，並維護金融體系穩定及消費者權益。

(二)持續檢視本會相關規範，並適時進行法規調適，以建立金融業應用 AI 系統之完善法規環境。

(三)利用 AI 技術發展監理科技(SupTech)，以提升金融監理之效率及有效性。

(四)與國際組織及其他國家金融監理機關進行交流及合作，以確保本會之監理政策與國際主流發展趨勢保持一致。

(五)持續鼓勵金融業積極參與 AI 的研發及應用，以提供客戶更優質之金融服務或發展法遵科技(RegTech)，並將透過舉辦研討會、工作坊等方式，協助金融機構導入最佳實務做法。

(六)持續就金融業應用 AI 系統之實際情況進行瞭解及檢視，必要時進行專案金融檢查，確保金融機構運用 AI 系統時，能確實遵守法規及落實風險管理，並有助於提升公眾信

任及社會福祉。

(七)責成各金融業公會制訂金融業運用 AI 系統相關自律規範及最佳實務守則，以強化金融機構資安、內控及公平待客，並持續協助員工因應 AI 進行轉型及於導入過程中重視勞工權益。

(八)持續督導金融機構落實公平對待客戶及金融友善準則，並透過金融知識宣導活動，提升民眾使用數位金融工具之觀念，降低數位落差及確保數位金融的公正轉型。

六、結語

隨著人工智慧的快速進步及應用範疇的持續擴大，可信賴之 AI 在未來金融領域日益重要。從改善金融服務效率至深化客戶關係，AI 具有極大的潛力。然而，要善用此一新科技之力量，必須確保 AI 系統的運用能夠在創新及責任之間取得平衡，確保公平、透明，並充分因應可能的風險及挑戰。

本次本會所推出金融業運用 AI 之原則與政策，係為引導金融機構在運用 AI 系統時，不僅能創造價值，更能保護消費者權益，維護金融穩定，並實現包容性的永續發展。本會期望，透過相關原則及政策之實施，金融業能在即將來臨的 AI 時代中，充分發揮其功能，為客戶及社會帶來更大的福祉。

附件：金融業運用 AI 之 6 項核心原則：

一、建立治理及問責機制：

(一)金融機構應對其使用之 AI 系統承擔相應之內、外部責任。內部責任包含指定高階主管負責 AI 相關監督管理並建立內部治理架構；外部責任則涉及對消費者與社會之責任，包括保護消費者之隱私及資訊安全等。

(二)金融機構應建立全面且有效的AI相關風險管理機制，並整合至現行風險管理及內部控制作業或流程中，且應進行定期的評估及測試。

(三)金融機構應確保其人員對 AI 有足夠之知識及能力，並應以風險為基礎做出適當之決策及監督。

二、重視公平性及以人為本的價值觀：

(一)金融機構在使用 AI 系統之過程中，應儘可能避免演算法之偏見所造成的不公平。

(二)AI 系統之運用應符合以人為本及人類可控之原則，並尊重法治及民主價值觀。

(三)生成式 AI 產出之資訊，仍需由金融機構人員就其風險進行客觀且專業的管控。

三、保護隱私及客戶權益：

(一)金融機構應充分尊重及保護消費者之隱私，並妥善管理及運用客戶資料。

(二)金融機構如運用 AI 系統向客戶提供金融服務，應尊重客戶選擇的權利，並提醒客戶是否有替代方案。

四、確保系統穩健性與安全性：

(一)金融機構在運用 AI 系統時，必須確保其系統之穩健性(robustness)與安全性，以避免對消費者或金融體系造成損害。

(二)若金融機構運用第三方業者開發或營運之 AI 系統提供金融服務，應對第三方業者進行適當之風險管理及監督。

五、落實透明性與可解釋性：

(一)金融機構在運用 AI 系統時，應確保其運作之透明性及可解釋性。

(二)金融機構使用 AI 與消費者直接互動時，應適當揭露。

六、促進永續發展

(一)金融機構在運用 AI 系統時，應確保其發展策略及執行與永續發展之原則相結合，包括減少經濟、社會等不平等現象，保護自然環境，從而促進包容性成長、永續發展及社會福祉。

(二)金融機構在 AI 系統運用過程中，宜對一般員工提供適當之教育及培訓，使員工能適應 AI 帶來之變革，並盡可能維護其工作權益。