「生成式 AI 對金融機構業務應用的挑戰與監管影響」期末報告

委託　單位：　財團法人台北外匯市場發展基金會
計畫主持人：　周振鋒教授（政治大學法律學系）
共同主持人：　莊弘鈺副教授（政治大學科技管理與智慧財產研究所）
共同主持人：　鄭婷嫻副教授（中央大學產業經濟研究所）
研究　助理：　王士愷、劉映廷、龔英斌

中　華　民　國　115　年　10　月　14　日

目錄

第一章　緒論 **1**

第一節　研究背景 1

第二節　研究目的 2

第三節　基本法律爭議與監管模式 3

一、　法律爭議 3

二、　監管模式 7

第四節　研究方法與架構 12

一、　研究方法 12

二、　研究架構 12

第二章　生成式 AI 在臺灣金融機構之應用現況 **13**

第一節　生成式 AI 之技術基礎與功能特性 13

一、　生成式 AI 之技術原理概述 13

二、　生成式 AI 之發展與演進 14

三、　生成式 AI 於金融應用之核心功能特性 17

四、　生成式 AI 應用之新興風險 17

第二節　生成式 AI 發展下 FinTech、RegTech 與 SupTech 之制度定位與互動關係 21

一、　FinTech、RegTech 與 SupTech 之定義與發展脈絡 21

二、　生成式 AI 發展下三類科技之角色分工與互動關係 22

三、　生成式 AI 對金融監理科技生態之影響 23

第三節　金融機構運用生成式 AI 之主要應用領域 24

一、　風險管理與監理高度敏感之應用 24

二、　投資決策支援與理財機器人之應用 25

三、　個人化客戶互動與行銷應用 25

四、　內部營運、報表生成與制度支援應用 25

第四節　生成式 AI 於金融機構之實務應用案例 26

一、　國內實際應用實例 26

二、　國外實際應用案例 32

第五節　生成式 AI 於金融機構之應用場景類型化分析 39

一、　內部行政與營運輔助型應用類型：高度採用生成式 AI 39

二、　對客互動與資訊提供型應用類型：中高度採用生成式 AI 39

三、　決策輔助與評估支援型應用類型：低度採用生成式 AI 40

第三章　歐盟 AI 監管法制 **42**

第一節 General Data Protection Regulation (GDPR) 42
一、 適用範圍 42
二、 資料處理原則與合法基礎 42
三、 剖析與自動化決策 43
四、 資訊揭露義務 45
五、 資料保護影響評估 45
第二節 Digital Operational Resilience Act（DORA） 46
一、 立法背景及立法目的 46
二、 適用範圍 47
三、 金融實體之治理與風險管理框架 47
四、 數位營運韌性測試制度 48
五、 ICT 第三方風險管理原則 48
六、 關鍵 ICT 第三方之監督制度 49
七、 違規裁罰框架 49
八、 執法現況 50
第三節 Consumer Credit Directive II（CCD II） 50
一、 適用範圍之擴張與數位化信貸納入 51
二、 前契約資訊揭露義務 51
三、 信用能力評估義務 52
四、 自動化處理之銜接規範 52
五、 執法現況 53
第四節 EU Artificial Intelligence Act（AI Act） 53
一、 立法目的與適用範圍 53
二、 風險分級制度 54
三、 生成式 AI 與通用型 AI 之特別規範 56
四、 裁罰機制 57
第五節 小結 57
第四章 美國 AI 監管法制 **59**
第一節 聯邦層級 59
一、 行政部門的治理工具 59
二、 既有聯邦法律對 AI 於金融產業之適用 62
第二節 州層級立法 69
三、 州法 69
四、 州層級行政機關之既有法規適用 73
第三節 小結 77

第五章　新加坡 AI 監管法制 **79**
第一節　國家級治理框架 79
一、　AI 治理框架第二版 79
二、　生成式 AI 治理框架 80
第二節　FEAT 原則與檢測工具 82
一、　FEAT 原則 82
二、　真理倡議 83
三、　登月計畫 85
第三節　強化生成式 AI 風險管理與聯盟合作 85
一、　與生成式 AI 相關網路風險資訊文件 86
二、　AI 模型風險管理資訊文件 87
三、　銀行業生成式 AI 護欄手冊 89
四、　關於生成式 AI 在銀行業面臨的新興風險與機會白皮書 91
五、　AI 風險管理執行手冊 92
六、　MindForge AI 風險管理營運手冊 94
第四節　既有法規範持續適用 100
一、　個人問責與行為準則 100
二、　科技風險管理指引 101
三、　銀行業委外指引 101
第五節　小結 103
第六章　臺灣金融產業現況與挑戰 **105**
第一節　臺灣 AI 監管法制發展現況 105
一、　《人工智慧基本法》 105
二、　金融業專屬規範 108
三、　其他規範 114
第二節　臺灣現行法制於金融實務之適用缺口 117
一、　各項原則欠缺量化的判斷標準 117
二、　生成式 AI 模型的黑盒問題（可解釋性問題） 117
三、　個資利用問題 118
四、　資安問題 118
五、　外包管理與監理 118
六、　對內問責機制集中於管理層 119
七、　小結 119
第三節　臺灣與國際法制之比較 120
一、　治理模式差異 120

二、 對黑盒與可解釋性問題之應對差異 120
三、 第三方管理與對跨國科技公司的監理差異 121
四、 可加強問責機制與職能劃分 122
五、 委外作業之「重大性」標準一致 123
六、 小結 123
第七章 結論與建議 **126**
參考文獻

第一章　緒論

第一節　研究背景

生成式 AI（Generative Artificial Intelligence）技術的快速演進，尤其以大型語言模型（Large Language Models, LLMs）為代表，正逐步進入金融業的多種應用場域，包括客服助理、自動化報告撰寫、投資諮詢、風險模型輔助、合規監控與內部流程自動化等。根據 Saha 等人在 2025 年的研究《Generative AI in Financial Institution: A Global Survey of Opportunities, Threats, and Regulation》，全球多家銀行、資產管理公司與 Fintech 公司正透過生成式 AI 整合客服、資料摘要、風險模型輔助等功能，但也面臨偏誤、模型不透明、責任認定與監管挑戰等問題[1]。

同時，國際監管機構也注意到 AI 在金融領域可能引發的系統性風險。以金融穩定理事會（Financial Stability Board, FSB）的報告為例，其指出隨著 AI 採用日益擴散，金融機構對 AI 的依賴可能帶來第三方供應商風險、模型風險、市場連動性增加、網路攻擊與系統脆弱性提升等問題[2]。

在法制層面，部分國家已開始對 AI 展開監理制度布局。例如，歐盟已通過《人工智慧法》（Artificial Intelligence Act, 下稱 AI Act），將金融用途中的信用評分、風險評估等 AI 系統列為「高風險 AI」，要求可解釋性、人為監督與風險管理等機制[3]。

然而，目前臺灣在金融法制及監理框架中尚未對生成式 AI 應用進行系統性制度規劃。金融監督管理委員會（下稱金管會）於 2024 年 6 月發布之《金融業運用人工智慧（AI）指引》[4]，雖對金融機構導入、使用及管理 AI 具有高度參考價值，但仍非完整的監理方式。至於 2026 年 1 月 14 日經總統公布並由國家科學技術委員會擔任主管機關之《人工智慧基本法》[5]，該法確立了人工智慧研發、應用及治理基本原則，以鼓勵技術創新及與國際接軌，然由該法第 10 條「政府得透過法令或指引建立標準、

[1] Bikash Saha et al., *Generative AI in Financial Institution: A Global Survey of Opportunities, Threats, and Regulation*, https://arxiv.org/abs/2504.21574 (last visited Jul. 5, 2026).
[2] FSB, *The Financial Stability Implications of Artificial Intelligence*, https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/?utm_source=chatgpt.com (last visited Jul. 5, 2026).
[3] EU Artificial Intelligence Act, *High-level summary of the AI Act*, 27. Feb. 2024, https://artificialintelligenceact.eu/high-level-summary/?utm_source=chatgpt.com (last visited Jul. 5, 2026).
[4] 金融監督管理委員會，金管會發布「金融業運用人工智慧(AI)指引」-金融監督管理委員會全球資訊網，2024 年 6 月 20 日，https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0,2&mcustomize=news_view.jsp&dataserno=202406200001&dtable=News。
[5] 行政院通過「人工智慧基本法」，2025 年 8 月 28 日，https://www.ey.gov.tw/Page/9277F759E41CCD91/5d673d1e-f418-47dc-ab35-a06600f77f07（最後瀏覽日：2026 年 7 月 5 日）。

驗證、溯源或問責機制，強化人工智慧決策之可驗證性及人為可控性」可知，目前生成式 AI 在金融監理與法制發展的領域尚屬初步，仍需政府建立更多法令或指引。

第二節　研究目的

本研究旨在釐清生成式 AI 在金融機構中的具體應用、法律風險與制度挑戰，並提出具體政策建議，以期協助監理機關與金融業在創新與風險之間取得平衡。事實上，許多國際金融機構早已積極開發與導入生成式 AI 技術，並廣泛應用於財務管理助理、信用分析與授信、潛在詐欺偵測與自動化客服等領域。例如：彭博（Bloomberg）於 2023 年所推出的金融專用大型語言模型 Bloomberg GPT[6]、萬事達卡（Mastercard）於 2024 年推出的異常交易偵測工具 DI Pro[7]，及摩根史丹利（Morgan Stanley）亦於 2024 年推出顧問助理 Debrief[8]。而根據金管會於 2025 年所公布的「金融業應用人工智慧(AI)調查結果」顯示，我國目前已有 126 家金融業者導入 AI 技術，占總調查對象總數的 33%，應用場域則集中於內部行政作業、智能客服及防制金融犯罪[9]。金融業導入 AI 科技，顯然已為全球產業的趨勢。

此外，2023 年哈佛商學院的一項研究更顯示，使用生成式 AI 顧問相較於未使用生成式 AI 顧問，其生產力提高 12.2%、速度加快 25.1%、品質提升 40.0%，足見生成式 AI 技術的應用，確實可以有效提升效率與客戶體驗[10]。然而，AI 的應用也並非毫無風險，2024 年史丹佛大學公布的 AI Index 報告便指出，AI 技術在隱私與數據安全、透明度與問責制度、公平性與偏見以及可解釋性等方面仍存在隱憂[11]。換句話說，對金融業而言，

---

[6] Bloomberg, *Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for Finance*, Bloomberg L.P., 30. March. 2023, https://www.bloomberg.com/company/press/bloomberggpt-50-billion-parameter-llm-tuned-finance/ (last visited Jul. 5, 2026).

[7] Mastercard, *Mastercard supercharges consumer protection with gen AI*, Mastercard, 1. February. 2024, https://publish-p58776-e1271253.adobeaemcloud.com/us/en/news-and-trends/press/2024/february/mastercard-supercharges-consumer-protection-with-gen-ai.html (last visited Jul. 5, 2026).

[8] Morgan Stanley, *Launch of AI @ Morgan Stanley Debrief*, Morgan Stanley, 26. June. 2024, https://www.morganstanley.com/press-releases/ai-at-morgan-stanley-debrief-launch (last visited Jul. 5, 2026).

[9] 金融監督管理委員會新聞稿，金管會公布金融業應用人工智慧(AI)調查結果，2025 年 5 月 20 日，https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0,2&mcustomize=news_view.jsp&dataserno=202505200001&dtable=News（最後瀏覽日：2026 年 7 月 5 日）。

[10] Fabrizio Dell'Acqua et al., *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, Sept. 15, 2023), https://doi.org/10.2139/ssrn.4573321 (last visited Jul. 5, 2026).

[11] Nestor Maslej et al., *Artificial Intelligence Index Report 2024*, 163, Stanford, CA: Stanford University, Apr. 2024, https://arxiv.org/abs/2405.19522 (last visited Jul. 5, 2026).

AI 技術的應用至少帶來兩大面向的挑戰：第一，技術及營運風險挑戰，與第二，監管挑戰。

在技術及營運風險挑戰方面，生成式 AI 需要大量高品質的數據進行訓練，若數據品質不足或存在偏見，可能導致決策錯誤或對低收入群體造成歧視。而生成式模型運作如黑盒（Black Box），將降低金融決策的透明度與可解釋性，進而削弱客戶信任。此外，生成內容也可能被惡意濫用，如深偽（Deepfake），並進一步引發詐欺和數據外洩的資安風險。

在監管挑戰方面，生成式 AI 亦對現行金融監管架構帶來挑戰，包括：如何確保 AI 系統的透明度、AI 系統錯誤時的責任歸屬、如何監控與管理新創科技進入金融市場、如何評估創新業務模式對金融穩定的影響，以及如何調和不同司法管轄區間的監管差異等。由於臺灣金融市場與國際具有高度連結性，監理政策除需與快速迭代的國際規範接軌，更需回應「金融服務業公平待客原則」的核心精神。綜上所述，本研究希望達成以下研究目的：

1. 系統性梳理生成式 AI 在金融機構各種業務場景中的可能應用模式與潛在法律風險。
2. 比較歐盟、美國與新加坡等法域對生成式 AI 在金融領域之監理體系與政策態度。
3. 分析臺灣現行金融法規（銀行法、金融控股公司法、金融消費者保護法、個人資料保護法等）在面對生成式 AI 應用時的適用困境與規範缺口。
4. 建構一套適合臺灣情境的生成式 AI 在金融機構的監理政策與法制調整建議，期兼顧創新、安全與金融消費者保護。

第三節　基本法律爭議與監管模式

一、　法律爭議

AI 作為一個創新技術而言，全新的科技帶來的不僅是快速且多變的發展，也對社會產生諸多變革與衝擊。對於現行法制而言，新的技術往往立於舊體制之上，從而可能發生無法律可監理之真空地帶。綜觀全球各地政府，因各國民情、立法體系迥然不同，其所面臨此一衝擊時的因應方式亦可能採取不同思維，反映出不同的文化價值影響人們對於 AI 的接受程度與如何與之共處[12]。

而環視全球作為，大致上可區分為硬法流派及軟法流派兩種。以歐盟為代表性之硬法流派，主張強烈監控方式，並要求明文化、實體化權利

[12] Aaron, J. B., Yuanyuan, Z. & Ana, *AI and culture: Culturally dependent responses to AI systems.* CURRENT OPINION IN PSYCHOLOGY, 58(4):101838. https://www.sciencedirect.com/science/article/pii/S2352250X24000514 (last visited Jul. 5, 2026).

標的，要求必須並確保能做到權利不被侵害或受保障[13]。反之則為軟法治理，以美國作為代表性國家，主張避免扼殺創新，並強調市場機制與企業自主，則以準則、原則等方式訂立最低規範[14]。由此可看出兩者間差異巨大，而為何會造成天壤之別，除本身思維外，更可觀察各國法律上的緣由歷程，以知曉各國政府於立法或監管時之考量。

（一） AI 屬性認定

AI 管制之前提，視當事之國家如何認定 AI 之屬性，並針對該屬性出發立法或監管，從而衍生相關法制。在牛津字典中，AI 是能夠執行通常需要人類智慧任務的電腦系統[15]，並基於「系統、數據、選擇行動」的流程運作[16]。有些國家則肯認其屬於「工具」的一環，同時認爲其產出之結果，將足以影響人類社會。

而在我國當中，根據金管會頒布《金融業運用人工智慧（AI）指引》[17]，將 AI 定義爲系統之一部分，同時也訂義生成式 AI，其為生成模擬人類創造系統之AI工具，由此可見，生成式AI應屬於AI系統下的子集合。

雖本文主軸為生成式AI，然一般AI系統之生命週期與生成式AI之生命週期基本並無不同，分別僅存於後續輸出之結果，屬於生成式或單純分析之結果。所謂 AI 系統之生命週期，為初步構想、資料搜集與處理、模型訓練、部署與監控等四大方面[18]。故於一般 AI 於前階段生命週期所遇到的監管治理問題，生成式 AI 理應亦受管制之，且在後續部署與監控部分，需探討除一般AI輸出情況外之個案。

由上述可知，AI 之輸出，會影響物理或虛擬環境，當此技術影響到環境時，亟需監管治理。各國雖對於 AI 之定義上差異不大，卻因個別國家之管制出發點有所不同，進而衍生不同管制模式，最根本的價值選擇決定未來的監管方向[19]，在這光譜的兩端中，分別是代表創新的自由軟法，與代表安全的拘束硬法。

於歐盟當中，雖也將其描述為基於機器之系統，然而更加強調的，是後續輸出的交互性與影響力，歐盟認為 AI 得以針對人類的定義目標，從而產出足以影響其交互環境的結果，不論是內容、預測、建議等，均會

[13] 林勤富，歐盟《人工智慧法》之制度設計、 規範內涵與治理侷限，中研院法學期刊，第 36 期，頁 12，2025 年。
[14] 林勤富，同前註。
[15] Oxford English Dictionary, *Software Used to Perform Tasks or Produce Output Previously Thought to Require Human Intelligence*. https://www.oed.com/dictionary/artificial-intelligence_n?tab=meaning_and_use#38531565 (last visited Jul. 5, 2026).
[16] Tambiama Madiega, *Artificial Intelligence Act.* European Parliamentary Research Service, Sep. 2024, https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf (last visited Jul. 5, 2026).
[17] 金管會，註 4。
[18] 金管會，同前註。
[19] *Id.*

產生影響[20]。而美國則將其定位為基於機器之系統，本質上屬於運算邏輯與機器規則的集合，仍屬於工具之地位，是運用機器或人類的輸入，進行感知並推理，最後發展行動方案之工具[21]。

在安全拘束為主的政府當中，歐盟處於領導之地位，更關注基本人權問題，高度重視並要求企業遵守法律，故制定嚴格的法律規範。相對的在自由經濟為主的政府中，美國處於領導地位，更關注經濟與市場調節，避免扼殺創新，故期望以軟法管制。

根據金管會公布之《金融業運用人工智慧(AI)之核心原則及政策》，當中第二點闡明，「強調金融機構在運用 AI 系統時，應儘可能避免演算法之偏見所造成的不公平，並符合以人為本及人類可控的原則。」故在討論管制議題，不妨由基本人權的角度出發，此為人權保障最低度的內涵，也是不容剝奪之處，進而衍生管制可能侵害基本人權之事務的正當性。參考歐洲議會研究服務（European Parliamentary Research Service, EPRS）所公布之資訊[22]，基於歐盟的基本權利憲章保障，同時重視各種基本人權。而以上人權保障落實，以下逐一分析。

1. 個資與隱私權

基本人權當中的隱私權做為平衡國家權力與私人權利之關鍵，進而要求管制有關隱私相關行為，當中自然包含 AI 訓練。而 AI 在訓練過程需大量資料，於對個資依賴型的 AI 訓練時不免觸及個資相關議題，如何劃下限制，不僅是保障基本人權，更是隱私權的落實方式。

具體來說，對於隱私權落實方式，可能遇見的問題有三，其一為資料蒐集困境、資料處理、資料利用三大方面。然而，隱私利益的本質，在於識別化[23]，該特定資料有無識別回特定當事人之可能，將成為關鍵。如一個資料並不能連接至特定個人，僅能認為其屬於一段資料，也可能符合個資法規範。故於資料蒐集困境中，並非全然拒絕利用，應視不同程度的蒐集利用方式探討。

---

[20] *Id.*

[21] Miklós Király, *Survival of the Fittest? AI and Its Different Regulatory Models*, 39 EMORY INT'L L. REV. 473, 969 (2025). "A machine-based system that, for a given set of human-defined goals, can make predictions, recommendations, or decisions that affect real or virtual environments."

[22] *See* Tambiama Madiega, *supra* note 16. 在引言部分，提及 AI 技術預計帶來的經濟和社會效益，為因應變化歐洲議會得出的結論為：「應由道德為本來制定法律框架」，故由歐盟的「基本權利憲章」所載應被保障的基本權利，包括但不限於不被歧視權、言論自由權、人格權、個資權和隱私權等，均須被重視，進而衍生不同風險等級應對。

[23] 憲法法庭 111 年憲判字第 13 號，第 35 段，「個資若經處理，依其資料型態與資料本質，客觀上仍有還原而間接識別當事人之可能時，無論還原識別之方法難易，若以特定方法還原而可間接識別該個人者，其仍屬個資。當事人就此類資料之自主控制權，仍受憲法資訊隱私權之保障。反之，經處理之資料於客觀上無還原識別個人之可能時，即已喪失個資之本質，當事人就該資訊自不再受憲法第 22 條個人資訊隱私權之保障。」

2. 不被歧視權

對於歧視上，可區分直接歧視與間接歧視，所謂的直接歧視，係指直接以歧視性指標作為差別待遇之事由，例如限定性別、種族等，所謂間接歧視，為採取一結構性的規範，但對特定群體有不同之社會意涵，而該意涵意味著不法之差別待遇或不利益[24]。然不歧視並非係指必須採取齊頭式平等，而係應為合理化之差別待遇，從事之活動或所需之條件、所列事由係屬重要與關鍵之職業資格、比例原則與合法目的性，才謂之平等。

可見於輸入資料本身是潛藏偏差的資料，抑或是演算法本身即具有不同權重等，進而影響機器輸出結果。AI 不見得屬於完全公平之客觀決定，其輸出取決於學習時蒐集之資料，亦有研究指出，AI 的自動化決策選才決定，也會發生就業歧視，因為人類提供 AI 學習的訓練資料中，就包含「歧視因素」[25]。

3. 知情權與黑盒

知情權是使人民知曉資訊的權利，而運用 AI 下的黑盒問題，有可能對此造成相當程度之侵害。申言之，雖數學算式是公正客觀之數據，遵循一定規則運行，無隨機的產生，無「主觀層面」介入，因之有部分人認為，AI 的介入得以彌平人類的歧視行為[26]，然由於上述提及到 AI 的產出結果，很大程度上取決於所接受到之資訊，以及演算法運算過程，而只視最終結果卻不知其過程為何，故仍有所謂黑盒問題[27]。

在對人之決策並直接對人發生效力之情形下，有可能因黑盒問題使其難以信服，甚至侵害基本人權。換言之，單獨 AI 在資訊蒐集與後續演算法撰寫上難以全數概括，使前期資料蒐集有所缺失，而無法完整說明，進而衍生不透明的黑盒問題。故同理，應用於個人上之 AI，應謹慎思考公開權與知情權的平衡原則，進而使顧客知曉其中的理由。

（二） 法律爭議小結

綜上說明，AI 所觸及之法律問題，大多數由基本人權出發，結合 AI 運作模式，顯現於資料蒐集與演算法之間，例如前期預訓練之資料來源，

[24] 陳靜慧，歐洲人權法院及歐洲法院對於間接歧視概念之適用與實踐，憲法解釋之理論與實務，第九輯，頁 394-395，2017 年 4 月。

[25] Michael Witteler & Lucas Moll, *Künstliche Intelligenz am Arbeitsplatz – Datenschutz und Rechte des Betriebsrats*, 2023 NZA 327.

[26] 許慧瑩，AI 偏差歧見的另類難題–使用者對 AI 風險評估工具的差異性互動，臺灣人工智慧行動網：https://ai.iias.sinica.edu.tw/how-risk-assessment-tools-influence-final-decision-exp/（最後瀏覽日：2026 年 7 月 5 日）。

[27] 范姜真媺，AI 系統運用與基本人權之衝突及治理，司法院大法官 111 年度學術研討會-「數位時代網路與憲法基本權保障」，憲法法庭：
https://cons.judicial.gov.tw/docdata.aspx?fid=5364&id=345738（最後瀏覽日：2026 年 7 月 5 日）。

是否符合隱私權相關規範，是否本就包含歧視資訊等，及在演算法運作過程中，是否得透明化公布資訊、符合人權規範等。由此可見，AI 影響社會方方面面，不單僅有科技進步的影響，是的確有必要進行管制。惟更深層次的問題在於，如何在確保管制下亦平衡創新動力，對此本研究認為可參考各國 AI 監管模式作為借鑑。

二、 監管模式

（一） 嚴格監管：歐盟

於 2024 年，歐盟通過並發布 AI Act，開啟 AI 管制明文規定的濫觴，而在這之前，由《Ethics Guidelines for Trustworthy AI》、《General Data Protection Regulation》（下稱 GDPR）保護原則等歐盟所訂立之規範可稽，歐盟極可能採取嚴格監管模式，並以區分風險的方式分別規範。

觀察歐盟 AI Act 的立法緣由，或許會認為係由基本人權出發，然而如再觀察其 AI Act 之立法即可知，歐盟並非直接賦予人權，而是以區分風險方式，試圖對每一種 AI 功能立法[28]，針對的是基於風險的管理架構。比起直接賦予權利後再行提出救濟，歐盟反係利用更為詳細的規範風險，規定何可為、何不可為，並以「造成損害的機率與損害嚴重程度的結合」作為風險之定義，以此區別四大風險類別，準此，整體而言，AI Act 係以針對每一個 AI 風險的監管框架[29]，進而逐一規範並保障基本人權。

1. 歐盟法規立法過程

歐盟之立法，需優先考量到歐盟的價值觀與基本權利，同時也思考身為領導者的身份，如何引領歐盟以外的其他國家，以及立法後的布魯塞爾效應（Brussels Effect），故應基於此設立準則[30]。而在實際立法過程中，歐盟乃採用風險管控方式立法，運用不同的風險掌控疏密程度，逐一放寬或限縮各 AI 運用場景。

觀察歐盟所提出之 AI 規範歷程，學者提出，始於 2018 年成立之高級人工智慧專家小組，後提出人工智慧白皮書，甫成 AI Act。由此歷程可知，最終的 AI Act 來源於前期法規的疊加，以及前期的智慧小組發想，後再建立人工智慧白皮書[31]。因此歐盟認爲，未來的 AI Act 規劃，應著重於使用

---

[28] Alice E. Keane, Carlos Ferran & Robert Garcia, *AI and the European Union Artificial Intelligence Act: An Overview*, 35 MIDWEST L.J.1 (2025).
[29] *Id*.
[30] European Commission, *Artificial Intelligence for Europe*, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018DC0237 (last visited Jul. 5, 2026). ‘As with any transformative technology, some AI applications may raise new ethical and legal questions, for example related to liability or potentially biased decision-making. The EU must therefore ensure that AI is developed and applied in an appropriate framework which promotes innovation and respects the Union's values and fundamental rights as well as ethical principles such as accountability and transparency. The EU is also well placed to lead this debate on the global stage. This is how the EU can make a difference – and be the champion of an approach to AI that benefits people and society as a whole.’
[31] 林勤富，註 13，頁 28。

者的角度出發，並兼顧隱私、言論自由等基本人權的保障，以及努力避免歧視問題產生[32]。如此可證於 AI Act 第一條[33]，明文顯示 AI 法規應以人為本，並基於《基本權利憲章》（Charter of Fundamental Rights of the European Union）所擁有的權利均應受保障。

2. 《基本權利憲章》

歐盟《基本權利憲章》具有法律約束力，其權源來自於《里斯本條約》（Treaty of Lisbon），屬於高於各國法律位階的「準憲法」[34]，如各國的國內法與上開基本權利憲章競合時，將具有優先適用的原則，故立法論上勢必得參考此基本權利憲章要求。該基本權利憲章提及六大項目，分別為尊嚴、自由、平等、團結、公民權、司法六大面向，故藉由更高位階的準憲法制度，要求歐盟法律必須明文化這些權利主體，並將其納入法條規範中。而實際上的作為，或許可參考過往立法歷程。

3. 網路規範

在 AI 尚未蓬勃發展前，上一個革新技術可視為網路技術，針對網路技術，歐盟亦對其進行立法。在快速變動的網路環境中，為了落實基本權利憲章等人權議題，更多的討論係基於個資隱私權的保護規範[35]，因大量資料處理，以及資料背後龐大的經濟價值，無一不驅動人性的貪婪，進而討論個資的重要性與影響，從而以GDPR作為網路訊息時代資料處理規則，當中最重要的，乃新興科技帶來的新型基本權利，例如被遺忘權等[36]，並訂下更詳細的規範，也由此而生不同的規範，例如境外傳輸原則禁止等。

4. 歐盟模式小結

如以隱私權角度為出發，不難看出歐盟模式背後的邏輯，仍是出於以人為本的思想，首重道德規範作為 AI 管理的核心價值，由先前的可信賴準則開始，發展 AI 信賴度，定義道德與價值為取向的框架，後公布人工

[32] 林勤富，同前註，頁 29。

[33] 歐盟，人工智慧法第一章第一條“ The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter, including democracy, the rule of law and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation. ”

[34] 陳顯武、連雋偉，從『歐盟憲法』至『里斯本條約』的歐盟人權保障初探──以『歐盟基本權利憲章』為重點，臺灣國際研究季刊，第 4 卷，第 1 期，頁 25-45，2025 年。

[35] 林勤富，註 13，頁 25。

[36] 林勤富，同前註，頁 26。

智慧白皮書，進一步確立未來應以人為本，最後頒布 AI Act，確立明確規範[37]。

由前述說明可知，歐盟在新型網路科技上，係基於基本權利憲章保障，以隱私權出發，進而管制公司等處理資料者作為，然而 AI 不同於資訊處理，僅資訊處理規範難以有效管控全部的議題，個資更偏向蒐集、保護與共享，AI 更需專注於偏見、透明度與社會影響[38]。故有 AI Act 後續立法，利用風險管理框架，將無明文定義的基本權利與其衍生議題，分散於不同的風險框架中保護，而非一刀兩斷，例如，於禁止侵犯基本人權的 AI，對於高風險 AI，評估並要求減少影響等，不難看出為了平衡基本人權與創新，歐盟所採取的基於風險治理模式[39]。（此處將於本研究第三章做詳細說明。）

（二） 軟法治理：美國

美國在 AI 監管方面，採取較為溫和態度，其監管主要透過聯邦政策、國家原則和商業領域的國際法構成[40]。與歐盟硬法規定不同，美國規範採取軟法規範為主，然此非謂全無規範，而是相較歐盟全面性的 AI Act，以及更為明確的規範制度而言，美國多數以原則、指導、準則等方式處理。雖美國處於科技上領先地位，惟為了確保創新、避免過度監管扼殺創意，不願意積極介入管制，採取較低限度監管模式 ，並期望企業自願式填補空白[41]。

循此，國家聯邦層次上，美國係利用原則性規範來約束大範圍的方向，並在各州的立法上給予權限，使各自能立法管制各州認為必須硬法管控之處，例如加州之規範透明度法規[42]，即係因認為 AI 生成的資料應被大眾知曉，為因應可能的假訊息，而定有強制附上浮水印等措施。換言之，比起歐盟上對下的風險管控，利用整體歐盟法規要求個別成員國遵守，美國更像由下到上的規範路徑[43]，由處於第一線的公司先行自我管控，而各州針對性立法，為鼓勵創新放於首位，確保美國在 AI 研發領域繼續保持領先地位[44]。

在美國法規歷程上，除了最初始的基本標準，更多的會因美國歷屆政府之不同，對於 AI 監管也有所不同，以現任川普政府和上任拜登政府為

[37] 陳奕璇，現階段歐盟與美國對人工智慧規範之比較分析，安全與情報研究，第九卷，第一期，頁 116，2026 年。
[38] Delaram Rezaeikhonakdar, *Artificial Intelligence in Healthcare Governance Compliance and Data Privacy Beyond Hipaa,* 29 QUINNIPIAC HEALTH L.J. 175 (2026).
[39] Heidi L. Frostestad, *AI Regulation in a ChatGPT Era: Cross-border Cooperation and Hope in a Sudden Storm*, 32 IND. J. GLOB. LEGAL STUD. 1 (2024).
[40] *Id.*
[41] *See* Delaram Rezaeikhonakdar, *supra* note 38, at 184.
[42] 陳奕璇，註 37，頁 127。
[43] *See* Frostestad. *supra* note 39, at 18.
[44] *See* Miklós Király, *supro* npte 21*,* at 969.

例，一者以國家利益導向為主，另一者以人文倫理導向為主[45]。美國於不同政府時期，所應對之方式也有所不同，目前川普政府傾向於美國本位主義，並採取寬鬆的 AI 管制政策，試圖在 AI 浪潮中佔據領頭地位[46]。

1. 基本標準

美國於 2016 年時，針對 AI 未來可能的治理需求，認為既有法規已可面對 AI 衝擊，為了保護創新而不應制定規範限制，但為了基本的安全與倫理，仍應建立基準以提供指引[47]。故美國認為並非不重視倫理道德，而是已有足夠規範限制，再為此進行限制顯有管制過當疑慮，仍應回歸市場機制運作，並提供最低限度的運作指引。

2. 拜登政府時之行政命令

在前任美國總統拜登任內，頒布行政命令指出[48]，對於 AI 監管系統提出演練模式，就安全性與可靠性都需要進行評估。就人力資源方面，大力發展人力資源，以擴大技術人才掌握。就人民方面，試圖解決應用上之歧視與不平等，在國際方面，要求國際合作之重要性[49]。因此，該命令彰顯第一次的政府政策轉變，並強調在科技變革的時代，消費者保護更值得受到重視[50]。

由此可見，拜登相較 2016 的川普政府而言，更為重視人民權利面向，然而，此監管思維仍趨近業者自律[51]，因其所做的各種要求，仍限於企業自我要求拘束性質，也與歐盟不同，並未明文列舉何種方式使用為禁止，故可視為折衷的辦法。

3. AI 權利法案藍圖與 AI 風險框架

在 2023 年時，美國政府提出兩項指標，分別為《AI 權利法案藍圖》（Blueprint for an AI Bill of Rights）和《AI 風險管理框架》(Artificial Intelligence Risk Management Framework, 下稱 AI RMF 1.0)。兩者均屬自發性接受性質。前者是安全、歧視、隱私、告知、選擇五大公民權利[52]，後

[45] 曾憲立，公部門人工智慧規範架構結案報告，數位治理研究中心：https://www.teg.org.tw/researches/69（最後瀏覽日：2026 年 7 月 5 日）。
[46] Rezaeikhonakdar. *supra* note 38, at 186.
[47] 林勤富，註 13，頁 12。
[48] 曾憲立，註 45。
[49] 文逢遠，美國人工智慧行政命令之新舊比較與可能影響，經貿法訊第 343 期，政大商學院國際經貿組織暨法律研究中心，頁 3，2025 年。
[50] *See* Miklós Király, *supra* note 21, at 981.
[51] 林勤富，註 13，頁 15。
[52] 許向華，美國人工智慧權利法案藍圖，科技發展觀測平台，2022 年 https://outlook.stpi.niar.org.tw/index/tdop/4b11410082b117250183a5d5bfd020e5（最後瀏覽日：2026 年 7 月 5 日）。

者針對風險性、可解釋性、隱私權、偏見等進行強調，認為發展 AI 的同時須遵守上述規則[53]。由此可見，仍屬於自願性質的大範圍框架。

4. 川普政府行政命令

美國總統川普第二次上任後公布多項行政命令，極力消除對 AI 發展的阻礙，例如公布美國 AI 行動計劃、重建顧問委員會等。根據其行政命令[54]，要求可能阻礙創新之政策、法規等均應被撤銷，同時認為現有反歧視法律即可應對歧視問題，也因此並不像歐盟規劃以人為本之方向，最後要求美國必須處於 AI 之領導地位，故現行法規傾向以聯邦政策優先，較低限度管控，為第二次的政策轉變。

5. 各州自主思維

雖聯邦政府層級未積極立法，然於州政府層級，仍有部分硬法管控，例如針對性犯罪、意圖引導選舉、保障名譽等[55]，由此可見，對於部分議題，特別是現行法律已有規定之罪，有部分州對此進行立法，惟並未聯邦性一體適用。然而，消除阻礙並非全面無視風險，仍會針對最低風險管控措施，故美國各地設立 Chief Artificial Intelligence Officers（CAIO），作為首席 AI 官員，落實風險分析，並定期公告檢驗報告，且各機關首長得自行決定是否終止使用[56]。

可見在州層級上，仍具有部分自主權利的彈性，比起擴大範圍拘束的監管治理，更小範圍的立法或許具有更多的彈性，但也引發是否足夠安全之隱憂。

6. 美國模式小結

綜上，美國在立法大方向上，採取規範性框架之立法，於此框架下，再交由各州自主決定。整體而言，美國之立法仍屬於以科技發展為主軸，相較於歐盟區分更多層次風險，美國認為足以面對未來風險，從而放任自由經濟市場發展，進而獲取更多利益。然而，除聯邦層級外，各州的立法亦有所不同，但各州立法時所強調之重點仍待探討。（此處將於本研究第三章詳細說明。）

[53] 劉汶渝，人工智慧風險管理框架(AI RMF)，臺灣人工智慧行動網，2024 年 https://ai.iias.sinica.edu.tw/ai-risk-management-framework.（最後瀏覽日：2026 年 7 月 5 日）。
[54] 曾憲立，註 45。
[55] 陳奕璇，註 37，頁 127。
[56] 陳奕璇，同前註，頁 129。

第四節　研究方法與架構

一、　研究方法

本研究將採用以下兩種分析方式：

1. 文獻分析法

彙整美國、歐盟 AI Act 及新加坡等地 AI 監管制度，並與臺灣既有金融相關法規（銀行法、個資法等）進行交叉比對。分析金融監理機構與銀行業者的公開指引與風險治理文獻資料。

2. 案例研究法

選擇銀行或 Fintech 應用生成式 AI 的實際案例（如應用生成式 AI 客服等），進行剖析、合規風險與監理對應策略。

二、　研究架構

本研究共分為七章。於第一章緒論時先進行概要說明，講述本研究動機、目的、問題意識，並進行文獻分析與回顧。第二章生成式 AI 在金融機構之應用現況，開始盤點現行金融機構，如何使用 AI 運行於系統內，包含國內外各大金融機構現況。第三章歐盟 AI 監管法制，論及目前歐盟對於 AI 監管的現況詳細說明，分別闡述歐盟之監管結構。第四章美國 AI 監管法制，區分聯邦層級與州層級審視，研析美國 AI 監管結構。第五章新加坡 AI 監管法制，彙整該國 AI 治理原則、實務手冊以及檢測工具，彙整出新加坡縝密的治理架構。第六章臺灣金融產業現況與挑戰，由《人工智慧基本法》開始，再論述金融業專屬規範與其餘規範，並分析現行法規落差與實務可能之困境和國際法制之比較。第七章結論與建議，將總結本研究成果與提出後續監管建議方向。

第二章　生成式 AI 在臺灣金融機構之應用現況

依金管會近期對金融業者及其周邊單位所進行之調查結果顯示，臺灣金融機構對AI之採用程度已呈現明顯提升趨勢，其中以銀行業之使用比例最高，約達87%，其次為壽險業與產險業[57]。進一步就生成式AI而言，目前已有約48%之金融業者實際導入相關技術，較前一年度顯著成長，惟其應用仍主要集中於內部行政作業與智能客服等相對低風險場域，且多數業者之自動化決策比例仍維持在低度水準[58]。可見金融機構對生成式AI之採用，整體仍採審慎漸進策略，亦顯現行監理制度與風險治理架構在生成式AI發展過程中所扮演之角色。

第一節　生成式 AI 之技術基礎與功能特性

近年來，生成式 AI 隨著大型語言模型之快速發展，已成為 AI 應用中最具突破性與產業影響力之技術。相較於傳統以分類、預測或規則比對為核心之 AI 系統，生成式 AI 強調對資料分布與語意結構之學習，能夠依據輸入脈絡生成具連貫性與可理解性之新內容，廣泛應用於文字、圖像、語音與程式碼等多元場景。金融產業因其高度資訊密集、法規複雜且業務流程標準化程度高，成為生成式 AI 最早且最積極導入之產業之一。 本節首先說明生成式 AI 之基本技術原理，繼而整理其於金融場域中展現之關鍵功能特性，作為後續探討金融機構實際應用情境與監理影響之理論基礎。

一、　生成式 AI 之技術原理概述

生成式 AI 係指一類以學習資料機率分布為目的之 AI 模型，其核心在於透過大量訓練資料，掌握語言、圖像或其他資料型態之潛在結構，並在推論階段生成新的、具語意合理性之輸出內容[59]。當前金融領域所應用之生成式 AI，主要以大型語言模型為核心，其技術基礎多建立於 Transformer 架構之上[60]。

Transformer 架構透過自注意力機制（Self-Attention Mechanism），使模型能夠同時考量輸入序列中不同位置之資訊關聯，突破過往循序式模

[57] 金融監督管理委員會，註 4。

[58] 金融監督管理委員會，註 4。

[59] Ian J. Goodfellow et al., *Generative Adversarial Nets*, *in* 27 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 2672 (Z. Ghahramani et al. eds., 2014), https://papers.nips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html(last visited Jul. 5, 2026).

[60] Tom B. Brown et al., *Language Models Are Few-Shot Learners*, *in* 33 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 1877 (Hugo Larochelle et al. eds., 2020).

型在長距離語意理解上的限制[61]。使模型得以有效理解金融文本中高度專業且結構複雜之內容，例如法規條文、契約文件、財務報表與研究報告等[62]。

大型語言模型通常採用預訓練加微調（Pre-training and Fine-tuning）之方式建構，其係模型先以大規模通用語料進行預訓練，建立基本語言理解能力，再透過金融領域資料進行微調，使其輸出內容更貼近特定產業語境[63]。近年亦廣泛採用檢索增強生成（Retrieval-Augmented Generation, RAG）技術，於生成過程中即時結合外部資料庫或內部知識庫，以補足模型知識時效性不足之問題，並降低生成內容失真或幻覺（Hallucination）之風險，對於法遵審查、監理報告與高風險決策輔助尤為重要[64]。另，生成式AI在金融領域之技術基礎，並非僅止於文字生成，而是結合語意理解、資料檢索與脈絡推論之複合式智能系統，為金融業務流程之重構奠定技術前提[65]。

二、 生成式 AI 之發展與演進

（一） 大語言模型與基礎模型之興起

自 2018 年以來，以 BERT、GPT 系列模型為代表之大語言模型，透過在大規模語料上進行預訓練，逐步展現出超越傳統任務導向式 AI 系統之語言理解與生成能力[66]。尤以 2020 年問世之 GPT-3 為例，其具備約一千七百五十億參數，已能生成具備高度語意連貫性之自然語言文本，並廣泛應用於文章撰寫、問答回應及程式碼生成等多元情境[67]。而相較於過往金融機構所採用之 AI 系統，多半侷限於特定功能或明確定義之業務任務（如詐欺偵測或風險評分），大語言模型係建立於「基礎模型

---

[61] Ashish Vaswani et al., *Attention Is All You Need*, *in* ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30, 5998 (2017).
[62] *Id.*
[63] Edward J. Hu et al., *LoRA: Low-Rank Adaptation of Large Language Models*, ARXIV:2106.09685 (2021), https://arxiv.org/abs/2106.09685.
[64] Gautier Izacard et al., *Unifying Large Language Models and Retrieval for Knowledge-Intensive Tasks*, ARXIV:2206.05235 (2022), https://arxiv.org/abs/2206.05235.
[65] Patrick Lewis et al., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, *in* ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 33, 9459 (2020).
[66] 參考 OpenAI 所提出之 AGI（Artificial General Intelligence）發展階段模型，其演進歷程可區分為五個層級。第一階段為「對話型 AI（Conversational AI）」，主要功能在於模擬人類語言互動與回應能力。第二階段為「理性思考者（Reasoners）」，除具備對話能力外，亦能進行邏輯推理與複雜問題分析。第三階段為「自主代理人（Agents）」，此階段之 AI 不僅能理解與推理，尚可自主規劃並執行多步驟任務，具備持續管理與完成工作流程之能力。第四階段為「創新者（Innovators）」，AI 將超越既有問題解決框架，能主動提出新構想、設計全新解決方案，甚至開發創新的產品或流程。第五階段則為「組織型 AI（Organizations）」，屬於高度整合之終極型態，AI 能統籌策略規劃、資源配置與日常營運管理，於組織層級發揮決策與執行功能。陳曉莉，OpenAI 把 AI 程度分為 5 級，最高等級將可處理整個組織的事務，iThome: https://www.ithome.com.tw/news/163913; Tiernan Ray, *OpenAI's Gigantic GPT-3 Hints at the Limits of Language Models for AI*, ZDNET, https://www.zdnet.com/article/openais-gigantic-gpt-3-hints-at-the-limits-of-language-models-for-ai/ (last visited Jul. 5, 2026).
[67] *Id.*

（Foundation Model）」之設計思維上，其核心特徵在於先於巨量通用資料中進行預訓練，後續再依不同應用場景進行微調或提示設計，因而具備高度通用性與可遷移性[68]。

在金融領域之初期應用上，金融機構多以直接採用開源大型語言模型，或透過商用 API（Application Programming Interface, API）將通用模型導入既有工作流程，例如財務報表分析、投資問題回應等[69]。然而，由於金融語言涉及高度專業之術語體系、制度背景與語境判斷，通用模型於此類任務中往往呈現理解不精確或語意偏差之情形，顯示單純移植通用模型難以充分回應金融場域之專業需求[70]。

基於前述限制，實務界遂逐漸轉向發展金融專屬之大型語言模型，以強化模型對金融語境之掌握能力。其中，Bloomberg 公司於 2023 年推出之 BloombergGPT，該模型擁有約五百億參數，並以 Bloomberg 長期累積之金融資料庫（約三千六百三十億詞彙）作為主要訓練來源，輔以一般通用語料進行混合訓練[71]。相關研究結果顯示，BloombergGPT 於多項金融專用任務中之表現，明顯優於一般通用模型，同時亦能維持一定程度之通用語言處理能力[72]。其凸顯「領域適應（Domain Adaptation）」於生成式 AI 發展的重要性，透過在金融專屬資料上進行預訓練或後續微調，大語言模型得以更精確地掌握財經語言中隱含之制度語境、專業語義與判斷邏輯，從而提升其於金融決策支援、法遵審查及專業分析等高風險場域中的實用性與可靠性[73]。

（二） 生成式 AI 技術的成熟與關鍵進展

除模型參數規模擴張與訓練資料量增加外，近年生成式 AI 之實際可用性與穩定性，亦有賴於多項技術之成熟與整合[74]：

1. 透過人類回饋機制，其輸出行為逐步被納入可控與可預期的範圍

OpenAI 在 ChatGPT 中引入人類反饋強化學習（Reinforcement Learning from Human Feedback, RLHF），透過人類評估者對模型輸出進行

---

[68] H. Suresh & John V. Guttag*, A Framework for Understanding Unintended Consequences of Machine Learning, ARXIV:1901.10002 (2019),* https://arxiv.org/abs/1901.10002; Edward J. Hu et al., *supra* note 63.

[69] Carlo Giovine, Larry Lerner, *One Year In: Lessons Learned in Scaling Up Generative AI for Financial Services*, MCKINSEY & CO. (2024), https://www.mckinsey.com/industries/financial-services/our-insights/banking-matters/one-year-in-lessons-learned-in-scaling-up-generative-ai-for-financial-services.

[70] Shayne Longpre et al., *BloombergGPT: A Large Language Model for Finance*, ARXIV:2303.17564 (2023), https://arxiv.org/abs/2303.17564.

[71] *Id.*

[72] *Id.*

[73] International Monetary Fund, *Generative Artificial Intelligence in Finance: Risk Considerations*, FINTECH NOTE NO. 2023/006, 2023, https://www.imf.org/en/Publications/fintech-notes/Issues/2023/08/18/Generative-Artificial-Intelligence-in-Finance-Risk-Considerations-537570.

[74] 李宜熹、王友珊，生成式 AI 在金融領域的技術演進與挑戰，期貨人，第 93 卷，頁 9，2025 年。

回饋與偏好標註，使模型得以在訓練過程中調整其回應策略，所生成內容更符合社會期待與使用情境，並降低不當或偏離任務目的之輸出風險[75]。

2. 藉由提示設計與工具整合，得以嵌入既有業務流程運作

提示設計（Prompt Engineering）與外部工具整合，讓使用者得以透過結構化指令引導模型產出符合特定需求之內容，同時，模型亦可結合外部資料庫、檢索系統或計算工具，於生成過程中即時取得外部資訊，以支援較為複雜或具脈絡依賴之任務[76]。

（三） 多模態模型的發展，生成式 AI 具備跨資訊型態之綜合理解能力

隨著多模態（Multimodal）模型之發展，生成式 AI 已逐漸突破單一資料型態之限制，透過同時處理文字、圖像與其他感測資料，模型得以建立跨模態之關聯理解能力[77]。例如，結合財務文本與圖表資料之模型，能同時解讀財報說明與視覺化資訊，進而形成更完整之分析判斷[78]。

（四） 模型輕量化與私有化部署，回應金融機構對資料治理與資安合規之需求

考量大型語言模型對運算資源與資料治理之高度需求，近年亦出現以模型輕量化與隱私保護為核心之技術發展方向，透過參數高效微調、知識蒸餾與量化等方法，金融機構得以在維持一定生成品質之前提下，部署較小規模之模型於內部環境中運作[79]。實務上，已有不少金融機構基於資訊安全考量，限制員工直接使用外部生成式 AI 服務，轉而採取內部模型建置或私有化部署之策略[80]。

---

[75] Hugging Face, *Illustrating Reinforcement Learning from Human Feedback (RLHF)* (2022/12/9), https://huggingface.co/blog/rlhf. (last visited Jul. 5, 2026).

[76] Tom B. Brown et al., *Language Models are Few-Shot Learners*, ARXIV:2005.14165 (2020), https://arxiv.org/abs/2005.14165.

[77] Alec Radford et al., *Learning Transferable Visual Models from Natural Language Supervision*, in PROCEEDINGS OF THE 38TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML 2021), https://arxiv.org/abs/2103.00020; Alexey Dosovitskiy et al., *An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale*, in INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR 2021), https://arxiv.org/abs/2010.11929.

[78] 多模態模型常見於數位開戶與 KYC 流程，系統可同時處理客戶上傳之身分證件、表單影像、自拍照或自拍影片（活體偵測），並結合對話介面引導完成驗證；另於文件自動化場景，亦可利用具視覺能力之模型直接理解含版面與表格之 PDF 文件，進行欄位抽取、分類與摘要，以提升作業效率與一致性。例如，金融機構可透過 Amazon Bedrock 所提供之大型語言模型建構對話式數位助理，於線上開戶流程中引導使用者上傳身分證明文件與相關資料，並結合文件理解與生成式對話能力，自動完成欄位抽取、資料驗證與流程指引。參照：AWS, *Automate User On-Boarding for Financial Services with a Digital Assistant Powered by Amazon Bedrock*, AWS MACHINE LEARNING BLOG (2023), https://aws.amazon.com/blogs/machine-learning/automate-user-on-boarding-for-financial-services-with-a-digital-assistant-powered-by-amazon-bedrock.

[79] Edward J. Hu et al., *LoRA: Low-Rank Adaptation of Large Language Models*, ARXIV:2106.09685 (2021), https://arxiv.org/abs/2106.09685.

[80] 如 Morgan Stanley Wealth Management 強調其解決方案「不是」直接使用公開 ChatGPT，而是以 GPT-4 在內部內容來源上生成回覆，並搭配控制措施以降低外洩與不當引用風險。參照：OpenAI, *Morgan Stanley Uses AI Evals to Shape the Future of Financial Services*, https://openai.com/index/morgan-stanley/ (last visited Jul. 5, 2026).

三、 生成式 AI 於金融應用之核心功能特性

相較於既有 AI 技術，生成式 AI 在金融機構實際應用中具有下列特性，使其不僅成為提升作業效率之工具，更逐步影響金融服務設計、風險管理與監理模式之運作邏輯[81]。

（一） 自然語言理解與生成能力，降低非結構化資訊處理成本

金融機構長期面臨大量非結構化資訊處理問題，例如客戶對話紀錄、法規解釋文件、內部報告與市場新聞等，生成式 AI 可將此類資訊轉化為摘要、分析或可供決策參考之內容，顯著降低人工作業負擔[82]。

（二） 高度可擴充性與平台化特性

透過 API、企業內部平台或 Copilot 形式，生成式 AI 可嵌入既有資訊系統與業務流程中，成為橫跨多部門之通用輔助工具，使金融機構得以逐步推動「AI 平台化」策略，而非僅限於單一專案或特定業務線應用[83]。

（三） 支援決策與知識輔助

透過整合歷史資料、即時資訊與專業規則，模型可提供分析建議、情境模擬與風險提示，協助專業人員聚焦於高價值判斷工作[84]。然而，生成式 AI 所產生之內容本質上仍屬機率性推論結果，並非具法律或財務確定性之結論，若未妥善設計人機協作機制，將導致錯誤決策風險[85]。

四、 生成式 AI 應用之新興風險

相較於傳統 AI 多聚焦於分類、預測、偵測或評分等「輸出結果」型決策支援，生成式 AI 與進一步發展之 AI 代理／代理型 AI 所帶來的變化，在於其不僅能生成內容，亦可能規劃並採取行動，使風險型態由單純的資訊層次，擴張至實際業務流程與交易行為層次[86]。此一轉變使金融監理所面對的風險，不再僅限於模型準確性或偏誤問題，而呈現出更高度的行為性、動態性與跨系統性特徵，以下分為幾個面向[87]。

（一） 內容生成與資訊可信度

生成式 AI 出現之「幻覺」或錯誤推論，且其輸出在語氣與結構上高度擬真，若應用於客服回應、理財機器人、法遵解釋或金融商品資訊揭

---

[81] International Monetary Fund, *supra note 73.*
[82] Fabrizio Dell'Acqua et al., *supra* note 10.
[83] McKinsey & Co., *supra* note 69 .
[84] Ross P. Buckley et al., *Regulating Artificial Intelligence in Finance: Putting the Human in the Loop*, 43 SYDNEY L. REV. (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3831758.
[85] International Monetary Fund, *supra* note 73.
[86] 許正乾、陳家駿，人工智慧最新應用及其法律議題系列 3──AI 代理 v.代理 AI，科技產業資訊室：https://iknow.stpi.niar.org.tw/post/Read.aspx?PostID=22296 (最後瀏覽日：2026 年 7 月 5 日)。
[87] Financial Stability Board (FSB), *The Financial Stability Implications of Artificial Intelligence* https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/. (last visited Jul. 5, 2026).

露，恐使消費者難以辨識其正確性，進而影響投資判斷或權益保障，並衍生不實或誤導資訊之責任爭議[88]。

（二） 歧視性風險與偏誤

若生成式 AI 所依據之訓練資料含有結構性偏誤，或模型於資訊不足之情境下仍作出確信度高之判斷，則其輸出結果將造成定價不公平、決策標準不一致，應用於保險業中，甚至會引發過度審查或調查等問題[89]。在缺乏完善驗證機制與人類監督框架下，金融機構若過度依賴此類系統，將對消費者信賴與機構聲譽產生負面影響[90]。

（三） 決策透明度與責任歸屬

生成式模型與代理型系統多具高度黑盒性與可解釋性不足之特徵，且其行為受到提示、上下文、工具回傳結果等多重因素影響，使金融機構難以清楚說明特定輸出或行動之依據與決策邏輯[91]。當 AI 系統進一步具備安排優先順序、持續迭代執行或跨系統調用工具之能力時，人類對其行為結果之可預期性將隨之降低，進而提高內外部稽核、行為留痕、問責與責任歸屬之治理難度[92]。

（四） 行動性、自動化與資安保護

AI 代理與代理型 AI 能主動調用內外部系統以完成任務，若其規劃或工具調用出錯，可導致錯誤資料寫入、流程誤觸發，甚至不當交易行為，且錯誤會因自動化流程而被迅速放大與擴散[93]。生成能力之普及亦降低深偽（Deepfake）、社交工程與偽造文件之技術門檻，擴大詐欺與資安攻擊面[94]。同時，隨著 RAG 架構與外部雲端模型服務之導入，金融機構在資

---

[88] 生成式 AI 的「幻覺（Hallucination）」指的是 AI 模型在回答時產生看似合理但實際錯誤或無根據的資訊，這類錯誤往往語氣自信且表面流暢，讓使用者不易察覺其虛假性。源於模型依據語言模式與機率計算產出文字，而非透過事實查證之結果。參照：生成式 AI 產生「幻覺」問題警惕虛假資訊風險，今周刊：https://www.businesstoday.com.tw/article/category/183015/post/202601200045/（最後瀏覽日：2026 年 7 月 5 日）。

[89] Dima Gritsenko, *Generative AI in Insurance: The Future of Claims – Proactive Problem-Solving and Fraud Detection at Your Fingertips*, MASTER OF CODE (2025), https://masterofcode.com/blog/generative-ai-in-insurance.

[90] *Id.*

[91] 「黑盒」一詞原本源自資訊控制理論，係指在分析某一系統時，不必揭示其內部結構與運作機制，而僅透過輸入與輸出之對應關係，推測其運行規律。應用於金融科技場景時，外界通常無法掌握演算法之設計架構、參數設定、模型權重或實際決策邏輯，亦無從了解交易執行之具體判斷過程，即便為模型設計者或開發者，亦未必能完整揭示其權重分配與推理路徑，其所能觀察者，僅為輸入資料與最終交易結果之對應。參照：EWSolutions, *Understanding Black Box AI*, https://www.ewsolutions.com/understanding-black-box-ai/ (last visited Jul. 5, 2026).

[92] 劉輝，大資料金融演算法的法律規制，財經理論與實踐，第 42 卷，第 2 期，頁 152-154，2021 年。

[93] 許正乾、陳家駿，註 86。

[94] FSB, *supra* note 87.

料傳輸與處理過程中，亦面臨更高之機敏資訊外洩、不當使用及第三方服務供應商風險管理壓力[95]。若 AI 代理工具得以代表使用者或機構發起行動並處理紀錄與任務，其在電子交易情境下甚至被類比為「電子代理人」（Electronic Agent），進一步牽動錯誤更正、撤銷機制與法律責任配置等問題[96]。

（五） 系統性風險與同步性效應

生成式 AI 技術在金融機構之廣泛採用時，當多家機構依賴相似之訓練資料來源、模型架構或第三方供應商時，其決策行為會有趨於同步，進而放大市場波動並加劇流動性緊縮之風險[97]。該現象係羊群效應、網路外部性、供應商高度集中，以及模型訓練資料來源缺乏透明揭露等因素所影響[98]。在交易、資產配置、信用評估與保險定價等領域，若多數機構使用高度相似之 AI 模型進行價格訊號辨識或風險校準，則於市場壓力情境下，會同時觸發類似之調整行為，形成順循環性（Pro-Cyclicality）與同步拋售風險，甚至提高閃崩發生之可能性[99]。而大型語言模型與生成式 AI 之底層架構高度相似，且可能依賴共同之網路資料進行訓練，亦加劇輸出結果之相關性[100]。

（六） 風險轉型與監管密度之調整

生成式 AI 與代理型 AI 所帶來之新興風險，不再僅限於模型準確性或資料偏誤控制，而須進一步關注輸出與行為之可追溯性、責任歸屬之明確性、人類監督之有效性，以及對外服務之資訊揭露義務。故其監管密度不宜僅停留於一般性風險管理原則，而應依應用情境與風險影響程度進行分級調整[101]。當生成式 AI 僅作為內部輔助工具時，監理強度可維持較低密度之原則型管理；惟當其涉及對外服務、交易決策或權益影響時，則有必要提高監理密度，強化查核標準與治理要求[102]。

---

[95] *Id.*

[96] Unif. Elec. Transactions Act (1999) § 2(6): "Electronic agent" means a computer program or an electronic or other automated means used independently to initiate an action or respond to electronic records or performances in whole or in part, without review or action by an individual.; *see* also: Emre Bayamlıoğlu, *Intelligent Agents and Their Legal Status*, 1 ANKARA BAR REV. 46 (2008), https://dergipark.org.tr/en/pub/abr/article/607011.

[97] Gary Gensler & Lily Bailey, *Deep Learning and Financial Stability*, SSRN (2020), https://ssrn.com/abstract=3723132.

[98] FSB*, supra* note 87, at 20. 關於金融業隻羊群效應，參照: Abhijit V. Banerjee, *A Simple Model of Herd Behavior*, 107 Q. J. ECON. 797 (1992).

[99] Gensler and Bailey, *supra* note 97; International Monetary Fund, *supra* note 73.

[100] International Monetary Fund, *supra* note 73; Jon Danielsson et al., *Artificial Intelligence and Systemic Risk*, 140 J. BANKING & FIN. 106290 (2022).

[101] 劉博涵，智慧投顧演算法透明的價值辯證、規制導向與實現路徑，深圳大學學報淵人文社會科學版，第 41 卷，第 1 期，頁 105，2024 年。

[102] 劉博涵，同前著。

生成式 AI 與代理型 AI 相較於傳統 AI 所引入之新興風險，核心在於內容生成能力與代理化行動能力的結合，使金融風險由「說錯話、給錯建議」進一步演變為「做錯事、動錯流程」。基此，金管會除延續既有以 AI 生命週期為核心之治理框架（如資料治理、模型驗證、持續監控與風險分級）外，尤有必要針對內容責任歸屬、輸出與行為之可追溯與可稽核性、對外服務之資訊揭露義務、資安反詐與深偽防制措施，以及第三方與雲端模型服務之控管機制，建立更具體且可操作之監理要求與查核標準，以回應生成式 AI 與代理型 AI 對金融監理所帶來之新型挑戰（請參照下表 1、各類型 AI 之比較）。

表 1、各類型 AI 之比較

風險	傳統 AI（規則/ML）	生成式 AI（LLM/RAG）	AI Agent/代理型 AI
風險本質	多集中在「輸出結果」是否準確、偏誤、歧視。	風險由「結果」延伸到「內容」。	風險由「說錯」進一步變成「做錯」。
內容可信度	可能因資料偏誤而錯判。	幻覺風險，且文字高度擬真，會放大不實誤導之資訊。	若將錯誤內容再去執行，錯誤會沿流程擴散、放大。
可解釋性	可用特徵重要性、規則追溯等方式。	黑盒效應更嚴重。	透過思維鏈，能說明其背後的數據依據、決策邏輯，提升信任度[103]。
資安風險	對抗樣本、模型竊取等風險[104]。	深偽、社交工程、偽造文件門檻下降，擴大詐欺與資安攻擊。	代理若被提示詞注入（Prompt Injection）可能使攻擊變成業務事件[105]。

[103] 思維鏈（Chain-of-Thought, CoT），係指一種提示工程（Prompt engineering）技術，透過引導大型語言模型逐步列出推理過程，使模型在回答問題前先生成中間推理步驟，有助於降低草率輸出之風險，並使推理步驟得以被外部觀察與審視，從而提升決策過程之可追蹤性與使用者信任感。然而，思維鏈所呈現之推理過程仍屬語言生成結果，並非模型內部權重機制之真正可解釋性。參照：許正乾、陳家駿，人工智慧最新應用及其法律議題 系列 1 -- 從生成式 AI 到 AI 代理，科技產業資訊室：https://iknow.stpi.niar.org.tw/Post/Read.aspx?PostID=22206（最後瀏覽日：2026 年 7 月 5 日）

[104] 對抗性攻擊（Adversarial attack），係指透過對輸入資料施加細微但刻意設計之擾動，使模型產生錯誤判斷或行為。參照：Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy, *Explaining and Harnessing Adversarial Examples*, ARXIV:1412.6572 (2015), https://arxiv.org/abs/1412.6572.

[105]「提示詞注入（Prompt injection）」係指攻擊者利用大型語言模型對自然語言指令之依循性，將惡意指令置入模型可取得之資料來源，誘使模型在工具調用或外部系統介接情境下執行非授權之資料存取或流程觸發。「ZombieAgent」漏洞即屬此類，即當 ChatGPT 透過 Connectors/Apps 連接 email/雲端服務後，攻擊者可於來信或檔案中藏入不可見提示，進而引導模型在「零點擊/一點擊」條件下進行資料外洩、植入記憶以取得持久性，甚至透過收件匣進行擴散。參照: Sead

資料外洩/第三方風險	主要在資料蒐集、模型訓練部署的治理。	RAG、雲端模型導入帶來資料傳輸處理時的外洩、不當使用、供應商風險壓力。	代理跨系統取資料、觸發後續作業，資料流更長、更難控，也更依賴外部工具與平台。
可稽核性/留痕	較聚焦模型版本、特徵、輸入輸出紀錄。	需要保存提示、上下文、檢索片段，才有可能重建輸出依據。	須記錄規劃步驟、工具呼叫、回傳結果、狀態變化，否則無法稽核整段行動鏈。
法律責任	為「模型錯誤導致決策錯誤」的責任歸屬問題。	內容若被認定為不實誤導資訊，會牽動揭露義務與責任分擔。	若為電子代理人，進一步牽動代理效力、撤銷機制與行為能力判定。

資料來源：本研究自行整理

第二節　生成式 AI 發展下 FinTech、RegTech 與 SupTech 之制度定位與互動關係

隨著 AI 技術，特別是生成式 AI 之快速演進，金融科技之發展已不再僅限於金融服務創新本身，而逐步延伸至法遵與監理層面，形成涵蓋金融機構、監理機關與科技供應商之多層次科技治理生態。在此脈絡下，金融科技（FinTech）、監理科技（RegTech）與監管科技（SupTech）成為理解生成式 AI 於金融體系中角色分工與制度影響之三項重要概念。

本節將分別說明 FinTech、RegTech 與 SupTech 之定義與發展背景，並進一步分析三者在生成式 AI 技術導入後之功能定位與互動關係，作為後續探討金融機構具體應用場域與監理影響之理論基礎。

一、　FinTech、RegTech 與 SupTech 之定義與發展脈絡

所謂金融科技，係指運用資訊科技與數位技術，提升金融服務效率、可及性與創新性之各類應用與商業模式，並對金融市場、金融機構及金融服務之提供產生重大影響[106]。其應用範圍涵蓋支付、放款、投資理財、保險、行銷與客戶服務等核心金融業務[107]。生成式 AI 納入 FinTech 架構

Fadilpašić, *This "ZombieAgent" Zero-Click Vulnerability Allows for Silent Account Takeover – Here's What We Know*, TECHRADAR (2026), https://www.techradar.com/pro/security/this-zombieagent-zero-click-vulnerability-allows-for-silent-account-takeover-heres-what-we-know

[106] Xavier Vives, *The Impact of FinTech on Banking*, 2 EUR. ECON. 97 (2017).

[107] FSB, *FinTech and market structure in financial services: Market developments and potential financial stability implications*,https://www.fsb.org/uploads/P140219.pdf (last visited Jul. 5, 2026).

後，使金融服務由既有的流程自動化與數據分析，邁向語意理解、內容生成與高度個人化之服務設計[108]。

監理科技則係指金融業者為因應法規遵循、內控制度與風險管理需求，所導入之科技工具與解決方案，其並非直接創造營收，而在於降低法遵成本、提升合規效率與減少人為錯誤[109]。傳統 RegTech 多以規則引擎、關鍵字比對或結構化資料檢核為主，惟隨著監理與合規情境中 AI 模型與風險判斷日益複雜，金融機構對「可解釋、可被稽核之輸出」需求提高，促使 RegTech 開始導入能將模型結果轉化為人類可理解敘述之技術與工具[110]。而面對法規文本複雜化與跨境監理要求增加，生成式 AI 的語意理解與文件摘要能力亦成為 RegTech 發展之推力[111]。

監管科技則係由金融監理機關所採用之科技工具，目的在於強化對金融市場與金融機構之監督能力[112]。SupTech 的應用範圍包括異常交易偵測、市場風險監控、監理申報自動化與政策分析等。相較於 FinTech 與 RegTech 主要由市場或業者端推動，SupTech 之發展更直接反映監理機關對於科技風險與系統性風險之治理需求[113]。

FinTech、RegTech 與 SupTech 雖各自起源於不同需求，但其共同特徵在於，皆以科技手段回應金融體系中效率、風險與治理之問題，且隨著生成式 AI 的導入，三者之界線逐漸產生交集。

二、 生成式 AI 發展下三類科技之角色分工與互動關係

在生成式 AI 成為通用型技術之背景下，FinTech、RegTech 與 SupTech 不再各自獨立運作，而形成相互影響、相互制衡之關係結構[114]。

就 FinTech 而言，生成式 AI 主要被應用於前台與中台業務，例如智能客服、虛擬助理、投資研究輔助與行銷內容生成等[115]。此類應用以提升客戶體驗與營運效率為主要目標，惟其輸出內容直接影響消費者決策，

---

[108] Financial Stability Board, *supra* note 87.
[109] Douglas W. Arner et al, *FinTech, RegTech, and the Reconceptualization of Financial Regulation,* 37 NW. J. INT'L L. & BUS. 371 (2017).
[110] Philipp Weber, Karl V. Carl & Oliver Hinz, *Applications of Explainable Artificial Intelligence in Finance—A Systematic Review of Finance*, INFORMATION SYSTEMS, AND COMPUTER SCIENCE LITERATURE, MGMT. REV. Q. (2023).
[111] FSB, *The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions: Market Developments and Financial Stability Implications* , https://www.fsb.org/uploads/P091020.pdf (last visited Jul. 5, 2026).
[112] Broeders and Prenio (2018)將 Suptech 定義為監理機關運用創新科技以支援監理工作的方式。本研究則進一步擴大該定義，將 Suptech 涵蓋至金融主管機關所使用之先進資料蒐集與分析工具，不僅限於傳統意義上的監理機關或監管機構，亦包括其他相關機關，例如金融情報單位（FIUs）等。*See* Dirk Broeders & Jermy Prenio, *Innovative Technology in Financial Supervision (Suptech): The Experience of Early Users*, FSB, BANK FOR INT'L SETTLEMENTS (2018), https://www.bis.org/fsi/publ/insights9.pdf.
[113] *Id.*
[114] FSB, *supra* note 87, at 11.
[115] FSB, *supra* note 69.

因而具有高度法遵與風險敏感性[116]。FinTech 應用規模愈大，對 RegTech 與 SupTech 之需求亦隨之提高[117]。

RegTech 則扮演承接 FinTech 創新與法規要求之中介角色，早期 RegTech 主要作為自動化合規與報告工具，著重於資料蒐集與事後檢核[118]。生成式 AI 使 RegTech 從過往的事後檢核工具，逐步轉型為即時輔助與預警系統[119]。例如，透過語言模型即時解析法規異動、比對內部流程文件，或於模型輸出階段加入合規檢核邏輯，以降低違規風險[120]。

SupTech 則運用生成式 AI 與進階分析技術，協助監理機關從大量申報資料與市場資訊中，快速識別潛在異常或系統性風險[121]。生成式 AI 所提供之摘要、分類與趨勢分析能力，使監理行為得以由傳統之事後稽核，逐步轉向近即時監控與預測式監理[122]。

因此，在生成式 AI 技術架構下，FinTech 著重於創新與服務、RegTech 著重於合規與風險控管，而 SupTech 則著重於整體金融穩定與市場秩序，三者形成一種動態平衡之治理結構。

三、 生成式 AI 對金融監理科技生態之影響

生成式 AI 的導入，使金融監理科技生態產生結構性變化。在技術層面，監理與法遵工具由「規則導向」逐步轉為「語意導向」，監理判斷不再僅依賴固定條件，而是結合語境理解與風險推論[123]。此一轉變雖提升監理效率，亦增加模型可解釋性與決策透明度之要求[124]。

在制度層面，生成式 AI 促使金融機構與監理機關之互動關係更加緊密。當金融業者廣泛使用生成式 AI 於核心業務時，監理機關勢必須理解其運作邏輯、風險特性與治理架構，否則將難以有效進行監理，此亦使 SupTech 與 RegTech 成為監理制度中不可或缺之基礎能力[125]。

而在治理層面，生成式 AI 的不確定性特質，使責任歸屬與風險分攤成為重要議題[126]。無論係金融機構使用生成式 AI 提供服務，或監理機關運用 SupTech 輔助決策，皆須建立適當之人機協作機制與審核流程，以避免過度依賴自動化判斷[127]。

---

[116] OECD & FSB, *OECD–FSB Roundtable on Artificial Intelligence (AI) in Finance* (2024/9/30), https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/digital-finance/OECD%20%E2%80%93%20FSB%20Roundtable%20on%20Artificial%20Intelligence%20%28AI%29%20in%20Finance.pdf.
[117] Dirk Broeders & Jermy Prenio, *supra* note 112.
[118] FSB, *supra* note 69, at 9-12.
[119] International Monetary Fund, *supra* note 73, at 12-15.
[120] *Id.*
[121] Dirk Broeders & Jermy Prenio, *supra* note 112 .
[122] International Monetary Fund, *supra* note 73.
[123] FSB, *supra* note 69, at 9-12.
[124] Philipp Weber, Karl V. Carl & Oliver Hinz, *supra* note 110 .
[125] *Id.*
[126] *Id.*
[127] *Id.*

第三節　金融機構運用生成式 AI 之主要應用領域

隨著生成式 AI 技術逐步成熟，其於金融機構之應用已由早期偏重客服或文件處理等低風險場域，擴展至涉及金融風險判斷、決策支援與監理敏感度較高之核心業務[128]。相較於傳統 AI 多著重於數值預測或規則比對，生成式 AI 具備語意理解、情境推論與內容生成能力，使其得以介入金融實務中高度依賴專業判斷與說明責任之場景[129]。本研究以下說明金融機構運用生成式 AI 之主要應用領域。

一、　風險管理與監理高度敏感之應用

（一）　詐欺偵測（Fraud Detection）

相較於傳統以規則或異常值為核心之偵測模型，生成式 AI 採取以行為模式為核心之異常辨識方式，能夠在大規模資料中動態識別偏離正常交易軌跡之可疑活動[130]。其亦能透過合成資料生成機制擴充訓練樣本，有助於提升模型泛化能力與準確性，並結合即時監控架構，能於交易發生當下即進行風險評估，從而降低潛在財務損失與詐欺風險暴露[131]。

（二）　授信與信用風險分析（Credit Risk Analysis）

生成式 AI 可透過處理大量非結構化資料（例如財報附註、產業報告、新聞資訊及內部信貸文件），協助金融機構進行借款人風險評估與授信決策。相較於傳統模型主要依賴結構化財務指標，生成式 AI 能夠進行語意層次之文本分析，快速彙整關鍵風險資訊，並支援信用備忘錄（Credit memos）之草擬與文件審查流程[132]。其應用亦延伸至投資組合監控與早期預警機制之建立，有助於強化信用風險管理之效率與精準度[133]。

（三）　投資組合與風險管理（Portfolio & Risk Management）

---

[128] FSB, *supra* note 69.
[129] Michael Kearns, *AI Reverberations Across Finance*, FIN. & DEV., at 38-41 (2023).
[130] OECD & FSB, *supra* note 116.
[131] 以 Mastercard 為例，其已將生成式 AI 應用於信用卡詐欺偵測領域。透過對大規模交易資料進行模式學習與異常分析，系統能更有效辨識遭盜刷之卡片與可疑交易行為。該技術不僅大幅提升受害卡片之識別能力，亦降低誤判情形，並顯著加快對高風險商戶之風險評估速度。Mastercard, *Mastercard Accelerates Card Fraud Detection with Generative AI Technology*, https://www.mastercard.com/news/press/2024/may/mastercard-accelerates-card-fraud-detection-with-generative-ai-technology/ (last visited Jul. 5, 2026).
[132] Andreas Kremer et al., *Embracing Generative AI in Credit Risk*, McKinsey & Co. (Jul. 1, 2024), https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/embracing-generative-ai-in-credit-risk.
[133] *Id.*

生成式 AI 有助於自動化例行性作業、強化決策品質，並發掘潛在的風險模式[134]。例如，AI 可在考量投資人風險承受度、市場趨勢與歷史績效資料的基礎上，優化資產配置策略，系統亦可模擬不同總體經濟情境，以評估投資組合之抗風險能力與韌性[135]。 另，透過整合即時市場數據與替代性資料來源，AI 模型得以在更全面之資訊基礎上進行判斷，同時亦透過可解釋性 AI 技術之導入，可提升模型決策過程之透明度，增強投資人對系統之信任[136]。

二、 投資決策支援與理財機器人之應用

生成式 AI 可大幅縮短資訊處理時間，使研究人員得以聚焦於高層次判斷，已被廣泛用於市場資訊彙整、研究報告摘要與策略分析輔助[137]。在此基礎上，理財機器人逐步結合生成式 AI 技術，形成新一代自動化投資服務模式，相較於早期以固定問卷與演算法模型為核心之理財機器人，結合大型語言模型之系統，已能透過自然語言互動方式理解投資人需求，並生成投資說明與風險揭露內容[138]。

三、 個人化客戶互動與行銷應用

透過分析客戶行為與偏好，系統可生成客製化金融內容、行銷訊息與服務建議，提升客戶體驗與行銷效率。惟此類應用涉及大量個人資料處理，對資料保護、目的限制與資訊透明度提出更高要求，若生成內容未清楚區分行銷與專業建議，具引發消費者保護與不當勸誘之疑慮[139]。

四、 內部營運、報表生成與制度支援應用

---

[134] Nishtha, *Generative AI for Finance*, PROJECTPRO (2024), https://www.projectpro.io/article/generative-ai-for-finance/1037 (last visited Jul. 5, 2026).

[135] 以 JPMorgan Chase 為例，其長期將 AI 技術整合於投資組合管理與風險控制流程之中。例如，其開發之 AI 交易系統（如 LOXM）透過分析市場數據與交易行為模式，以優化大宗交易執行策略並降低市場衝擊成本。生成式 AI 技術亦被運用於法律與監管文件之自動化解析，協助抽取關鍵條款與識別潛在合規風險。A. Ahmed Raza, *How JPMorgan Uses AI to Save 360,000 Legal Hours a Year*, MEDIUM (2025), https://medium.com/@arahmedraza/how-jpmorgan-uses-ai-to-save-360-000-legal-hours-a-year-6e94d58a557b (last visited Jul. 5, 2026).

[136] Nishtha, *supra* note 134.

[137] Fabrizio Dell'Acqua et al., *supra* note 10.

[138] Vanguard 於 2025 年推出其首項面向顧問端、可用於客戶溝通的生成式 AI 能力「Client-Ready Article Summaries」，可將其市場觀點文章自動生成「可直接提供客戶」的摘要，並可依客戶的金融素養程度、投資人生階段與語氣需求進行客製化，同時自動產出應隨附之必要揭露。雖不直接替代核心資產配置演算法，但可視為理財機器人／數位投資顧問流程中「前台互動與解釋層」的生成式強化：以自動化方式提升投資資訊傳遞效率與個人化程度，並支援顧問進行更一致、可規模化的投資溝通。*See* Vanguard, *Vanguard Unveils Generative AI Client Summaries for Financial Advisors*, https://corporate.vanguard.com/content/corporatesite/us/en/corp/who-we-are/pressroom/press-release-vanguard-unveils-generative-ai-client-summaries-for-financial-advisors-050525.html (last visited Jul. 5, 2026).

[139] OECD, *Recommendation of the Council on Artificial Intelligence*, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (last visited Jul. 5, 2026).

（一） 財務與監理報表生成（Financial Report Generation）

生成式 AI 已逐步被應用於內部管理報表、風險報告與監理申報文件之草擬與摘要，有助於降低人工作業負擔，並提升文件一致性，惟其輸出內容仍須經人工審核，以確保正確性與合規性[140]。

（二） 合成資料生成（Synthetic Data Generation）

運用生成式 AI 模型，在保留真實資料統計特性與結構關聯之情況下，人工產製具高度相似性的模擬資料[141]。金融機構可透過此技術補足稀缺事件樣本（如詐欺或違約案例）、進行壓力測試與模型驗證，並於資料受隱私或法規限制時，降低直接使用敏感個資之風險[142]。然而，合成資料之運用亦涉及偏誤傳遞、可稽核性及隱私再識別風險等治理議題，須納入既有模型風險管理與資料治理架構加以控管[143]。

（三） 內部流程與營運管理之自動化

除前台與中台業務外，生成式 AI 亦被廣泛應用於金融機構內部流程與營運管理，例如文件摘要、會議紀錄整理、內部知識問答與教育訓練內容生成等[144]。此類應用多屬低風險場景，亦常作為金融機構導入生成式 AI 之起點。惟即使屬於內部使用，仍須注意模型輸出正確性與資訊安全，以避免錯誤資訊擴散或敏感資料外洩[145]。

第四節 生成式 AI 於金融機構之實務應用案例

一、 國內實際應用實例

金融業為受到高度監管的行業之一，故多採取低風險、內部輔助之策略。2024 年 6 月 20 日金管會發布《金融業運用人工智慧（AI）指引》後，金融機構除導入傳統 AI 外，亦開始積極發展生成式 AI，並陸續推出各項運用 AI 技術的服務。不過，導入生成式 AI 的初期需要支出大量的建置或

[140] McKinsey & Co., *The Economic Potential of Generative AI: The Next Productivity Frontier* https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier (last visited Jul. 5, 2026).
[141] Jordan, J. et al., *Synthetic Data – What, Why, and How?,* ARXIV:2205.03257 (2022), https://arxiv.org/abs/2205.03257.
[142] *Id.*
[143] Financial Conduct Authority, *Generating and Using Synthetic Data for Models in Financial Services: Governance Considerations* (2025), https://www.fca.org.uk/publications/research/generating-using-synthetic-data-models-financial-services-governance (last visited Jul. 5, 2026).
[144] 例如 Goldman Sachs 於公司內部推出名為「GS AI Assistant」之生成式 AI 工具，並向全體員工開放使用。其可協助員工進行文件摘要、內容撰寫、資料整理與分析等工作，且可整合公司既有資訊系統與內部資料資源，並作為員工日常工作之輔助應用，涵蓋投資研究、行政作業及內部溝通等場景。*See Goldman Sachs launches AI assistant firmwide*, REUTERS, https://www.reuters.com/business/goldman-sachs-launches-ai-assistant-firmwide-memo-shows-2025-06-23/ (last visited Jul. 5, 2026).
[145] International Monetary Fund, *supra* note 73.

採購成本，因此大部分都還是由財力較雄厚的金融控股公司或其旗下子公司率先導入。以下針對臺灣 14 家金融控股公司（台新金控與新光金控合併後為 13 家）及其子公司與其他金融機構導入生成式 AI 的情形進行彙整。

（一） 玉山銀行「GENIE」AI 平臺、「AI 房貸顧問」與「投資 i-chat」

玉山銀行於 2024 年對內部推出通用型生成式 AI 平臺「GENIE」[146]，串接內部金融知識與法規資料，並透過 API 與現有的系統整合，快速嵌入既有業務流程。目前已於內部落地超過十項應用，包含純文字和圖片生成、會議記錄生成、翻譯等，並透過檢索增強生成技術（RAG），提升針對金融領域的回應準確性與可追溯性。例如，讓模型能讀懂會議中出現的專有名詞、辨識產品名稱、分辨出發言者身份等。具體而言，員工可以透過「GENIE」協助快速蒐集、摘要客戶的負面新聞，快速撰寫徵審報告；導入生成式 AI 協助行員分析來自客服 Chatbot 和留言板的文字訊息，也能將語音訊息轉為文字，讓行員直接從文字分析顧客情緒和來電意圖；將顧客的信用卡刷卡資料匯入生成式 AI ，以分析特定顧客樣態與行為模式；協助新進行員快速查詢教育訓練講義和規章，省下翻閱大量文件資料的時間；當行員遇上法遵問題時，可以使用法規文件比對服務快速了解公司內部法條和外部規範。未來將持續開發支援非文字形式的音檔、影像資料萃取服務。GENIE 係採用微軟 Azure OpenAI 的封閉模型服務[147]，以確保客戶敏感資料的安全性。

此外，玉山銀行亦與 Google Cloud 合作[148]，推出兩項以生成式 AI 為核心的金融顧問應用「AI 房貸顧問」與「投資 i-chat」。「AI 房貸顧問」可提供房貸基本知識與房市資訊，另外若舊客戶有增貸意願，亦提供一鍵申請、節省約九成的資料填寫時間；「投資 i-chat」則是即時檢索玉山投資研究團隊的市場研究及市場新聞，透過對話協助投資人理解市場變化，並透過問答互動判斷顧客屬性，進而提供推薦的產品資訊。

（二） 國泰金控「GAIA 2.0」框架

國泰金控於 2024 年首次提出其生成式 AI 技術發展框架「GAIA（Gen AI Architecture）」，運用集團各個子公司的技術與數據，建立專屬的金融

[146] 李昀璇，玉山怎麼用生成式 AI ？12 項應用服務大解析，iThome：https://www.ithome.com.tw/news/162551（最後瀏覽日：2026 年 7 月 5 日）。
[147] 張浩祥，生成式 AI 原理及在金融領域中之應用與突破，頁 9，中央銀行資訊處：https://www.tpefx.com.tw/uploads/download/tw/0721-11.pdf（最後瀏覽日：2026 年 7 月 5 日）。
[148] 玉山銀行，玉山與 Google Cloud 合作推出 AI 金融顧問 引領金融數位轉型，玉山銀行：https://www.esunbank.com/zh-tw/about/news-center/detail?id=114999119E9E4FABA44B89DE8F87A60F（最後瀏覽日：2026 年 7 月 5 日）。

領域知識庫[149]。此外，國泰金控更表示已建立 AI 模型中心（Model-Hub），匯集 50 多種開源、閉源及國泰專屬模型，並自建 AI 護欄，設有 70 道安全防護檢查點[150]，使 AIaaS（AI as a Service，AI 即服務）能夠在實際場域合規落地。2025 年國泰金控再提出「GAIA 2.0」技術框架，導入 AI Agent 概念，讓員工運用生成式 AI 打造內部應用服務，以完成行銷文案、各類產品知識、趨勢摘要等工作；對外則運用生成式 AI 於「CUBE」應用程式中打造能以自然語言檢索服務的功能[151]。

（三） 永豐金控「永豐雲 Chat!」、「Orbit.AI 投資水晶球」和「iWish」

永豐金控於 2022 年開始運用 AI 模型[152]篩選信用貸款行銷名單，也透過文件辨識（OCR）和流程自動化技術（RPA），優化旗下銀行、證券和租賃公司的紙本作業流程。2024 年陸續推出「永豐雲 Chat!」和「Orbit.AI 投資水晶球」服務[153]，前者以生成式 AI 提供生成文案、文字摘要、語言翻譯、查找規章等功能，協助員工節省時間、提升工作效率；後者則利用生成式 AI 彙整多個新聞來源，並產生投顧報告，協助客戶掌握市場趨勢。另外，針對親臨分行辦理業務的客戶，永豐銀行亦推出「iWish」的對話填單服務[154]，讓客戶可以用自然口說的方式完成臨櫃辦理業務需要填寫的單據。

永豐金控並表示[155]，未來將朝向「系統 API 化」邁進，讓系統可以被 AI 呼叫，並建立可信賴的數位邊界，以保護顧客與公司的機敏資料。

（四） 富邦金控「Fubon＋」、「富邦 AI PRO」

富邦金控自 2023–2025 年間組成「生成式 AI 應用推動團隊」，並在銀行、證券、人壽等多個子公司導入生成式 AI[156]，並與 IBM 合作[157]，打造

[149] 許靜之，CUBE 成本增 50%，收益成長 8 倍！國泰首發布「GAIA」生成式 AI 框架，曝數位轉型成果：https://www.bnext.com.tw/article/80649/cathay-holdings-ai-frameworks-gaia（最後瀏覽日：2026 年 7 月 5 日）。

[150] 黃于庭，國泰金控技術年會揭露生成式 AI 框架 GAIA 2.0 宣告進入雲端優先階段，經濟日報：https://money.udn.com/money/story/5613/9082615（最後瀏覽日：2026 年 7 月 5 日）。

[151] 國泰金控，AI 代理時代已至！國泰金控以 GAIA 2.0 框架加速 AI 應用百花齊放，數位時代：https://www.bnext.com.tw/article/85468/cathay1209（最後瀏覽日：2026 年 7 月 5 日）。

[152] 王宏仁，永豐金控揭露多 AI 代理戰略布局，關鍵是 Agent 宇宙藍圖和五大基礎建設，iThome：https://www.ithome.com.tw/news/173012（最後瀏覽日：2026 年 7 月 5 日）。

[153] 永豐銀行，MongoDB Atlas 與生成式 AI 完美結合 永豐銀行數位金融服務再進化，iThome：https://www.ithome.com.tw/pr/168603（最後瀏覽日：2026 年 7 月 5 日）。

[154] 鄭芝珊，業界首家！永豐 iWish 創新應用 GAI ，直覺對話讓臨櫃存提轉匯變得更快速，經濟日報：https://money.udn.com/money/story/5635/8996799（最後瀏覽日：2026 年 7 月 5 日）。

[155] 王宏仁，註 152。

[156] 藍鈞達，富邦金控積極擁抱 AI 組成「生成式 AI 應用推動團隊」，經濟日報：https://money.udn.com/money/story/5613/9125014（最後瀏覽日：2026 年 7 月 5 日）。

[157] IBM，IBM 諮詢助力台北富邦銀行推出 Fubon+，IBM Taiwan Newsroom：https://taiwan.newsroom.ibm.com/2025-04-15-IBM-Fubon（最後瀏覽日：2026 年 7 月 5 日）。

「Fubon＋」應用程式，整合銀行、保險、證券等多項金融服務，並提供資產負債分析、消費洞察、投資提示及個人化推播（如匯款提醒、外幣匯率低點通知）等功能。此外，金控旗下的台北富邦銀行亦開發「AI 鷹眼識詐模型」[158]，為全台首創 AI[159]預警機制，能快速識別異常交易，提升臨櫃關懷及線上攔阻的準確率，並偕同刑事局及金融同業成立「鷹眼聯盟」，無償分享鷹眼模型，共組反詐生態系統。富邦證券亦推出新一代應用程式「富邦 AI PRO」，透過生成式 AI 即時解析市場訊息與快速彙整法說會重點[160]，協助投資人進行交易決策。而富邦產險則以 AI 人臉辨識[161]技術推出「FacePass 極速保」，將身分驗證與線上簽署流程數位化[162]，實現快速投保的體驗。

2024 年富邦金控更舉辦「首屆生成式 AI 黑客松」[163]，以「將素人培訓為 AI 達人」為目標，多支參賽隊伍針對行政與人資流程（如履歷審閱、自動化行政流程等）、客戶互動強化、程式開發與檢核輔助、財務模型解讀工具、法遵與洗錢防制應用等方面進行提案與實作競賽，未來有機會將競賽成果實際部署運用。

（五） 中國信託「CTBC GPT」和「AI 信用卡秘書」[164]

中國信託透過 ATM 的攝影機影像結合 AI[165]辨識，針對同時講電話與操作 ATM、配戴遮蔽臉部特徵物品等高風險行為，系統將自動暫緩交易

---

[158] 呂欣芷，富邦金秀 AI 創新力 銀行全新 APP「Fubon＋」亮相，工商時報：https://tw.stock.yahoo.com/news/%E5%AF%8C%E9%82%A6%E9%87%91%E7%A7%80ai%E5%89%B5%E6%96%B0%E5%8A%9B-%E9%8A%80%E8%A1%8C%E5%85%A8%E6%96%B0app-fubon-%E4%BA%AE%E7%9B%B8-052614445.html（最後瀏覽日：2026 年 7 月 5 日）。

[159] 需特別說明，此類用於分析與預測的 AI，其底層多使用決策樹或卷積神經網路架構，屬於傳統機器學習（傳統 AI）領域；而用於內容生成的生成式 AI 則是使用 Transformer 架構。*See* Rahul Mundlamuri et al., *The Evolution of AI: From Classical Machine Learning to Modern Large Language Models*, 13 IEEE ACCESS 178302-7 (2025).

[160] 唐可欣，數位理財新升級 富邦金子公司推新 APP 與線上投保服務：https://www.storm.mg/article/11099961（最後瀏覽日：2026 年 7 月 5 日）。

[161] 需特別說明，人臉辨識（Facial Recognition）本質上屬於傳統機器學習與深度學習的範疇，主要仰賴卷積神經網路（CNN）架構，而非以生成或創造為目的的生成式 AI 。*See* Mundlamuri et al., *supra* note 159.

[162] 戴玉翔，保險業首家！富邦產險「FacePass 極速保」開創 AI 刷臉投保服務，經濟日報：https://money.udn.com/money/story/5613/9007939（最後瀏覽日：2026 年 7 月 5 日）。

[163] 富邦金融控股股份有限公司，富邦金控舉辦首屆生成式 AI 黑客松，富邦金控：https://www.fubon.com/financialholdings/home/index.html（最後瀏覽日：2026 年 7 月 5 日）；富邦金融控股股份有限公司，生成式 AI 改變職場！富邦金首度舉辦黑客松，2 個月將素人培訓為 AI 達人，中央社：http://www.cna.com.tw/postwrite/chi/390486（最後瀏覽日：2026 年 7 月 5 日）。

[164] 中國信託，從「AI 智能防線」到「AI 行員」中信打造全台第一個「AI 防詐金融生態系」，商業周刊：https://www.businessweekly.com.tw/business/indep/1006271（最後瀏覽日：2026 年 7 月 5 日）。

[165] 需特別說明，人臉辨識（Facial Recognition）本質上屬於傳統機器學習與深度學習的範疇，主要仰賴卷積神經網路（CNN）架構，而非以生成或創造為目的的生成式 AI 。*See* Mundlamuri et al., *supra* note 159.

或要求移除遮蔽物。此外，ATM 亦與警政單位連線，若偵測到警示帳戶或特定名單上的帳戶正在嘗試透過 ATM 進行交易，系統則會即時通報轄區派出所派員前往盤查。

另外，中國信託於企業內部推出「CTBC GPT」，透過生成式 AI 提供會議記錄生成、翻譯文件、生成洞察市場初步報告等服務；未來對外亦將推出「AI 信用卡秘書」服務，透過生成式 AI 與客戶進行口語化對答，提供信用權益與分析信用卡消費紀錄。

（六） 台新金控（現為台新新光金控）「信用貸款自動文件審核系統」、「台新新光腦」與「Sunny 智能客服」

台新金控於 2023 年導入微軟 Azure OpenAI 服務，嘗試運用雲端生成式 AI 提供投資商品建議[166]；2024 年進一步推出「信用貸款自動文件審核系統[167]」，系統將案件進行分類，並自動依風險等級完成文件審查與撥款，大幅提升放款效率。另外，台新金控亦建置地端 AI 模型「台新新光腦」[168]，讓員工可以透過和台新新光腦對話快速查找產品知識庫、生成產業分析報告；也透過 RAG 技術大幅提升「Sunny 智能客服」的對話能力和識別客戶意圖的能力，並生成擬人化的回應。

（七） 第一銀行「Project GALA」與「AI 智能鑑價」

第一銀行「Project GALA」係運用生成式 AI 建立企業問答助理[169]，整合行內知識庫，幫助分行員工快速查找規範和流程指引；亦導入 M365 Copilot，可自動摘要、整理重點並生成簡報。另外，第一銀行也打造「AI[170]智能鑑價」系統[171]，提高客戶取得房屋預估可貸金額及每月還款金額的效率。

（八） 凱基人壽「智能業務對練 2.0」

---

[166] 李昀璇，台新金控首度公開導入生成式 AI 的智能客服，提供理財投資建議和信用卡建議，iThome：https://www.ithome.com.tw/news/160151（最後瀏覽日：2026 年 7 月 5 日）。

[167] 台新國際商業銀行，台新銀行奪 2025《數位銀行家》五項大獎 以 AI 與數據核心引領金融創新體驗，台新國際商業銀行：https://www.taishinbank.com.tw/TSB/personal/common/news/TSBankNews-006851/（最後瀏覽日：2026 年 7 月 5 日）。

[168] 陳以謙，台新孫一仕：AI 進入民主化時代，臺灣銀行家：https://taiwanbanker.tabf.org.tw/paperDetail?id=4516（最後瀏覽日：2026 年 7 月 5 日）。

[169] 李昀璇，第一銀行公開最新生成式 AI 行動方案 GALA 計畫，優先賦能員工素養，再聚焦兩大發展重點，iThome：https://www.ithome.com.tw/news/169052（最後瀏覽日：2026 年 7 月 5 日）。

[170] 需特別說明，此類用於分析與預測的 AI，其底層多使用決策樹或卷積神經網路架構，屬於機器學習（傳統 AI）領域。*See* Mundlamuri et al., *supra* note 159.

[171] 第一銀行，從競賽激發創新力，第一銀行如何以 AI 打造差異化競爭優勢、型塑數位生活圈？，數位時代：，https://www.bnext.com.tw/article/83952/firstbank202507（最後瀏覽日：2026 年 7 月 5 日）。

凱基人壽運用生成式 AI 及 RAG 技術模擬客戶行為[172]，與業務員進行情境式對答，以訓練業務員的應對能力。此外，業務員得以問答的方式取得業務手冊內容與產品摘要，以便快速掌握產品特性和銷售技巧[173]。

（九） 華南金控生成式 AI 、GitHub Copilot 與防詐、防洗錢模型[174]

華南銀行與雲端廠商合作，運用各部門的數據和文件，以生成式 AI 產生人資和存匯實務手冊、建立內部知識管理平臺，以利行員快速查找相關規範，並開始發展 AI 代理人（AI Agent）服務 ，生成產經徵信研究、授信報告和投資研究等報告。此外，華南銀行亦採用 GitHub Copilot 輔助 IT 團隊進行弱點修補、系統手冊撰寫。另外，華南銀行亦透過機器學習技術，開發防詐、防洗錢模型，針對可疑交易與風險客戶進行標記。

（十） 元大金控「元先生」智能客服

元大銀行結合 AI 技術，推出「元先生」智能客服[175]，讓客服人員能透過 AI 輔助，快速查詢信用卡帳務及變更等類似問題，以及其他金融相關資訊，以提高回覆客戶的精準度。

（十一） 臺灣企銀樂齡行動銀行

臺灣企銀結合 AI 與樂齡友善設計，建置 i-Money 智能語音小幫手[176]，支援以語音完成找功能、查詢餘額、設常用功能、轉帳與繳費等操作，降低文字與多層選單對高齡族群的負擔，也加入「跌倒偵測與緊急求助功能」，讓使用者預先設定緊急聯絡人，當系統偵測到疑似跌倒時會自動跳出求助視窗，提高銀髮族群使用行動銀行時的安全感。

（十二） 兆豐金控與合作金庫金控正在導入中

---

[172] 李昀璇，【LLM 模型如何落地金融業？技術與非技術難題並存】凱基人壽導入生成式 AI 的三大挑戰，iThome：https://www.ithome.com.tw/news/162052（最後瀏覽日：2026 年 7 月 5 日）。

[173] 凱基人壽保險股份有限公司，壽險業第一家！凱基人壽導入生成式 AI 打造頂尖業務部隊：https://www.kgilife.com.tw/zh-tw/-/media/files/kgil/homepage/new-folder1/2024/20240201-ai.pdf（最後瀏覽日：2026 年 7 月 5 日）。

[174] 李昀璇，高層每月緊追生成式 AI 績效，華南銀揭生成式 AI 兩大應用模式，一年創造 50 多名數位勞動力，iThome：https://www.ithome.com.tw/news/169617（最後瀏覽日：2026 年 7 月 5 日）。

[175] 財管專案小組，元大銀行 AI 技術導入智能客服，提供精準服務，元大銀行：https://www.wealth.com.tw/articles/d9ccb8e8-3505-454f-9110-31c752f69c05（最後瀏覽日：2026 年 7 月 5 日）。

[176] 臺灣中小企業銀行，2024 台北金融科技展 臺灣企銀邀民眾探索「Hokii 星球」：https://www.tbb.com.tw/zh-tw/about/information/announcement/news/press-announcements/20241101（最後瀏覽日：2026 年 7 月 5 日）。

報導指出[177]，兆豐銀行、合作金庫銀行均正在導入生成式 AI 及代理式 AI 服務，預計運用在客服、業務輔助、知識庫問答與行銷分析等領域。

（十三） 金融大語言模型（FinLLM）專案

有鑒於金融業高度仰賴語言與文本處理，但同時也是高度監理產業，難以直接套用國際通用的大語言模型，發展在地化金融大語言模型，已成為產業升級的必要工程[178]。此專案由金融科技產業聯盟[179]（Taiwan FinTech Alliance, TFA）推動，集結 16 家金融機構與產、官、學資源，避免重複投資，並建立共用的 AI 基礎設施。此有助於降低金融業對境外科技公司的依賴，也有助於資料治理[180]。專案預計 2026 年底推出首版銀行通用模型，並於 2027 年第 1 季推出以 FinLLM 為基礎的 AI Agent。

（十四） 其他金融業導入生成式 AI 之情況

除上開金融控股公司外，部分非金控旗下的銀行、壽險公司與證券公司亦有導入生成式 AI 並運用於其業務中，如：群益金鼎證券與精誠資訊公司合作推出的群益贏家 Pro「AI 問問」功能[181]、南山人壽針對法務人員所設計的 AI 智能助理[182]、台中銀行預計與群聯電子合作[183]自建的地端大型語言模型[184]aiDAPTIV+等。

二、 國外實際應用案例

（一） JPMorgan Chase「LLM Suite」

摩根大通（JPMorgan Chase）以每年約 180 億美元的科技預算為基礎，推動包含代理式 AI 與生成式 AI 在內的多步驟工作自動化，並建置自有的生成式 AI 平台 LLM Suite（由多家第三方大型語言模型支援），並將 AI 工

[177] 葉佳華，公股銀拚 AI 運用多頭並進 今年聚焦兩大應用發展，經濟日報：https://money.udn.com/money/story/5613/9240342（最後瀏覽日：2026 年 7 月 5 日）。
[178] 廖珮君和任珮云，金融業打團體戰自建 AI 模型 金管會主委彭金隆看好三效益，經濟日報：https://money.udn.com/money/story/5613/9458638（最後瀏覽日：2026 年 7 月 5 日）。
[179] 金融科技產業聯盟是金融總會在金管會指導下所成立的跨機構合作平台，目標是聯合臺灣金融業推動金融科技與 AI 應用。
[180] 廖珮君和任珮云，註 178。
[181] 陳怡瑄，決策式 AI 引領投資新速度 群益金鼎證券攜手精誠資訊「AI 問問」一秒看懂個股重點，卓越雜誌：https://ecf.com.tw/tw/article/show.aspx?num=8813（最後瀏覽日：2026 年 7 月 5 日）。
[182] 南山人壽，用 AI 開外掛！一窺南山人壽法遵 AI 智能助理的架構工法，數位時代：https://www.bnext.com.tw/article/80404/nanshanlife202409（最後瀏覽日：2026 年 7 月 5 日）。
[183] 黃登榆，台中銀行與群聯電子跨界聯手 打造 AI 金融服務新平台，經濟日報：https://money.udn.com/money/story/5613/9043861（最後瀏覽日：2026 年 7 月 5 日）。
[184] 群聯電子，Phison | aiDAPTIV+：https://phisonaidaptiv.com/zh-tw/#（最後瀏覽日：2026 年 7 月 5 日）。

具直接提供給員工使用，使生成式 AI 從少數團隊試點轉向「全員可用」的生產力基礎設施[185]。該平台已讓接近 25 萬名員工取得使用權，且「略低於一半」員工每日使用生成式 AI，實際應用涵蓋法務契約掃讀、比對與生成、授信條款與契約條款資訊擷取、業務與前線銀行人員的客戶會議準備與資訊濃縮等多元情境[186]。為在高度監管環境中擴散採用，摩根大通採取「先控風險再擴規模」的節奏，將資料安全視為最優先考量，不以強制方式推動，而採分階段、員工自願（opt-in）導入並搭配訓練與變革管理，利用口碑與社群分享（如 prompt library、prompt of the week 等）形成擴散式學習機制[187]。該行亦提出「雙支柱」策略：一方面由上而下鎖定對經濟效益最關鍵的流程與領域（如授信、詐欺、行銷、技術開發、營運與前線賦能）進行端到端（end-to-end）再設計；另一方面則提供員工自助式工具，鼓勵長尾工作場景的自發創新累積成整體生產力提升[188]。LLM Suite 亦從早期聊天機器人進化為生態系，目標是讓核心 AI 智能連結團隊知識系統、全行資料與應用程式，支援簡報、資料分析與報告生成，並透過持續增加資料集與系統連結擴大可解決問題的範圍[189]。

（二） Morgan Stanley「AI@MS」系列

摩根士丹利（Morgan Stanley）2023 年 9 月將「AI @ Morgan Stanley Assistant」全面推廣給財富管理顧問與支援人員，協助旗下 1 萬 6 千多名財務顧問快速彙整研究報告和文件[190]。2024 年 6 月進一步推出「AI @ Morgan Stanley Debrief」，其以 OpenAI 模型為基礎，在客戶事前同意之前提下，協助財務顧問於客戶會談後自動生成會議摘要、行動項目及後續電子郵件草稿，並將相關紀錄回寫至客戶關係管理系統（Customer Relationship Management, CRM）Salesforce[191]。相較於直接參與投資判斷或產品建議，其明確聚焦於會談後的行政整理與溝通支援，所有生成內容仍須由財務顧問進行最終審閱與確認後方可使用[192]。摩根士丹利將生成式 AI 的成功視為「變革管理（Change Management）」問題，而非單純技術部署，透過大量內部教學影片與實務指導，從基礎提示設計到進階

[185] McKinsey & Co., *JPMorgan Chase's Derek Waldron on AI and banking*, MCKINSEY & CO. (2025).
[186] Lindsey Wilkinson, *JPMorgan Chase to equip 140K workers with generative AI Tool*, CIO DIVE, (2024).
[187] McKinsey & Co., *supra* note 185.
[188] *Id.*
[189] *Id.*
[190] Morgan Stanley, *Launch of AI @ Morgan Stanley Debrief*, https://www.morganstanley.com/press-releases/ai-at-morgan-stanley-debrief-launch (last visited Jul. 5, 2026).
[191] 李昀璇，Fintech 周報第 233 期：摩根士丹利揭露生成式 AI 發展策略，要以 AI 助理促進顧問、顧客和平臺三方無縫互動，iThome: https://www.ithome.com.tw/news/161710（最後瀏覽日：2026 年 7 月 5 日）。
[192] Hugh Son, *Morgan Stanley wealth advisors are about to get an OpenAI-powered assistant to do their grunt Work*, CNBC (2023), https://www.cnbc.com/2024/06/26/morgan-stanley-openai-powered-assistant-for-wealth-advisors.html.

代理 AI 應用，協助員工在日常工作中養成使用 AI 工具的習慣，並仰賴高階管理層的持續支持以加速文化轉型[193]。

（三） Citigroup「Citi Assist」、「Citi Stylus」、「Citi Squad」

花旗（Citigroup）自 2024 年起於多個市場（包含美國、英國、新加坡、香港等）分階段推出多項生成式 AI 內部應用工具，其推出三項功能分工明確的應用[194]。Citi Assist 被定位為內部政策與流程的智慧查詢助理，協助員工以自然語言快速搜尋並理解行內制度性文件與作業程序，涵蓋人力資源、風險、法令遵循與財務等領域，藉此縮短員工查詢規範文件的時間，並降低因資訊理解不足所產生的作業或合規風險[195]。Citi Stylus 為文件處理導向的生成式 AI 工具，支援多文件摘要、比對、搜尋與翻譯，並延伸提供工作區功能，用於草擬電子郵件、製作簡報講者備註與建立問答內容，協助員工在知識密集與行政性任務中提升處理效率[196]。Citi Squad 則面向內部軟體開發人員，作為 AI 程式碼助理協助撰寫與優化程式碼，以縮短開發週期並加速產品部署[197]。而在推動方式上，花旗並未將生成式 AI 視為單點技術導入，而是結合組織治理與變革管理同步推進[198]。透過分國別、分角色逐步開放工具使用，觀察實際工作情境與風險[199]，另亦搭配大規模員工教育訓練，強調如何在合規前提下正確使用生成式 AI 工具，並培養提示設計與結果判讀能力[200]。

（四） Wells Fargo「Fargo」

富國銀行（Wells Fargo）在生成式 AI 上的應用，可概括為「對客數位助理」與「對內代理式工作流程工具」兩條主軸並行推進[201]。對外方面，富國銀行於行動銀行應用程式中提供 Fargo 虛擬助理，用於回應客戶日常銀行相關問題，並採用 Google 的對話式 AI（如 Dialog flow）理解使

---

[193] Hugh Son, *Morgan Stanley kicks off generative AI era on Wall Street with assistant for financial Advisors*, CNBC (2023), https://www.cnbc.com/2023/09/18/morgan-stanley-chatgpt-financial-advisors.html (last visited Jul. 5, 2026).

[194] 鄭妤安，花旗 Citi AI 工具正式在台推出 打造 AI 智慧團隊，工商時報：https://www.ctee.com.tw/news/20251111701333-430304（最後瀏覽日：2026 年 7 月 5 日）。

[195] Tatiana Bautzer, *Citigroup rolls out artificial intelligence tools for employees in eight Countries*, REUTERS, https://www.reuters.com/technology/artificial-intelligence/citigroup-rolls-out-artificial-intelligence-tools-employees-eight-countries-2024-12-04/ (last visited Jul. 5, 2026).

[196] *Id.*

[197] *Id.*

[198] Ben Udell, *We're Speeding Past the Tipping Point of Gen AI in Banking*, THE FINANCIAL BRAND, https://thefinancialbrand.com/news/artificial-intelligence-banking/were-speeding-past-the-tipping-point-of-gen-ai-in-banking-193065 (last visited Jul. 5, 2026).

[199] The Digital Banker, *Citi rolls out AI tools to employees in Hong Kong*, The Digital Banker, https://thedigitalbanker.com/citi-rolls-out-ai-tools-to-employees-in-hong-kong/ (last visited Jul. 5, 2026).

[200] *Id.*

[201] Eric Best, *Artificial Intelligence and You*, WELLS FARGO STORIES, https://stories.wf.com/our-impact/wells-fargo-artificial-intelligence-and-you/ (last visited Jul. 5, 2026).

用者意圖，以提升自助服務的可用性與即時性[202]。根據公開資訊，Fargo 的互動量已累積達約 2.45 億次，並可在「零人工交接（no human handoffs）」的情境下運作[203]；同時，該系統透過資料去識別與個人可識別資訊（Personally Identifiable Information, PII）檢測等機制，避免敏感資料直接送入大型語言模型[204]。對內方面，富國銀行則與 Google Cloud 合作，作為 Google Agentspace 的早期採用者，推動代理 AI 在企業內部的規模化應用[205]。透過可控且統一的安全平台，該行得以建立與管理多個代理式 AI，協助員工更快速地檢索與綜整跨系統資訊，並將重複性任務納入自動化工作流程，以提升組織敏捷性與作業效率，亦可用於改善客戶體驗、支援例行作業自動化，並輔助行內團隊的日常決策與營運作業[206]。

國內外銀行實際應用案例整理，請參照下表 2、生成式 AI 於金融機構之實務應用。

[202] *Id.*
[203] Matt Marshall, *Wells Fargo's AI assistant just crossed 245 million interactions – no human handoffs, no sensitive data Exposed*, VENTUREBEAT, https://venturebeat.com/business/wells-fargos-ai-assistant-just-crossed-245-million-interactions-with-zero-humans-in-the-loop-and-zero-pii-to-the-llm (last visited Jul. 5, 2026).
[204] *Id.*
[205] Tracy Kerrins and Rohit Bhat, *Wells Fargo brings the agentic era to financial services with Google Cloud AI*, GOOGLE CLOUD, https://cloud.google.com/blog/topics/financial-services/wells-fargo-agentic-ai-agentspace-empowering-workers (last visited Jul. 5, 2026).
[206] *Id.*

表 2、生成式 AI 於金融機構之實務應用案例

類別	金融機構	應用名稱	是否觸及客戶個資	是否觸及交易資料	是否觸及內部機密	功能簡述
對客互動與資訊提供型應用	玉山銀行	AI 房貸顧問	可能	可能	無	提供房貸基本知識與房市資訊，以及舊客戶一鍵增貸
		投資 i-chat	可能	可能	無	即時檢索玉山投資研究團隊的市場研究及市場新聞
	國泰世華銀行	CUBE Intelligence	可能	可能	無	CUBE App 內的智能助理，具備對話式搜尋與個人化理財建議功能
	永豐金證券	Orbit.AI 投資水晶球	無	無	無	自動摘要上千篇財經新聞，生成「市場輿情」重點
		iWish	可能	可能	無	對話式金融服務，用對話方式完成臨櫃填單
	富邦證券	富邦 AI PRO	可能	可能	無	透過 GenAI 即時解析市場訊息與快速彙整法說會重點
	中國信託銀行	理財智囊團	可能	可能	無	運用 GenAI 摘要市場財經新聞，提供理專與客戶投資建議
	台新銀行	Sunny 智能客服	可能	可能	無	精準識別客戶意圖並進行如同人類般的對話
	元大銀行	「元先生」智能客服	可能	可能	無	客服人員透過 AI 輔助，快速查詢信用卡帳務及變更等類似問題
	臺灣企銀	i-Money 智能語音小幫手	可能	可能	無	支援以語音完成找功能、查詢餘額、設常用功能、轉帳與繳費等
	台中銀行	aiDAPTIV+（規劃中）	可能	可能	無	代理式 AI 處理顧客問題、智慧投資助理
	群益金鼎證券	群益贏家 Pro	無	無	無	AI 自動閱讀新聞與財報，生成 6 大重點摘要
	Wells Fargo	Fargo 虛擬助理	可能	可能	無	理解使用者意圖，並回應客戶日常銀行相關問題
內部行政與營運輔助型應用	國泰金控	GAIA Studio	可能	可能	可能	員工可以利用 GenAI 建立撰寫信件、自動生成會議紀錄工具
	永豐金控	永豐雲 Chat!	可能	可能	可能	自建 GenAI 平台，可進行文書處理

	中信金控	CTBC GAI	可能	可能	可能	生成會議紀錄與翻譯
	第一金控	Project GALA	可能	可能	可能	建立問答助理，幫助分行員工快速查找規範和流程指引
	富邦金控	GenAI 黑客松	無	無	無	利用 GenAI 建立行政、人資、法遵文書工具
	華南金控	華寶 AI 助理	無	無	可能	產出人資、存匯實務手冊與系統文件
		GitHub Copilot	無	無	無	輔助 IT 團隊進行弱點修補、系統手冊撰寫
	台新金控	台新新光腦	可能	可能	可能	繁體中文 LLM 用於法遵檢索與內部問答
	凱基人壽	KGI Life Copilot	可能	可能	可能	協助摘要保險文件與法規
	南山人壽	AI 智能助理	可能	可能	可能	針對法務設計，自動閱讀裁罰書、生成案件事實摘要、繪製圖表
	凱基人壽	智能業務對練 2.0	無	無	無	業務員可透過問答進行銷售演練
	新光人壽	AI 講師	無	無	無	業務員可隨時與 AI 講師進行商品知識問答與銷售演練
	兆豐金控	Mega 理財小秘書（規劃中）	可能	可能	可能	法遵檢索、文書摘要
	台中銀行	aiDAPTIV+（規劃中）	可能	可能	可能	自動生成文件、處理內部表單與流程
	玉山金控	GENIE	可能	可能	可能	自建 GenAI 平台，可處理徵審資料、客服紀錄、刷卡資料等
	JPMorgan Chase	LLM Suite	可能	可能	可能	法務契約掃讀、比對與生成、授信條款與契約條款資訊擷取、業務與前線銀行人員的客戶會議準備與資訊濃縮
	Morgan Stanley	AI @ Morgan Stanley Assistant	無	無	無	快速彙整研究報告和文件
		AI @ Morgan Stanley Debrief	可能	可能	無	自動生成會議摘要、行動項目及電子郵件草稿，並將相關紀錄回寫至客戶關係管理系統

	Citigroup	Citi Assist	可能	可能	可能	協助員工以自然語言快速搜尋並理解行內制度性文件與作業程序，縮短員工查詢規範文件的時間，並降低因資訊理解不足所產生的作業或合規風險
		Citi Stylus	可能	可能	無	支援多文件摘要、比對、搜尋與翻譯，並延伸提供工作區功能，用於草擬電子郵件、製作簡報講者備註與建立問答內容
		Citi Squad	無	無	無	為 AI 程式碼助理協助撰寫與優化程式碼，以縮短開發週期並加速產品部署
	Wells Fargo	Google Agentspace	可能	可能	可能	建立與管理多個代理式 AI，協助員工更快速地檢索與綜整跨系統資訊，並將重複性任務納入自動化工作流程
決策輔助與評估支援型應用	國泰金控	GAIA	可能	可能	可能	自建 GenAI 平台，可自建工具產出風險與趨勢摘要
	華南金控	AI Agent（規劃中）	可能	可能	可能	可生成徵審、授信、徵信報告
	富邦銀行	AI 鷹眼識詐模型	可能	可能	無	快速識別異常交易，提升臨櫃關懷及線上攔阻的準確率
	富邦產險	FacePass 極速保	可能	可能	無	將身分驗證與線上簽署流程數位化，實現快速投保的體驗
	台中銀行	aiDAPTIV+（規劃中）	可能	可能	無	異常交易偵測、防詐與交易監控模型

資料來源：本研究自行整理

第五節　生成式 AI 於金融機構之應用場景類型化分析

由上述國內外金融機構的生成式 AI 應用情況可知，多數金融機構在 2024 年之後才公開推出生成式 AI 的相關應用，顯見金融機構於運用生成式 AI 方面確實較爲謹慎。而觀察金融機構生成式 AI 之導入策略可以發現，呈現顯著的「風險導向」特徵。依據技術對金融業務之實質影響程度、個人資料之涉入深度，以及內部機密之敏感性[207]，目前金融業的 AI 應用場景大致可歸納為三大類別，且在發展進程上呈現高度的集中化現象。

一、　內部行政與營運輔助型應用類型：高度採用生成式 AI

依前述之案例分析可知，臺灣金融機構目前導入生成式 AI 應用樣態具有高度共通性：其一，系統主要供內部人員使用，不直接對外提供資訊或建議；其二，生成內容並未直接構成金融交易、授信、投資或商品銷售之最終決策；其三，金融機構在制度設計上，多仍要求人員對生成內容進行審閱與確認。從風險性質與法制評價的角度觀察，內部行政與營運輔助型應用之所以在金融實務與政策討論中，普遍被視為生成式 AI 的低風險切入點，並非僅因其技術複雜度較低，而係源於其對金融決策與個人權益之影響程度相對有限。相較於直接介入授信核准、投資建議或交易執行之系統，此類內部應用原則上僅影響組織內部之作業效率與資訊處理方式，其風險後果多由金融機構自身承擔，而不立即外溢至金融消費者或市場層次。

現行金融機構在此類應用中普遍仍保留人為審閱與最終確認機制，生成式 AI 僅作為內容生成或資訊整理工具，而非自動作成決定之主體。此種人機分工結構，使其在形式上不構成自動化決策，亦不容易直接引發歧視、誤導或不公平交易等典型金融監理關切。因此，不論在實務導入策略或風險治理討論中，內部行政與營運輔助型生成式 AI，經常被視為一個可供組織逐步學習、測試與建立治理經驗的初階場域。

二、　對客互動與資訊提供型應用類型：中高度採用生成式 AI

相較於前述僅限於組織內部使用之行政與營運輔助型應用，部分臺灣金融機構已開始將生成式 AI 部署於直接面向金融消費者之互動與資訊提供場域。此類應用主要出現在銀行與證券業者之數位通路中，包含智能客服系統、線上即時問答介面，以及用於回應消費者關於金融商品、投資資訊或理財知識之生成式內容模組。例如，銀行導入生成式 AI 智能客服之案例，顯示該等系統已可即時回應消費者對帳戶服務、產品條件、

[207] 金融監督管理委員會，金融業運用人工智慧(AI)指引，頁 4：https://www.fsc.gov.tw/websitedowndoc?file=chfsc%2F202408231741530.pdf&filedisplay=%E9%99%84%E4%BB%B6_%E9%87%91%E8%9E%8D%E6%A5%AD%E9%81%8B%E7%94%A8AI%E6%8C%87%E5%BC%95.pdf（最後瀏覽日：2026 年 7 月 5 日）。

費用結構或交易流程之詢問；而在理財或投資相關應用中，亦可見生成式 AI 被用於整理市場資訊、說明投資選項或提供一般性之理財知識內容。此類系統雖多被業者定位為「資訊提供」或「客服輔助」，而非直接提供投資建議或交易指示，然其共同特徵在於，生成內容已直接進入金融機構與消費者之互動關係中。

生成式 AI 一旦被部署於對客互動場域，即使該等系統並未直接替消費者作成投資、授信或交易決定，其所回應與生成之資訊內容，仍可能對消費者之理解、預期與行為選擇產生實質影響。此一風險性質之改變，主要根源於金融服務本身所具有之高度專業性與不對稱性。

首先，金融商品與投資資訊本質上具有高度複雜性，消費者往往仰賴金融機構所提供之說明與資訊，作為判斷風險與利益的重要依據。當生成式 AI 被置於客服或資訊回應位置時，其語言能力與即時性，反而可能強化消費者對該內容之信賴程度。即便業者主張其僅提供一般性資訊，消費者在實際互動情境中，未必能清楚區分「資訊說明」與「建議意味」之界線。

其次，與內部行政應用不同，對客互動型生成式 AI 已嵌入金融機構與消費者之信賴關係之中。消費者並非與中立工具互動，而是基於對金融機構專業能力與合規義務之期待，接受其所提供之資訊。當生成內容出現錯誤、偏頗或過度簡化時，其後果不僅是資訊瑕疵，更可能引發誤信、誤判，甚至在事後產生責任歸屬與爭議解決上的困難。因此，即使形式上未構成自動化決策，對客互動與資訊提供之生成式 AI，仍可能因其實質影響消費者決策環境，而成為金融監理與消費者保護法制高度敏感之應用場景。

三、　決策輔助與評估支援型應用類型：低度採用生成式 AI

前述關於銀行於授信與信用風險分析流程中，運用生成式 AI 協助彙整客戶資料、生成風險摘要或模擬不同授信情境之案例，係於核心金融判斷流程之前端或中介階段，導入生成式 AI 作為決策輔助與評估支援工具。此外，在證券與資產管理相關場域，前述提及金融機構運用生成式 AI 整理研究資料、摘要市場資訊、排序投資標的，或作為投資決策前之分析支援工具。此類系統未直接執行下單或核貸動作，但已在事實上影響決策者可得資訊的範圍、呈現方式與注意焦點。因此，凡屬生成式 AI 被嵌入授信、風險管理或投資決策支援流程之情境，即構成本項分析的核心對象。

生成式 AI 於本項之實質金融決策（如信用評分、授信額度核准、風險定價等）之應用比例極低，多數金融機構對此仍抱持高度觀望態度。從法令遵循與內部稽核的標準觀察，營運決策無可避免地將同時觸及「客戶個人機敏資料」、「核心交易數據」以及「內部商業機密」三大高

風險領域。此外，決策結果將直接影響客戶之金融實體權益，在生成式 AI 仍面臨黑盒特性的可解釋性不足[208]與幻覺[209]等技術瓶頸的情況下，金融機構於缺乏明確監理標準前，多不願貿然將生成式 AI 引入核心決策流程，以免發生不可預期之違規裁罰與客訴賠償風險。

[208] 李昀璇，註 191。

[209] 孫彬訓，銀行導入生成式 AI 內用先行，工商時報：https://www.ctee.com.tw/news/20241218700140-439901（最後瀏覽日：2026 年 7 月 5 日）。

第三章　歐盟 AI 監管法制

第一節　General Data Protection Regulation (GDPR)

《General Data Protection Regulation》（GDPR）為歐盟個人資料保護之基本規範，自 2018 年 5 月 25 日起適用[210]。有鑑於金融機構使用 AI 系統通常涉及個人資料之自動化分析與決策，本研究僅就與自動化處理、剖析及資料主體權利保障直接相關之條文加以整理，包括適用範圍、資料處理原則與合法基礎、自動化決策規範及資料保護影響評估等規定。

一、　適用範圍

GDPR 第 2 條與第 3 條分別規範其實質與地域適用範圍。依第 2 條規定，本規則適用於全部或部分自動化處理之個人資料，亦適用於非自動化處理之個人資料，只要該等資料構成歸檔系統之一部分，或擬成為歸檔系統之一部分[211]。換言之，凡透過自動化方式進行之個人資料分析、評估或建檔活動，原則上均屬 GDPR 之規範範圍。第 2 條並列舉若干排除情形，例如處理活動不屬於歐盟法適用範圍、涉及共同外交與安全政策，或屬於純粹個人或家庭活動者，則不在本規則適用之列[212]。

在地域適用方面，第 3 條規定凡資料控制者或處理者於歐盟境內設有機構者，其於該機構活動中所進行之個人資料處理，即受 GDPR 規範，不論實際處理行為是否發生於歐盟境內[213]。此外，即使控制者或處理者未於歐盟境內設立機構，只要其處理活動涉及向歐盟境內之資料主體提供商品或服務，或監控其在歐盟境內之行為，亦屬本規則之適用範圍[214]。另依第 3 條第 3 項，在依國際公法適用會員國法律之特定情形下，相關個人資料處理活動亦適用 GDPR[215]。

二、　資料處理原則與合法基礎

GDPR 第 5 條確立個人資料處理之基本原則。依該條規定，資料處理須具合法性、公平性與透明性，並應基於特定、明確且合法之目的進行，不得與原目的不相容。資料應適當且限於達成目的所必要之範圍，並保持正確與安全，且不得保存超過必要期間。資料控制者並負有責任證明其遵守上述原則（問責原則）[216]。

---

[210] Council Regulation 2016/679, art. 99(2), 2016 O.J. (L 119) 1, 88.
[211] *Id.* art. 2(1).
[212] *Id.* art. 2(2)(a)–(c).
[213] *Id.* art. 3(1).
[214] *Id.* art. 3(2)(a)–(b).
[215] *Id.* art. 3(3).
[216] *Id.* art. 5(1)(a)–(f), 5(2).

在此基礎上，第 6 條規定個人資料處理之合法基礎。資料處理僅於具備法定事由時方屬合法，包括資料主體同意、履行契約所必要、履行法律義務、保護重大利益、執行公共任務，或基於控制者或第三方之合法利益且未凌駕於資料主體基本權之情形[217]。若處理係基於法律義務或公共任務，其法律依據應明確載於歐盟法或會員國法律[218]。另若進一步處理目的與原蒐集目的不同，且非基於資料主體同意或法律規定時，控制者應依第 6 條第 4 項評估其相容性[219]。

三、 剖析與自動化決策

GDPR 將以自動化方式評估個人特徵之處理活動納入「剖析」（profiling）之範圍。依第 4 條第 4 款，凡利用個人資料進行自動化分析，以評估自然人之經濟狀況、行為或其他個人面向者，即屬剖析[220]。此一定義涵蓋透過演算法進行信用評分、風險分級或其他個人化評估之情形。

復參第 22 條規定，資料主體有權不受僅基於自動化處理（包括剖析）所作，且對其產生法律效果或類似重大影響之決定所拘束[221]。法律效果通常指對權利義務之直接影響；類似重大影響則包括對經濟狀況、信用評價或其他重要生活面向產生實質影響之決定。

就此，Article 29 Data Protection Working Party 所通過、並經 European Data Protection Board（EDPB）於 2018 年正式背書之指引（WP251rev.01）進一步指出，所謂「僅基於自動化處理」之決定，係指決策過程中並無具實質意義之人為介入；若控制者僅在形式上保留人工審查程序，實際上卻未由具備權限與能力之人員真正檢視、評估並得改變決定結果，則不得據此主張該決策不屬於第 22 條規範範圍[222]。該指引並認為，第 22 條所稱「類似重大影響」亦包括足以對個人之金融取得、就業機會、健康、教育或其他重要生活面向造成重大影響之決定，故若信用評分、授信核准、保費定價或其他金融條件係由自動化系統作成，且足以影響個人取得金融服務之機會，即可能落入第 22 條之規範範圍[223]。

第 22 條同時設置三種例外情形，包括契約必要、法律授權及明確同意，但即使符合例外，控制者仍須提供人為介入與救濟機制，以確保資料主體得參與決策過程[224]。上開指引亦指出，控制者應提供有關決策邏

[217] *Id.* art. 6(1)(a)–(f).
[218] *Id.* art. 6(3).
[219] *Id.* art. 6(4).
[220] *Id.* art. 4(4).
[221] *Id.* art. 22(1).
[222] Article 29 Data Protection Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, WP251rev.01*, at 21–22 (Feb. 6, 2018)
[223] *Id.* at 22–23.
[224] GDPR, *supra* note 202, art. 22(2)–(3).

輯之「有意義資訊」，使資料主體得理解該決策之基本形成方式、主要考量因素及其對自身可能產生之影響[225]。

此外，第 22 條第 4 項規定，自動化決策原則上不得以第 9 條第 1 項所列特殊類型個人資料為基礎，除非符合第 9 條第 2 項第(a)或(g)款之條件，即資料主體已「明確同意」處理該等特殊類型資料，或為重大公共利益（Substantial Public Interest）所必要，並基於歐盟法或會員國法律，且該法律應規定適當保障措施[226]。

2023 年 12 月 7 日，歐盟法院（Court of Justice of the European Union, CJEU）首次就 GDPR 第 22 條之適用範圍作出解釋。本案源於德國信用評分機構 SCHUFA Holding AG 向一名借款人 OQ 之往來銀行提供負面信用評分，導致其貸款申請遭拒[227]。OQ 向黑森邦資料保護及資訊自由專員（HBDI）提出申訴，惟遭駁回[228]。案件上訴至威斯巴登行政法院後，該院向 CJEU 申請先決裁判，請求解釋 GDPR 第 22 條之適用範圍。CJEU 裁定信用評分機構於以自動化處理方式計算信用還款機率值，且貸款機構高度依賴該評分值以決定是否締結、履行或終止契約之情形下，即構成第 22 條所稱之「自動化個人決策」[229]。CJEU 明確認定第 22 條之規範對象不限於作成最終決定之機構（即銀行），同樣及於在決策鏈中提供決定性輸入資訊之上游機構（即信用評分機構）[230]。換言之，即便形式上係由銀行作成最終核貸決定，只要其實際上幾乎完全依賴信用機構之評分結果，則信用評分機構本身亦須承擔第 22 條之合規義務。

2024 年 11 月 22 日，義大利資料保護機關就 Foodinho S.r.l.對超過 35,000 名外送員非法處理個人資料之情事，裁處歐元 500 萬元罰鍰[231]。Foodinho 存在多項重大 GDPR 違規事項，包括：Foodinho 透過數位平台採用自動化演算法系統（Excellence Score）對外送員之工作表現進行評分，並據此分配工作班次及訂單，惟該系統欠缺任何人為介入機制，外送員既無法向人工審核人員申請複查，亦無法對系統決定提出異議[232]。此舉直接違反 GDPR 第 22 條所規定之人為介入權、表達意見權及決定爭議權；Foodinho 不僅在外送員工作時間外仍持續追蹤其地理位置，亦在未事先告

[225] WP251rev.01, *supra* note 214, at 25–26.
[226] GDPR, *supra* note 202, art. 9(1), 9(2)(a), 9(2)(g), 22(4).
[227] Case C-634/21, *OQ v. Land Hessen (SCHUFA Holding)*, ECLI:EU:C:2023:957, [2023] (Dec. 7, 2025).
[228] *Id.*
[229] *Id.; see also Case C-634/21, SCHUFA Holding (Scoring), GDPRHUB,* https://gdprhub.eu/index.php?title=CJEU_-C%E2%80%91634%2F21-SCHUFA_Holding(Scoring) (last visited Jul. 5, 2026).
[230] Case C-634/21, *supra* note 219.
[231] EUROFOUND, *Italy: Privacy Authority Fines Foodinho 5 Million Euro for Rider Data Breaches*, PLATFORM ECONOMY DB. (July 16, 2021).
[232] *Id.*

知外送員之情況下，將個人資料（包括地理位置資訊）分享予第三方機構，嚴重違反透明性原則與資料主體知情權[233]。

上述裁罰案例雖均涉及傳統演算法系統，然其所確立之合規原則對金融機構導入生成式 AI 之情境同樣具有直接參考意義。生成式 AI 因大型語言模型之黑盒特性，較傳統評分系統更難提供 GDPR 所要求之「有意義資訊」，亦更易使業務人員因模型輸出之高度擬真性而形成實質依賴，導致形式上存在人工複核、實質上卻未達有效人為介入之合規缺口。此外，SCHUFA 案所確立之決策鏈上游責任原則，於金融機構仰賴外部 AI 供應商或第三方基礎模型提供決策支援之情境下，亦意味著責任歸屬不因形式上由機構作成最終決定而得以免除。換言之，生成式 AI 之技術特性並未使其脫離第 22 條之規範射程，反而因可解釋性不足與人機協作邊界模糊，使相關合規義務之履行更具挑戰性。

四、 資訊揭露義務

GDPR 第 13 條至第 15 條建立資料主體之資訊取得與存取權架構，規範控制者於蒐集或處理個人資料時之資訊揭露義務，以及資料主體對其資料處理情形之查詢權利[234]。

第 13 條與第 14 條分別規範在「資料直接自資料主體蒐集」與「資料非自資料主體取得」兩種情形下，控制者應提供之資訊內容，旨在確保資料主體知悉：處理之目的與法律依據、資料保存期間、接收者類別、跨境傳輸情形，以及其所享有之權利[235]。若處理涉及自動化決策或剖析，控制者並應提供有關該處理存在之事實、所涉邏輯之有意義資訊，以及該處理對資料主體可能產生之重要性與後果[236]。

第 15 條進一步賦予資料主體主動存取其個人資料之權利，資料主體得確認其資料是否正在被處理，並取得該資料之副本及相關處理資訊，包括處理目的、資料類別、保存期間與是否存在自動化決策[237]。若處理涉及自動化決策或剖析，第 15 條規定亦要求控制者提供有關所涉邏輯及可能後果之資訊[238]。

五、 資料保護影響評估

GDPR 第 35 條建立資料保護影響評估（Data Protection Impact Assessment, DPIA）制度，要求在特定高風險處理情形下，控制者於處理

---

[233] *Id.*
[234] GDPR, *supra* note 202, arts. 13–15.
[235] *Id.* arts. 13(1)–(2), 14(1)–(2).
[236] *Id.* arts. 13(2)(f), 14(2)(g) (requiring controllers to provide, where automated decision-making including profiling is involved, "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject").
[237] *Id.* art. 15(1)(a)–(e).
[238] *Id.* art. 15(1)(h).

前對該等活動對個人資料保護之影響進行評估[239]。該義務適用於考量處理之性質、範圍、情境及目的後，可能對自然人之權利與自由產生高度風險之情形[240]。

第 35 條第 3 項列舉須進行評估之典型情形，包括：基於自動化處理（含剖析）對自然人個人面向進行系統性且廣泛之評估，並依該評估作出對其產生法律效果或類似重大影響之決定；大規模處理特殊類型個人資料；以及對公開可進入區域進行大規模且系統性監控[241]。依第 35 條第 7 項規定，資料保護影響評估應至少包括：對預計處理活動及其目的之說明、對處理必要性與比例性之評估、對資料主體權利與自由所面臨風險之分析，以及為處理風險所規劃之保障措施與安全措施[242]。若處理活動之風險情況發生變化，控制者並應進行審查[243]。

第二節　Digital Operational Resilience Act（DORA）

《Digital Operational Resilience Act》（下稱 DORA）於 2022 年 11 月 28 日獲歐盟理事會正式通過，最終於 2022 年 12 月 27 日公布於歐盟官方公報[244]。

一、　立法背景及立法目的

隨著科技發展，歐盟金融部門因高度數位化，導致資訊與通訊科技（Information and Communication Technology, ICT）風險日益升高。然而，原有歐盟法規對 ICT 安全和數位韌性的處理分散、不一致，且側重於財務韌性而非全面的營運韌性，會員國間的立法差異也阻礙了跨境金融服務的順暢運作[245]。故 DORA 旨在整合並升級所有關於 ICT 風險的要求，建立一個統一且全面的數位營運韌性框架，其目標係確保金融實體能有效管理、應對和從 ICT 相關事件中復原，從而維護歐盟金融市場的穩定性與完整性，並保護投資人及消費者權益[246]。該法規透過統一的 ICT 風險管理、事件通報、韌性測試、資訊共享和第三方風險管理等規定，來減少監管複雜性，促進監管協調，並提高法律確定性[247]。

---

[239] *Id.* art. 35(1).
[240] *Id.*; *see also* Article 29 Data Protection Working Party, *Guidelines on Data Protection Impact Assessment (DPIA) and Determining Whether Processing Is "Likely to Result in a High Risk" for the Purposes of Regulation 2016/679*, WP 248REV.01, at 6–7 (Oct. 4, 2017) (endorsed by EUROPEAN DATA PROTECTION BOARD, Endorsement 1/2018 (May 25, 2018) (identifying nine criteria to assess whether processing is likely to result in high risk requiring a DPIA).
[241] GDPR, *supra* note 202, art. 35(3)(a)–(c).
[242] *Id.* art. 35(7)(a)–(d).
[243] *Id.* art. 35(11).
[244] Council Regulation 2022/2554, 2022 O.J. (L 333) 1.
[245] *Id.* Recitals 5, 8–10.
[246] *Id.* Recitals 11–12; *id.* art. 1(1).
[247] *Id.* Recitals 13–14.

二、 適用範圍

DORA 適用範圍涵蓋廣泛之金融實體，包括銀行、投資公司、支付機構、電子貨幣機構、中央交易對手、交易場所、證券結算系統及其他受歐盟金融監管體系規範之機構[248]。除金融實體本身外，向該等實體提供 ICT 服務之第三方服務供應商，亦納入規範架構，特別是被認定為關鍵 ICT 第三方者，將受特定監督制度約束[249]。

三、 金融實體之治理與風險管理框架

DORA 第 5 至 16 條建立金融實體之 ICT 風險管理框架，金融實體須設計並維持一套完整且持續運作之 ICT 風險管理制度，涵蓋資產識別、風險評估、保護控制、持續監控、事件應變、業務持續與復原機制[250]。管理機構須定義並批准該風險管理架構，監督其執行，並確保投入必要之資源，並負擔管理金融實體 ICT 風險之最終責任[251]。相關策略需設定 ICT 風險容忍度，並說明數位中斷對關鍵功能之影響承受範圍[252]。管理機構亦須定期審查該架構，並在重大 ICT 事件後重新檢視其有效性[253]。

在制度運作層面，金融實體首先須建立 ICT 資產清冊，辨識所有資訊系統、資料與相互依存關係，並依其對關鍵功能之重要性進行分類[254]。在此基礎上，建立存取控制、資料保護、加密、監控與漏洞管理機制，以防止未授權存取與系統干擾[255]。同時設置持續監測與預警機制，以即時辨識異常活動與潛在威脅[256]。當 ICT 相關事件發生時，金融實體須依既定程序進行分類、升級與處理，並啟動應變與復原計畫，包括業務持續安排與資料復原措施[257]。復原計畫應事先設定復原時間與復原點目標，並定期進行測試與演練[258]。事件處理完成後，金融實體須進行事後檢討，分析中斷原因與制度缺口，並將相關發現納入風險評估與策略調整之中[259]。資深 ICT 人員應定期向管理機構報告相關發現與改進建議[260]。

此外，第 17 至 19 條建立重大 ICT 相關事件之分類與通報制度。金融實體須對所有 ICT 事件進行記錄與分類，並於重大事件發生時，向主管機關通報；如事件影響客戶財務利益，亦須通知客戶[261]。第 16 條則對特定

[248] *Id.* art. 2(1)(a)–(t).
[249] *Id.* art. 2(1)(u); *id.* arts. 31–44.
[250] *Id.* arts. 5–16.
[251] *Id.* art. 5(2)(a), (d).
[252] *Id.* art. 6(8)(b).
[253] *Id.* art. 6(5).
[254] *Id.* art. 8(1), (4).
[255] *Id.* art. 9(2)–(4).
[256] *Id.* art. 10(1)–(2).
[257] *Id.* art. 11(1)–(3).
[258] *Id.* art. 12(6); *id.* art. 11(6).
[259] *Id.* art. 13(2)–(3).
[260] *Id.* art. 13(5).
[261] *Id.* arts. 17–19.

小型金融實體設置比例適用之簡化要求，但仍須維持基本之安全與監控機制[262]。

四、 數位營運韌性測試制度

第四章要求金融機構將測試制度化，金融實體必須依其風險概況與關鍵功能的重要性，建立一套定期執行的測試機制[263]。測試結果不得停留於技術報告，而須進入內部優先排序、補救與驗證程序，並由管理層級掌握其改善情形，使測試與治理形成閉環，而非一次性檢驗[264]。

此外，本章將測試分為一般測試與進階測試兩個層次。前者涵蓋日常風險評估與技術檢驗，包括漏洞評估與掃描、開放原始碼分析、網路安全評估、差距分析、實體安全審查等[265]；後者則針對特定風險程度較高之金融實體，要求以威脅導向滲透測試（Threat-Led Penetration Testing, TLPT）方式模擬真實攻擊情境，並於正式運作系統上驗證防禦與應變能力，且至少每三年執行一次[266]。測試範圍並不限於內部系統，若關鍵功能依賴外部ICT服務提供者，相關底層系統亦可能納入測試；即使第三方參與測試，金融實體仍對整體合規負責[267]。測試完成後，於測試報告與補救計畫達成共識後，金融機構及（如適用）外部測試人員應向指定之主管機關提交測試摘要、補救計畫及證明測試符合規定之文件[268]。主管機關應向金融機構出具確認測試符合規定之證明文件，該證明得作為主管機關間相互承認測試結果之依據[269]。

五、 ICT 第三方風險管理原則

金融實體須依比例原則管理ICT第三方風險，並依服務之性質與關鍵性程度區分管理強度[270]。為此，金融實體須建立ICT第三方風險策略，並維持一份持續更新之契約安排資訊登錄冊，記錄所有使用ICT服務之契約，並區分支援關鍵或重要功能之安排[271]。主管機關得要求提供完整登錄冊或特定部分，以利監督[272]。

在訂立契約前，金融實體須進行風險評估與盡職調查，包括評估該服務是否涉及關鍵或重要功能、是否可能導致集中風險、是否存在利益衝突，以及該提供者之資訊安全能力[273]。若涉及關鍵或重要功能，金融

[262] *Id.* art. 16(1)–(2).
[263] *Id.* art. 24(1), (3).
[264] *Id.* art. 24(5).
[265] *Id.* art. 25(1).
[266] *Id.* art. 26(1)–(2).
[267] *Id.* art. 26(3).
[268] *Id.* art. 26(6).
[269] *Id.* art. 26(7).
[270] *Id.* art. 28(1)(b).
[271] *Id.* art. 28(2)–(3).
[272] *Id.* art. 28(3).
[273] *Id.* art. 28(4).

實體尚須規劃退出策略，確保在契約終止、服務失敗或品質惡化時，不致中斷營運或影響法規遵循與客戶服務[274]。

契約內容方面，金融實體與ICT第三方服務提供者之權利義務須明確書面化，並載明服務範圍、資料處理地點、資料保護條款、事件協助義務、終止條款及主管機關配合義務[275]。對支援關鍵或重要功能之契約，尚須納入具體服務水準目標、重大事件通報義務、業務持續與安全測試配合義務、查閱與稽核權，以及退出與過渡安排[276]。

六、 關鍵ICT第三方之監督制度

DORA將特定ICT第三方服務提供者納入聯盟層級的監督架構。關鍵提供者之指定須考量該等提供者可能帶來的潛在系統性風險，以及ICT第三方集中度所造成之風險，指定標準涵蓋服務失敗對金融穩定之系統性衝擊、依賴該提供者之系統重要性金融機構數量、相關服務之可替代程度等因素[277]。主導監督者會特別關注相互依存的程度，並識別高集中度可能對聯盟金融體系穩定性和完整性造成壓力的具體情況[278]。一旦指定為關鍵ICT第三方服務提供者，該提供者須接受聯盟層級的監督活動，並由一名主導監督者負責，該主導監督者將依本法規之規定任命[279]。主導監督者之職責在於持續監控關鍵ICT第三方服務提供者的活動，並協助金融監管機構量化、評估和補救關鍵ICT第三方服務提供者所造成的ICT風險後果[280]。為確保跨部門的一致性，歐洲監理機構（ESAs）被賦予權力，透過其聯合委員會設立監督論壇[281]。DORA 強調，主導監督者應與關鍵ICT 第三方服務提供者保持對話，以應對可能對聯盟金融體系穩定性和完整性造成壓力之特定風險[282]。

七、 違規裁罰框架

DORA 第 50 條課予各會員國義務，須制定適當之行政罰鍰與補救措施，並確保其有效執行，相關罰鍰須符合有效性、比例性與嚇阻性之要求[283]。第 51 條進一步規定，主管機關於決定行政罰鍰之種類與金額時，須考量違規之故意或過失程度，以及違規之重大性、嚴重性、持續期間、責任主體之財務實力、所獲利益或所避免損失之規模、對第三方造成之

[274] *Id.* art. 28(8).
[275] *Id.* art. 30(1)–(2).
[276] *Id.* art. 30(3).
[277] *Id.* art. 31(2)(a)–(d).
[278] *Id.* Recital 67.
[279] *Id.* art. 31(1)(b).
[280] *Id.* art. 33(1)–(2).
[281] *Id.* art. 32(1).
[282] *Id.* Recital 67.
[283] *Id.* art. 50(3).

損害，以及責任主體與主管機關之合作程度等因素[284]。就關鍵ICT第三方服務提供者之違規而言，主導監督者得依第 35 條第 6 項課予每日定期罰款，以迫使其遵守透明度及查閱相關義務，罰款金額以關鍵ICT第三方服務提供者前一營業年度全球日均營業額之 1%為上限，且最長不得超過六個月[285]。

就非財務措施而言，第 50 條第 4 項授權主管機關發布勒令停止違規行為之命令、要求暫停或永久停止特定業務行為、採取確保金融實體持續遵法之各類措施，以及發布揭示違規法人或自然人身分及違規性質之公開聲明[286]。第 54 條規定，主管機關須於其官方網站公告無爭訟之行政罰鍰決定，以確保透明度與公眾問責[287]。此外，第 52 條允許會員國就嚴重違規行為另行訂定刑事處罰，並授權主管機關與司法機關相互傳遞資訊，形成行政裁罰與刑事追訴並行之雙軌執法機制[288]。

八、 執法現況

DORA自2025年1月17日起正式適用，各會員國主管機關之行政裁罰權於同日啟動[289]。然而，由於DORA第50條採取原則性授權而非統一訂定明確罰鍰上限，各會員國在行政罰鍰之最高金額及罰鍰計算細節上存在相當大的差異，跨境營運之金融機構於進行合規風險評估時，須特別留意所在地法律對裁罰金額的影響[290]。2025 年 11 月 18 日，EBA、EIOPA及ESMA聯合公布首批19家關鍵ICT第三方服務提供者名單，包含 AWS、Microsoft Azure、Google Cloud 等超大規模雲端供應商，以及Bloomberg、London Stock Exchange Group、IBM、Tata Consultancy Services、Orange 等金融資料與技術服務商，上述提供者已受聯合審查[291]。此外，上開雲端供應商亦為現階段金融機構部署生成式 AI 應用之主要平台，其受監督之情形或可作為後續觀察生成式AI治理發展之參考背景。

第三節 Consumer Credit Directive II（CCD II）

《Consumer Credit Directive (EU) 2023/2225》（下稱CCD II）係歐盟理事會與歐洲議會於 2023 年 10 月 18 日通過之消費者信貸協議指令，旨

[284] *Id.* art. 51(2).
[285] *Id.* art. 35(6)–(8).
[286] *Id.* art. 50(4).
[287] *Id.* art. 54(1)–(2).
[288] *Id.* art. 52.
[289] *Id.* art. 64.
[290] DLA Piper, *DORA Penalty Regimes: Overview of Divergence Among Member States*, https://www.dlapiper.com/en-us/insights/publications/2025/10/divergence-in-administrative-penalties-under-dora (last visited Jul. 5, 2026).
[291] Orbiq, *DORA Compliance: Complete Guide*, https://www.orbiqhq.com/eu-regulations/dora-compliance (last visited Jul. 5, 2026).

在取代並廢除原 2008/48/EC 指令，其立法背景主要源於數位化帶來的市場發展，尤其線上環境中新信貸產品的迅速普及，原指令 2008/48/EC 因其條文措辭不精確及未涵蓋某些信貸市場層面，在確保高標準消費者保護和促進單一信貸市場方面效益有限，導致各成員國間監管框架出現碎片化，進而限制了消費者受益於跨境信貸的機會，並造成保護水平的不一致[292]。新版指令旨在提升消費者信任與保護，並促進跨境活動的發展[293]。因此，CCD II 在維持消費者保護架構的基礎上，擴大了適用範圍，並強化信用評估義務與資訊透明機制[294]。

一、 適用範圍之擴張與數位化信貸納入

相較於原指令，CCD II 縮減排除範圍，將部分原本因金額過低（如 200 歐元以下）或期間過短而未納入規範之信用產品納入調整[295]。蓋因數位化導致了市場的重大發展，催生了新的產品並改變了消費者的行為與偏好，這些現象在 2008/48/EC 指令通過時未能預見[296]。例如，新的信貸產品特別是在線上環境中大量出現，這使得 2008/48/EC 指令的適用範圍產生了法律上的不確定性[297]。因此，立法者藉由擴大適用範圍的方式，將這些新型信貸模式（包括那些因高成本或未付款高額費用而可能對消費者造成損害的產品）納入規範，以確保消費者能夠獲得更高的保護和透明度，進而提高消費者信心[298]。

二、 前契約資訊揭露義務

CCD II 持續並強化締約前資訊揭露制度，要求債權人應在消費者受契約拘束的「足夠時間前」，提供標準化的《標準歐洲消費者信貸資訊表》（Standard European Consumer Credit Information, SECCI）[299]。透過此標準化格式，旨在有效降低資訊不對稱，使消費者能有充分時間閱讀、理解資訊，比較不同信貸方案並做出知情決定[300]。此資訊表內容主要包含：信貸總額、年利率（APR）、總成本、付款安排及逾期後果等[301]。

此外，為確保透明度，如果消費者所獲得的報價價格是基於自動化處理（包括分析/profiling）而個性化的，債權人或信貸中介機構必須以清

[292] Council Directive 2023/2225, recals. 3, 5–6, 2023 O.J. (L 2225) 1.
[293] *Id.* recitals 9–10.
[294] *Id.* recitals 9–10, art. 1.
[295] *Id.* recital 15; *see also id.* art. 2(1).
[296] *Id.* recital 4.
[297] *Id.* recital 7.
[298] *Id.* recital 15–16.
[299] *Id.* art. 10(1)–(2), annex I.
[300] *Id.* recital 36.
[301] *Id.* annex I, Part I.

晰易懂的方式，明確告知消費者此一情況，以便消費者在做出購買決策時能考慮潛在風險[302]。

三、 信用能力評估義務

為防止不負責任的放貸行為和消費者過度負債，CCD II 要求債權人於訂立信貸契約前，必須基於充分且適當資訊，評估消費者履行契約之能力。該義務具有強制性，不得透過契約排除或限制[303]。

此外，CCD II 要求成員國確保債權人建立並記錄其信用評估程序，並妥善保存相關資訊[304]。債權人僅在信用能力評估結果顯示消費者有能力履行信貸協議下的義務時，才能向其提供信貸[305]。若消費者被拒絕信貸，債權人必須立即且免費告知消費者信用評估資料庫查詢的結果[306]。一旦信貸協議簽訂後，債權人不得以信用評估不正確為由，單方面取消或修改協議，除非能夠證明消費者故意隱瞞或偽造了資訊[307]。

四、 自動化處理之銜接規範

在數位授信環境下，信用能力評估往往透過演算法模型完成。CCD II 並未自行建構完整之自動化決策制度，其明文規定，其關於自動化處理之條文皆在「不影響 GDPR 之前提下」實施，強調本指令在消費者資料保護上與 GDPR 的一致性與補充性[308]。

（一） 個性化報價之揭露義務

當債權人或信用仲介機構依賴自動化處理對特定消費者或消費者類別進行「價格個人化」時，必須以清晰易懂的方式事前告知消費者[309]。此資訊揭露須明確標示所呈現之價格係基於個人資料自動化處理之結果，以協助消費者於決策時評估潛在的演算法溢價風險。同時，配合 GDPR 第 14 條第 2 項第(f)款之精神，債權人與信用仲介機構亦有義務告知消費者用於個人化報價之資料來源[310]。

（二） 信用能力評估中之自動化決策保障

[302] *Id.* art. 13; recital 46.
[303] *Id.* art. 18(1).
[304] *Id.* art. 18(4).
[305] *Id.* art. 18(6).
[306] *Id.* art. 19(6).
[307] *Id.* art. 18(7).
[308] *Id.* recital 30; *see also* Council Regulation 2016/679, art. 5, 2016 O.J. (L 119) 1 (EU).
[309] GDPR recital 46.
[310] *Id.* recital 46; GDPR art. 14(2)(f).

CCD II 立法時，歐盟執委會已於 2021 年提出 AI Act 提案，並將用於自然人信用評分之 AI 系統列為高風險類型[311]，CCD II 進一步在信貸實務中落實對應之當事人程序保障。當信用能力評估涉及自動化處理時，債權人必須確保消費者：(1)有權要求債權人提供真人實質審查、(2)有權取得針對評估結果及自動化處理運作方式之「有意義且可理解的解釋」內容應涵蓋主要變數、運作邏輯、涉及之風險，以及該決策對消費者之意義與實質影響、(3)有權向債權人表達自身立場或補充有利資訊、(4)有權針對信用評估結果及是否核貸之決定請求重新審查[312]。債權人必須主動告知消費者上述權利；特別是當信貸申請遭自動化決策拒絕時，必須明示告知其權利及提出異議（Contest）之具體程序[313]。

（三） 資料處理之邊界

CCD II 確立信用能力評估不應僅基於消費者過去的信用歷史，而應進行全面性之財務考量[314]。在處理個人資料以進行評估時，必須以適當、相關，並以達成處理目的所必要者為限[315]。CCD II 更進一步全面禁止在信用評估中使用特種個人資料（如健康數據），並明文排除從「社群網路（Social networks）」爬取或獲得之資訊，藉此防堵演算法歧視與過度侵犯隱私[316]。

五、 執法現況

CCD II 依 Article 48 規定，各成員國須於 2025 年 11 月 20 日前完成國內轉置（Transposition），並自 2026 年 11 月 20 日起全面適用，原 2008/48/EC 指令同日廢止[317]。由於 CCD II 直至 2026 年 11 月方才對金融機構正式發生效力，現階段尚無任何金融機構因直接違反 CCD II 而遭裁罰之案例。

第四節 EU Artificial Intelligence Act（AI Act）

一、 立法目的與適用範圍

[311] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), annex III, COM (2021) 206 final* (Apr. 21, 2021); CCD II recal. 56.
[312] CCD II art. 18(8)(a)–(c).
[313] *Id.* art. 18(8), (9).
[314] *Id.* art. 18(11); recital 55.
[315] *Id.* art. 18(3); recital 55.
[316] *Id.* art. 18(3); art. 19(5); recital 55, 57.
[317] CCD II art. 47–48.

歐盟 AI Act 於 2024 年 8 月 1 日正式生效[318]，其立法目的在於建立一套統一之法律框架，以規範 AI 系統於歐盟境內之開發、上市與使用，促進內部市場之順暢運作，避免因會員國規範差異而造成市場碎片化，並提升業者之法律確定性[319]。其次，該法以人為本與可信賴 AI 為核心原則，確保 AI 之發展與運用不致侵害健康、安全及歐盟基本權利憲章所保障之基本權利。立法者特別意識到 AI 可能對個人及公共利益造成物質或非物質損害，因而透過風險分級制度，對高風險 AI 系統設定較為嚴格之義務，以維持高水準之權利保障。此外，AI Act 亦透過提供明確之規範架構，支持創新發展，並強化歐盟在全球 AI 治理中的領導地位。

AI Act之適用對象涵蓋整個AI供應鏈[320]，無論是否設立於歐盟境內，凡開發或以其名義將AI系統或通用AI模型投放市場之提供者（Providers），均受本法規範。其次，在歐盟境內使用AI系統之部署者（Deployers），只要其使用屬於專業或業務活動範圍，亦屬適用對象[321]。此外，授權代表、進口商與經銷商等供應鏈參與者，同樣被納入規範範圍並負有相應義務。

在排除範圍方面，軍事、國防或國家安全目的之 AI 系統原則上不適用本法，因該等領域屬會員國專屬責任[322]。然而，如相關系統同時或後續被用於民用、執法或其他非排除目的，則仍可能落入本法適用範圍。部署者於個人非專業活動中使用 AI 系統者，亦不在規範之列。至於依自由與開源授權（Free and Open-source Licenses）發布之 AI 系統，原則上不適用本法，但若其被投放市場為高風險用途，或屬於特定禁止或受監管之系統，則仍應遵守相關規定[323]。

二、 風險分級制度

AI Act 並未僅針對特定產業，而係以風險分級（Risk-Based Approach）為核心，將 AI 系統依其對基本權、健康、安全及公共利益可能產生之影響程度區分為不同風險層級，並對高風險系統課予較為嚴格之義務[324]。

（一） 不可接受風險（Unacceptable Risk）

不可接受風險為風險級別最高的類別，屬於不可接受風險之 AI 系統，因其本質上對基本權利或民主秩序構成重大威脅，而被全面禁止。此類系統包括：(1) 透過潛意識操控，或蓄意操縱或欺騙性技術的人工智慧系統；(2) 利用自然人或特定群體脆弱性（因其年齡、殘疾或特定

[318] *Regulation 2024/1689 of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, recals. 1, 4–5, 2024 O.J. (L 1689) 1.
[319] *Id.* recal. 27; arts. 1, 3(1).
[320] *Id.* art. 2(1)(a)–(c).
[321] *Id.* art. 3(4).
[322] *Id*. arts. 2(3), 2(4).
[323] *Id*. art. 2(6); recal. 22.
[324] *Id.* recal.s 14–17; art. 9.

社會或經濟狀況）；（3）進行社會評分（social scoring）；（4）僅基於對自然人之分析或性格特徵評估而預測其犯罪風險之人工智慧系統；（5）透過非目標性方式自網際網路或閉路電視影像抓取人臉圖像，以建立或擴展人臉辨識資料庫；（6）於工作場所或教育機構中使用人工智慧系統推斷自然人之情緒等用途[325]。

（二） 高風險 AI 系統（High-Risk AI Systems）

AI Act 規定凡屬高風險之 AI 系統，均須遵守一套強制性合規義務，包括風險管理、資料治理、技術文件保存、人類監督機制、透明性要求，以及合格評定程序等[326]。

高風險 AI 系統主要分為兩大類：第一，AI 系統作為歐盟既有產品安全立法所規範產品之安全組件，或本身即屬需經第三方合格評定之產品，此類規範多涉及機械設備、醫療器材或交通安全系統[327]；第二，AI 系統屬於附錄三所列舉之特定用途，例如：生物識別（Biometrics）、獲取和享用基本私人服務及基本公共服務和福利（Access to and enjoyment of essential private services and essential public services and benefits）、執法（Law enforcement）、司法和民主程序（Administration of justice and democratic processes）等。附錄三明確指出若干對基本權具有重大影響之應用場景，凡屬該等用途，應被視為高風險[328]。

（三） 非高風險 AI 系統與透明義務

對於不屬於高風險，但仍可能影響使用者知情權或判斷之 AI 系統，AI Act 採取透明度（Transparency）導向之規範模式，旨在確保使用者知悉其正在與 AI 互動，或所接觸之內容係由 AI 所生成，從而避免混淆與欺瞞。依據該法第 50 條，金融機構在應用生成式 AI 時，須於以下情境下履行相關透明度義務：

1、提供者必須確保與自然人直接互動的 AI 系統在設計和開發上能夠告知自然人，他們正在與一個 AI 系統互動，除非從一個理性、消息靈通、觀察仔細且謹慎的自然人角度來看，這種互動是顯而易見的，並考慮到使用情境和背景[329]；

2、若 AI 系統生成了構成一般公眾利益資訊的合成音訊、影像、視訊或文字內容，部署者必須以機器可讀（Machine-readable）格式進行標記，並可被偵測為人工生成或操縱的[330]；

[325] *Id.* arts. 5(1)(a)–(g).
[326] *Id.* arts. 6, 9–15, 17, 26.
[327] *Id.* art. 6(1); Annex I.
[328] *Id.* art. 6(2); Annex III, items 1, 5, 6, 8.
[329] *Id.* art. 50(1).
[330] *Id.* art. 50(2).

3、若 AI 系統生成或操縱的圖像、音訊或視訊內容，與現有人物、物體、地點或其他實體極為相似，致使人誤認為真實（即深偽技術），部署者必須揭露該內容是人工生成或操縱的[331]。

（四） 其他非高風險 AI 系統

若 AI 系統不屬於被禁止實踐，亦不屬於高風險或特定透明義務範圍之 AI 系統，AI Act 不施加任何強制性的法律義務，企業可自由開發與使用。歐盟僅鼓勵業者採行自願性行為準則與最佳實務以促進負責任之使用[332]。

三、 生成式 AI 與通用型 AI 之特別規範

AI Act 並未以「生成式 AI」（Generative AI）作為獨立風險分類，而係以「通用型 AI 模型」（General-Purpose AI Models, GPAI models）為規範核心。AI Act 特別增訂通用型 GPAI 之規範章節，並進一步區分一般 GPAI 與具系統性風險（Systemic Risk）之 GPAI[333]。

（一） 所有通用型 AI 模型(All GPAI Models)之基礎義務

無論風險等級為何，所有 GPAI 模型供應商（如 OpenAI、Google）皆須建立並更新技術文件，並向整合該模型的下游供應商（如開發金融客服 App 的銀行或軟體商）提供充分資訊，使其能理解模型能力與限制；遵守歐盟著作權法，特別是關於文字與資料探勘（Text and Data Mining, TDM）之保留聲明；發布一份「足夠詳細的摘要（Detailed Summary）」，揭露用於訓練該模型的內容，以解決數據來源合法性與潛在偏見問題[334]。

（二） 具系統性風險之通用型 AI 模型(GPAI Models with Systemic Risk)

AI Act 進一步在通用型 AI 模型之中區分「具系統性風險」之模型。若模型依客觀技術標準被認定具有高度影響能力，例如其訓練所使用之累積運算能力（cumulative compute）達到一定門檻，原則上即推定其具有系統性風險。[335]此外，歐盟執委會亦得依技術能力、應用範圍及潛在社會影響等標準，主動指定特定模型屬於具系統性風險之類型[336]。

[331] *Id.* art. 50(4).
[332] *Id.* recal. 46; art. 95.
[333] *Id.* recals. 97–101, arts. 51–55.
[334] *Id.* art. 53(1)(a)–(c).
[335] *Id.* art. 51(1).
[336] *Id.* art. 51(2).

所謂「系統性風險」，係指模型之能力或規模可能對公共安全、基本權保障、市場秩序或民主程序產生廣泛且跨領域之影響。此類風險不僅來自特定應用場景，而係源於模型本身之通用性與可擴散性質[337]。

對於具系統性風險之 GPAI 模型，提供者須履行較一般通用型模型更為嚴格之義務，包括：模型評估，即應根據符合最新技術水平的標準化協議和工具進行模型評估，包括實施和記錄對抗性測試，以識別和緩解系統性風險；建立持續性的系統性風險評估與緩解機制，評估模型可能導致之大規模錯誤資訊傳播、歧視性輸出、資安風險或其他重大社會影響，並採取適當控制措施；確保模型具備充分之資安防護能力，並於發生嚴重事故時向歐盟人工智慧辦公室通報[338]。

四、 裁罰機制

就裁罰架構而言，AI Act 第 99 條確立三級裁罰制度：違反禁止性規定者，最高可處 3,500 萬歐元或全球年度總營業額 7%之罰鍰（以較高者為準）；違反高風險 AI 系統相關要求者，最高可處 1,500 萬歐元或全球年度總營業額 3%之罰鍰；提供不正確或誤導性資訊者，最高可處 750 萬歐元或全球年度總營業額 1%之罰鍰[339]。上開罰鍰金額超越 GDPR 之上限[340]，並得與 GDPR、DORA 等其他歐盟法規之裁罰於不同違規事實下同時併行之[341]。

截至本研究撰寫時，AI Act 尚未有對金融機構公開之正式裁罰決定。AI Act 對金融業之信用評分、保險定價、貸款審核及反洗錢風險分析等金融業常見 AI 應用，均被明確列入附錄三高風險 AI 系統範疇，自 2026 年 8 月起須遵守完整合規義務[342]。

第五節 小結

歐盟對 AI 於金融領域之監管，係同時從基本權保障、交易秩序、營運安全及風險治理等不同層面進行制度設計，透過資料保護法、金融部門法、數位營運韌性規則與 AI 專門立法所共同構成之規範架構。

其 GDPR 構成整體架構之基礎層，凡金融機構使用 AI 系統涉及個人資料之蒐集、分析、剖析或自動化決策，即須受 GDPR 之合法性、透明性、目的限制與資料最小化等原則拘束。尤其第 22 條對僅基於自動化處理所作

[337] *Id.* recal. 110; art. 51(1).
[338] *Id.* art. 55(1)(a)–(d).
[339] *Id.* art. 99(3)–(5).
[340] *EU AI Act & Data Sovereignty: A Guide for Financial Institutions*, RASA：https://rasa.com/blog/eu-ai-act-financial-services. (Apr. 8, 2026),
[341] *EU AI Act Fines and Penalties: What Non-Compliance Will Cost You*, MATPROOF：https://matproof.com/blog/eu-ai-act-fines-penalties (Mar. 22, 2026).
[342] *Id.* art. 6(2); annex III; *see also 2026 EU Financial Services Compliance*, NORTAL, https://nortal.com/insights/eu-financial-services-compliance (last updated 2026).

且對個人具有法律效果或類似重大影響之決定設有限制，並透過資訊揭露義務與資料保護影響評估制度，將資料主體權利保障嵌入決策流程之中。GDPR 所處理者，係金融 AI 運作所依賴之個人資料處理基礎及其程序界線。

DORA 則關注金融機構在高度數位化環境下，如何維持 ICT 系統之安全、持續性與可復原性。從治理與風險管理框架、韌性測試，到第三方風險管理與關鍵 ICT 第三方之直接監督，DORA 將金融機構對數位基礎設施之依賴轉化為可識別、可測試、可監督之法定義務。CCD II 則進一步將數位化與自動化處理之議題帶入消費信貸場景。其一方面透過擴大適用範圍與強化締約前資訊揭露義務，回應線上信貸與新型態消費信貸產品之興起；另一方面，則透過信用能力評估義務、個性化報價之揭露，以及涉及自動化處理時之程序保障，將消費者保護、資料處理透明與信用決策之正當性要求結合於同一制度場域中。

至於 AI Act，則代表歐盟開始以 AI 系統本身作為獨立之規範對象，建立橫跨各產業之風險分級與合規義務架構。其透過禁止性規範、高風險系統義務、透明要求及對通用型 AI 模型之特別規範，使 AI 自原本散見於各部門法之監管對象，進一步成為具有統一法定分類與義務內容之規範客體。在金融領域中，若 AI 系統涉及信用評分、基本金融服務取得或其他對個人權利具有重大影響之用途，即可能落入高風險 AI 系統之範疇，而受更嚴格之事前與持續性合規要求。因此，若從法體系觀察，歐盟之監管結構並非由 AI Act 取代既有法制，而是形成分層分工之架構，各法規雖有不同規範標的與功能定位。

第四章　美國 AI 監管法制

美國 AI 監管法制呈現「分散式治理（Decentralized Governance）」，在聯邦層級以政策工具與既有法律適用為主，並由多個主管機關依職權進行功能性監理與執法；同時，州與地方政府透過 AI 專門立法或情境化規則補足特定風險場景之合規要求。

第一節　聯邦層級

在聯邦行政體系內，白宮與跨部會機制多以行政命令或政策方案設定治理方向，其他主管機關則依職權進行功能性監理與執法。

一、　行政部門的治理工具

（一）　總統行政命令（Executive Orders, E.O.）

1. E.O. 14110（Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence）[343]：

E.O. 14110 係由美國前總統拜登（Joseph R. Biden, Jr.）於 2023 年 10 月 30 日發布，其直接規範主體為聯邦行政機關。各行政機關須於其既有法定職權範圍內，依該命令之指示採取相應行動，以回應 AI 所引發之風險與治理需求[344]。

就適用範圍而言，E.O. 14110 所涵蓋者為聯邦行政機關於其職掌事項中，涉及 AI 之開發、使用、採購與監督等相關活動[345]。在規範內容上，E.O. 14110 並未設定具體禁止或裁罰規定，而係透過任務指派、期限設定與跨機關協調機制，要求行政機關建立可操作之治理與風險管理措施[346]。其具體作法包括：命令相關機關制定政策指引、推動標準化風險管理工具、調整政府採購與使用規範，以及運用既有法律權限處理 AI 相關風險[347]。該行政命令並未將 AI 視為獨立之規範客體，而係要求行政機關在促進 AI 發展與創新之同時，必須同步處理安全（Safety）、資訊安全（Security）、民權（Civil Rights）與消費者保護（Consumer Protection）等公共利益，並要求行政機關於其法定職權範圍內建構可操作的風險治理機制[348]。

---

[343] *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,* Exec. Order No. 14110, 88 Fed. Reg. 75191 (Oct. 30, 2023).
[344] *Id.* at 75191.
[345] *Id.* at 75191, 75218.
[346] *Id.* at 75191.
[347] *Id.* at 75201, 75218-220.
[348] *Id.* at 75217-221.

此外，該行政命令於「Policy and Principles」部分明確提出八項指導原則，分別為[349]：（1）促進負責任的創新與競爭（Promoting Innovation and Competition）；（2）支持勞工與勞動權益（Supporting Workers）；（3）推進平等與民權（Advancing Equity and Civil Rights）；（4）AI 必須安全且具資訊安全保障（Safe and Secure）；（5）保護消費者（Consumer Protection）；（6）維護隱私與公民自由（Privacy and Civil Liberties）；（7）確保政府自身使用 AI 之有效治理（Government Use of AI）；（8）強化國際領導與合作（International Leadership）。

2. E.O. 14365 (Ensuring a National Policy Framework)：

E.O. 14365 由美國總統川普（Donald J. Trump）於 2025 年 12 月 16 日發布，其目的在於建立一套國家 AI 政策框架，以統整聯邦政府於 AI 領域之政策方向、治理原則與跨機關協調機制。該命令之制定，係延續川普政府於 2025 年初發布之 E.O. 14179 所揭示之政策轉向方針[350]。
E.O. 14179 之政策目的在於調整聯邦 AI 政策取向，強調移除可能被認為阻礙美國維持 AI 領導地位之既有政策與行政措施，明確撤回前任政府於 2023 年發布之 E.O. 14110，並要求各聯邦機關立即檢視其依 E.O. 14110 所採取之政策、指令、規則或其他行動，若與新政策方向不一致，則應採取暫停、修正或撤回等調整措施[351]。除撤回既有政策外，E.O. 14179 亦提出若干具體行政推進機制，例如命令要求在發布後 180 日內制定「人工智慧行動計畫（AI Action Plan）」，以落實新的政策目標並促進美國 AI 之發展與國際競爭力；同時，命令亦指示白宮管理及預算局（OMB）於 60 日內調整包括 OMB M-24-10 與 M-24-18 在內之跨部會政策備忘錄，使其與本命令之政策方向一致[352]。

美國國會近年雖曾嘗試透過多項跨黨派法案建立統一之聯邦 AI 監管架構，並意圖藉此排除州層級規範以實現統一治理。惟因國會立法程序長期陷入僵局，行政部門遂轉而仰賴擴張之總統行政權限，作為主導 AI 監管政策之核心工具，E.O. 14365 即是在各州紛起立法（如加州 AI 法案）導致監管碎片化的背景下，為尋求聯邦一體化管理而發布[353]。

就規範主體而言，E.O. 14365 之直接拘束對象仍為聯邦行政機關。該命令要求各機關於其既有法定職權範圍內，依循國家 AI 政策框架所揭示之原則與目標，檢視並調整其 AI 相關之政策、計畫與行政措施，以確保聯邦政府於 AI 之開發、使用與監督活動中，具備一致性與可協調之治理

[349] *Id*. at 75191-193.
[350] *Ensuring a National Policy Framework for Artificial Intelligence*, Exec. Order No. 14365, 90 Fed. Reg. 58499 (Dec. 11, 2025).
[351] *Removing Barriers to American Leadership in Artificial Intelligence*, Exec. Order No. 14179, 90 Fed. Reg. 8741 (Jan. 23, 2025).
[352] *Id*.
[353] Exec. Order No. 14365, 90 Fed. Reg. 58499–501 (Dec. 11, 2025).

方向[354]。在適用範圍上，E.O. 14365 仍以聯邦行政機關於其職掌事項中，涉及 AI 之開發、部署、使用、採購及監督等相關活動為限，該命令並未直接對私部門或一般人民創設新的法律義務[355]。

在規範內容上，E.O. 14365 並未設定具體禁止或裁罰規定，而係著重於制度設計與政策協調層面。其主要作法包括：要求建立跨機關之政策協調機制、明確國家層級之 AI 治理目標、強化聯邦政府內部對 AI 使用之治理與監督，以及確保相關政策措施得與既有法律架構相互銜接[356]。

3. NIST Artificial Intelligence Risk Management Framework: Generative AI Profile（NIST AI 600-1）

有鑑於生成式 AI 具有不同於傳統型 AI 的特殊風險，美國國家標準與技術研究院（National Institute of Standards and Technology, NIST）於 2024 年發布 Artificial Intelligence Risk Management Framework: Generative AI Profile（NIST AI 600-1）。該文件係建立於 Artificial Intelligence Risk Management Framework 1.0（AI RMF 1.0）[357]提出之風險治理架構之上[358]。

NIST AI 600-1 延續 AI RMF 1.0 所採取之四大核心功能治理（Govern）、情境界定（Map）、衡量（Measure）與管理（Manage），但 NIST AI 600-1 進一步針對生成式 AI 因內容生成性、輸出不可預期性及高度依賴訓練資料與模型推論所衍生之風險加以具體化。相較於傳統以分類、預測或推薦為主之 AI 系統，生成式 AI 所產生之文字、圖像或建議內容，往往具有高度擬真性，易被使用者誤認為事實或專業判斷，從而放大錯誤輸出、誤導資訊與責任歸屬不明等問題[359]。

NIST AI 600-1 特別指出 12 項生成式 AI 獨有的風險類別，包括重大危害資訊與能力（CBRN Information or Capabilities）、虛構與幻覺(Confabulation/Hallucination)、危險、暴力或仇恨內容（Dangerous, Violent, or Hateful Content）、資料隱私（Data Privacy）、環境衝擊（Environmental Impacts）、偏見與同質化（Harmful Bias or Homogenization）、人機互動配置風險（Human-AI Configuration）、資訊

[354] *Id.* at 58499.
[355] *Id.*
[356] *Id.* at 58499–501.
[357] NIST 於 2023 年 1 月發布之 Artificial Intelligence Risk Management Framework, AI RMF 1.0，提供一套供政府機關及私部門參考之非強制性風險管理指引，用以協助組織在設計、開發、部署或使用 AI 系統時識別、評估及管理可能風險。AI RMF 1.0 列舉七項 AI 系統之信賴度特徵（trustworthiness characteristics），包括有效且可靠（valid and reliable）、安全性（safe）、資安與韌性（secure and resilient）、可歸責與資訊透明度（accountable and transparent）、可解釋性與可詮譯性（explainable and interpretable）、隱私保護（privacy-enhanced）以及公平性與有害偏見管理（fair—with harmful bias managed）。此外，本框架亦提出以四大功能為主軸之核心架構，即治理（Govern）、對照（Map）、量測（Measure）及管理（Manage）。
[358] NAT'L INST. OF STANDARDS & TECH., NIST AI 600-1, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile* 1 (2024).
[359] *Id.* at 1-2.

完整性（Information Integrity）、資訊安全（Information Security）、智慧財產權（Intellectual Property）、猥褻、貶抑或濫用內容（Obscene, Degrading, and/or Abusive Content）、供應鏈與組件整合風險（Value Chain and Component Integration）[360]。組織於設計、開發與部署相關系統時，應將風險管理納入整體 AI 生命週期之核心考量，而非僅以模型效能或準確率等技術性指標作為評估依據。具體而言，相關治理措施應涵蓋生成內容之來源與歷程追蹤（Content provenance）[361]、系統知識範圍與限制之揭露（Knowledge limits）[362]，以及透過人機互動配置（Human-ai configuration）[363]、人類監督與對抗性測試（如 Red-teaming）[364]等方式，強化人類對 AI 輸出之理解與控制。

NIST AI 600-1 雖然並未針對金融業建立具體之法律義務，而係以跨產業適用之風險治理框架形式，提出一般性之技術治理原則。惟於金融應用情境下，上述原則得轉化為具體之監理與合規要求，例如：要求金融機構於對客互動時適當揭露其使用生成式 AI 之情形，以避免消費者誤認[365]；透過建立內容來源追溯機制（如浮水印或數位簽章）以提升金融文件之可驗證性並降低深偽詐欺風險[366]；以及運用對抗性測試等方法，於模型部署前後持續檢驗其在特定應用領域（如金融法規解釋）之風險與侷限。

二、 既有聯邦法律對 AI 於金融產業之適用

（一） Equal Credit Opportunity Act

《平等信用機會法》（Equal Credit Opportunity Act, ECOA）為一聯邦法規，其明文禁止任何債權人於信用交易之任何方面，基於種族、膚色、宗教、國籍、性別、婚姻狀況、年齡，或申請人依法行使消費者權利等受保護事由，於信用申請之核准、拒絕或授信條件設定上，對申請人為不利之差別待遇[367]。

從法規文義觀之，ECOA 對差別待遇之禁止，係以授信結果是否對特定受保護群體產生不利影響作為判斷基準，而未區分信用決策係由人工

[360] *Id.* at 4-5.
[361] *Id.* at 52.
[362] *Id.* at 24, 32.
[363] *Id.* at 16-45.
[364] *Id.* at 50-51; red-teaming 為一種結構化之對抗性測試機制。不同於傳統資安著重於網路與系統權限之滲透，AI red-teaming 測試係由專業測試人員扮演惡意攻擊者，透過「提示詞注入（Prompt Injection）」、「越獄（Jailbreak）」或刻意輸入極端邊緣案例（Edge Cases）等手段，主動誘騙、攻擊或測試大型語言模型之邊界。其目的在於系統正式部署前，強制暴露出模型可能產生之機密資料外洩、有害內容輸出、幻覺或演算法偏見。
[365] *Id.* at 27.
[366] *Id.* at 19, 33-34.
[367] Equal Credit Opportunity Act § 701(a), 15 U.S.C. § 1691(a).

判斷、統計模型或自動化系統作成[368]。是以，金融機構於授信審查過程中使用 AI 進行信用評分、風險排序或核准建議，其相關行為仍屬 ECOA 所規範，並不因決策過程之自動化或技術複雜性而排除法律適用。

除禁止歧視性信用交易外，ECOA 及其施行細則 Regulation B 亦進一步課予債權人於拒絕信用申請或提供不利信用交易條件時，須向申請人提供具體理由之說明義務（Adverse action notice）[369]。此一規範設計顯示，信用交易決策之合法性不僅取決於最終結果是否構成差別待遇，亦涉及債權人是否能就其決策基礎提出具體且可理解之說明。故縱使債權人採用高度複雜之演算法或 AI 系統，亦不得因模型運作不透明或技術性質而免除其法定說明義務。

承前所述，ECOA 與其施行細則已就信用交易中之反歧視義務及不利處分說明義務建立明確規範。然而隨著演算法與 AI 系統被廣泛導入信用審查流程，實務上逐漸出現金融機構是否得以模型複雜性、專有性或第三方技術供應為由，主張其無法充分說明信用決策基礎，進而降低合規責任之疑義。針對此一趨勢，美國消費者金融保護局（Consumer Financial Protection Bureau, CFPB）於 2022 年發布 Circular 2022-03，CFPB 於該 Circular 中指出，金融機構不能僅僅因為其用於評估申請的技術過於複雜或晦澀難懂，作為未能履行法定說明義務之理由。無論模型係由金融機構自行開發，抑或委由第三方提供，只要該模型被用以影響消費者之信用決策，其最終合規責任仍應由金融機構承擔[370]。

2025 年 7 月，麻薩諸塞州檢察長（Massachusetts Attorney General, MAG）就學生貸款業者 Earnest Operations LLC 採取執法行動，此案明確援引 ECOA 就 AI 授信承保模型進行追訴並以和解告終。Earnest 係一採用演算法模型作成學生貸款再融資授信決策之業者，其模型於核准、拒絕及定價各環節均高度仰賴自動化運算[371]。MAG 指控 Earnest 以申請人就讀學校之「同期違約率（Cohort Default Rate, CDR）」作為加權評分變數。MAG 指出此一變數間接導致黑人及西班牙裔申請人，蓋其中許多人曾就

[368] 15 U.S.C. § 1691(a) (prohibiting discrimination in "any aspect of a credit transaction"). *See* also Consumer Fin. Prot. Bureau, Consumer Financial Protection Circular 2022-03, Adverse Action Notification Requirements in Connection with Credit Decisions Based on Complex Algorithms (May 26, 2022), https://www.consumerfinance.gov/compliance/circulars/circular-2022-03-adverse-action-notification-requirements-in-connection-with-credit-decisions-based-on-complex-algorithms/ (clarifying that the use of artificial intelligence or complex machine-learning models does not exempt creditors from ECOA compliance).

[369] 15 U.S.C. § 1691(d); Regulation B, 12 C.F.R. § 1002.9(a)–(b).

[370] CONSUMER FIN. PROT. BUREAU, Consumer Financial Protection Circular 2022-03, *Adverse Action Notification Requirements in Connection with Credit Decisions Based on Complex Algorithms* (May 26, 2022).

[371] Assurance of Discontinuance Part I, In re Earnest Operations LLC, No. 2584-cv-01895 (Mass. Super. Ct. (July 10, 2025). *See* also Press Release, Massachusetts Attorney General Andrea Joy Campbell, AG Campbell Announces $2.5 Million Settlement With Student Loan Lender For Unlawful Practices Through AI Use, Other Consumer Protection Violations：https://www.mass.gov/news/ag-campbell-announces-25-million-settlement-with-student-loan-lender-for-unlawful-practices-through-ai-use-other-consumer-protection-violations (2025).

讀歷史上服務低收入族群之院校，在相似財務條件下更可能遭拒絕或獲得不利條件，構成以種族為基礎之差別影響，違反 ECOA § 701(a)[372]。此外，Earnest 以未持有永久居留身分作為自動拒絕依據，涉及以國籍為由之歧視，同屬 ECOA 明文禁止之行為[373]。再者，Earnest 所提供之拒絕理由通知因系統設計缺陷，未能揭示真正影響決策之變數，未符合 ECOA 及 Regulation B 之說明義務。此案最終以 250 萬美元達成和解，並要求其須建立公平貸款風險管理機制、對 AI 模型進行差別影響測試[374]。

上述案例雖係針對傳統演算法授信模型之執法行動，然其所確立之合規要求對金融機構導入生成式 AI 之情境同樣具有參考意義。

（二） OCC Bulletin 2011-12 與 2025-26

1. OCC Bulletin 2011-12

美國通貨監理局(Office of the Comptroller of the Currency, OCC)與聯準會共同發布 OCC Bulletin 2011-12，旨在為銀行提供關於有效模型風險管理的全面指導，凡銀行依賴模型[375]所做成的金融決策，原則上均應納入該指引所建構之模型風險管理（Model Risk Management, MRM）架構中加以治理[376]。各銀行落實該指引的方式可因情形而異，實務適用應依銀行風險暴露、業務活動，以及模型使用的複雜性與廣度而調整；並以社區銀行為例，指出其採用模型較少且複雜度中等時，所需步驟可顯著低於大型銀行[377]。此外，指引要求銀行應對其使用供應商產品進行驗證，而模型驗證（Validation）則強調驗證須具有一定程度的獨立性（通常由非開發/非使用者執行或至少受獨立方嚴格審查）、驗證人員應具備足夠專業能力，且驗證的嚴謹程度應與模型之複雜度與重大性、以及銀行整體營運規模與複雜度相稱[378]。

值得注意的是，該指引要求銀行至少應對每一模型進行定期檢視（至少每年一次，必要時更頻繁），以判斷模型是否如預期運作、既有驗證是否足夠；並提到一般良好實務是讓所有模型在某一固定週期接受完整驗證流程。同時，其亦具體列舉「全面性驗證」三大核心元素：概念健全性評估（含開發證據）、持續監控（含流程驗證與基準比較）、結果分析（含回溯測試）[379]。就外部資源而言，指引亦指出模型驗證同樣適

---

[372] *Id.* Part III.B.
[373] *Id.* Part III.C.
[374] *Id.* Part III.D.
[375] 模型（Model）指一種應用統計、經濟、金融或數學理論、技術和假設，將輸入數據處理為量化估計的方法、系統或途徑，並包含輸入、處理和報告三個組成部分
[376] BOARD OF GOVERNORS OF THE FED. RESERVE SYS. & OFFICE OF THE COMPTROLLER OF THE CURRENCY, OCC 2011-12, *Supervisory Guidance on Model Risk Management* 1-2 (Apr. 4, 2011).
[377] *Id.* at 5.
[378] *Id.* at 10, 16.
[379] *Id.* at 10, 12.

用於外部供應商或顧問提供/開發之模型，且政策應載明供應商模型之驗證要求與納入整體框架之方式[380]。

然而，在實務適用上，部分社區銀行曾將該指引理解為具有高度處方性要求，例如將固定週期完整驗證解讀為必須每年全面驗證。為澄清此類理解偏差，OCC 於 2025 年發布 OCC Bulletin 2025-26[381]。

2. OCC Bulletin 2025-26

OCC 發布此公告旨在強調，社區銀行管理層在實際應用 OCC 模型風險管理指南時，應根據銀行的風險敞口、業務活動以及模型使用的複雜性和範圍進行客製化。使用相對較少且複雜度適中的模型的社區銀行，其模型風險管理活動可能遠少於模型使用範圍更廣或更複雜的銀行。同樣，社區銀行的模型驗證頻率通常也低於模型使用範圍更廣、更複雜的大型銀行[382]。

此外，OCC 明確表示：其模型風險管理指引不應被解讀為要求社區銀行進行年度模型驗證，OCC 不會僅因銀行根據自身風險敞口、業務活動以及模型使用的複雜性和範圍合理確定的模型驗證頻率或範圍而向其提供負面監管反饋[383]。

（三） SEC 對自動化投資建議與演算法互動之監理

在投資建議與財富管理領域，美國證券交易委員會（Securities and Exchange Commission, SEC）雖尚未建立專門針對 AI 之成文法規則，然其逐步透過投資顧問法（Investment Advisers Act of 1940），處理自動化投資建議、演算法互動及數位投資介面所生之合規問題[384]。

SEC 於 2017 年即透過幕僚指引（Staff Guidance）（非法律層級但極具指標性參考）指出，機器人投顧（Robo-Advisers）作為註冊投資顧問，仍受投資顧問法之實質義務與受託義務拘束[385]。該指引特別聚焦於數位投資顧問模式所衍生之問題，包括：是否已充分揭露演算法提供建議之方式與限制、是否蒐集足夠資訊以支持適當之投資建議、演算法問卷或客戶介面設計是否足以反映客戶之投資目標與風險偏好，以及其合規計畫是否足以處理透過技術提供建議所產生之新型風險。SEC 強調即使投資建議係透過自動化方式生成，投資顧問對其向客戶提供之建議內容與程序設計，仍須負擔既有法律責任[386]。

---

[380] *Id.* at 16, 18, 20.

[381] OFFICE OF THE COMPTROLLER OF THE CURRENCY, OCC Bulletin 2025-26, *Model Risk Management: Clarification for Community Banks*： https://www.occ.gov/news-issuances/bulletins/2025/bulletin-2025-26.html (Jul. 5, 2026).

[382] *Id.*

[383] *Id.*

[384] U.S. Sec. & Exch. Comm'n, Robo-Advisers, *IM Guidance Update No. 2017-02*：https://www.sec.gov/investment/im-guidance-2017-02.pdf (Feb. 2017).

[385] *Id.* at 1, 2, 6, 10.

[386] *Id.* at 1-8.

2018 年 SEC 即首次對機器人投顧業者採取執法行動，分別對 Hedgeable, Inc.與 Wealthfront Advisers LLC 作成處分。SEC 認定，Hedgeable 因對投資績效進行具誤導性之比較，且僅選取少數高績效帳戶作為宣傳基礎，而構成不實或誤導性陳述；Wealthfront 則因其聲稱透過系統全面監控客戶帳戶以避免洗售（wash sale），實際上卻未執行該機制，導致大量帳戶發生洗售情形，並同時涉及違反廣告與合規義務。此等案件顯示，即便投資顧問係透過演算法與自動化系統提供服務，其對資訊揭露之真實性、投資建議之適當性以及內部合規機制之有效性，仍須負完全法律責任，並不因技術介入而有所減輕[387]。

其後，SEC 檢查部門於 2021 年發布有關電子化投資建議（electronic investment advice）之風險提醒。該文件係基於實際檢查經驗所提出，適用對象包括向退休計畫或零售投資人提供機器人投顧服務之顧問機構、向第三方授權數位投資平台之業者，以及為互動式數位投資平台提供顧問或次顧問服務之機構[388]。SEC 在檢查中所觀察到之問題，主要集中於揭露內容與實際運作不一致、演算法與投資方法之說明不足、問卷設計與客戶風險輪廓間之連結不明、持續監督與合規測試不足，以及數位行銷與宣稱內容可能造成誤導等事項[389]。

此外，SEC 於 2023 年曾提出關於預測性數據分析(predictive analytics)、及其他類似技術之規則草案，擬要求經紀交易商(broker-dealers)與投資顧問(investment advisers)就其使用相關技術與投資人互動所生之利益衝突，建立特定之評估與處理機制，以避免機構利用數位互動、個人化推送或演算法優化，使自身利益優先於投資人利益[390]。該提案所欲處理之核心問題，在於機構若利用技術工具預測、引導或影響投資人行為，則其技術設計本身可能形成證券法上之衝突來源。雖然該提案已於 2025 年遭 SEC 正式撤回，並未成為現行有效規則，但其仍反映出美國證券監理機關已將預測性數據分析、演算法推薦及類似 AI 工具視為投資人互動與衝突管理問題之一環。

SEC 於 2024 年 3 月 18 日對 Delphia (USA) Inc.及 Global Predictions, Inc.提出執法行動，此為 SEC 就投資顧問之 AI 相關揭露所提出之首批執

[387] U.S. Sec. & Exch. Comm'n, *SEC Charges Two Robo-Advisers with False Disclosures* (Dec. 21, 2018), https://www.sec.gov/newsroom/press-releases/2018-300.
[388] SEC. DIV. OF EXAMINATIONS, Risk Alert, *Observations from Examinations of Advisers that Provide Electronic Investment Advice* 1：https://www.sec.gov/files/exams-eia-risk-alert.pdf (Nov. 9, 2021).
[389] Id at 3, 5, 7.
[390] U.S. Sec. & Exch. Comm'n, *SEC Proposes New Requirements to Address Risks to Investors from Conflicts of Interest Associated with the Use of Predictive Data Analytics by Broker-Dealers and Investment Advisers*：https://www.sec.gov/newsroom/press-releases/2023-140 (July 26, 2023).

法案例，亦是 SEC 明確以漂 AI（AI washing）[391]為由追究聯邦證券法反詐欺條款責任之首開先例[392]。

就 Delphia 之違規事實而言，SEC 命令認定，Delphia 自 2019 年至 2023 年間，在其向 SEC 申報之揭露文件、對外發布之新聞稿及公司網站上，就其聲稱運用 AI 與機器學習整合客戶數據以進行投資決策一事，作出虛假與具誤導性之陳述；Delphia 宣稱其「將集體數據轉化為 AI 訓練資料，藉以預測即將崛起之公司與趨勢，並搶先投資」，然實際上 Delphia 並不具備其所聲稱之 AI 與機器學習能力。Delphia 於 2021 年 SEC 例行檢查期間已自承尚未建立相關演算法，並進行若干更正性揭露，然其後仍繼續就演算法在投資流程中之運用發出不實陳述，包括於 2022 年發布新聞稿，聲稱其「專有演算法結合成員投資數據與商業可用數據」，惟彼時相關演算法實際上並不存在[393]。

就 Global Predictions 而言，SEC 認定 Global Predictions 在其網站、社群媒體及致現有與潛在客戶之電子郵件中，虛稱自身為「首家受監管之 AI 金融顧問」，並在公開網站上不實聲稱其技術採用「AI 驅動之專家預測」；SEC 進一步認定，Global Predictions 無法提出文件以實質支撐其所謂「首家受監管 AI 金融顧問」之聲稱。此外，Global Predictions 亦被認定未能揭露其與特定提供推薦語（testimonials）人士之利害關係所形成之重大利益衝突，以及在未建立並落實投資顧問法（Investment Advisers Act of 1940）第 206(4)條授權 SEC 所訂定之行銷規則，所要求之政策與程序之情況下，即於公開網站刊登假設性績效數據[394]。

在裁罰結果上，Delphia 與 Global Predictions 分別同意接受 SEC 之譴責及停止並糾正令；Delphia 同意支付 225,000 美元之民事罰款，Global Predictions 同意支付 175,000 美元之民事罰款，合計 400,000 美元[395]

（四） Federal Trade Commission Act § 5

---

[391] *AI Washing*, WIKIPEDIA, https://en.wikipedia.org/wiki/AI_washing (last visited Jul. 5, 2026) (定義「漂 AI」為一種商業欺詐手法，指企業為了拉抬股價、吸引投資或迎合市場潮流，誇大甚至虛假宣傳其產品、服務或營運中使用了 AI 技術，本質上只是利用行銷話術將傳統技術或人力包裝成「AI 驅動」)。

[392] *In the Matter of Delphia (USA) Inc.*, INVESTMENT ADVISERS ACT RELEASE NO. 6573, at 1 (Mar. 18, 2024), https://www.sec.gov/files/litigation/admin/2024/ia-6573.pdf; *In the Matter of Global Predictions, Inc.*, INVESTMENT ADVISERS ACT RELEASE NO. 6574, at 1 (Mar. 18, 2024), https://www.sec.gov/files/litigation/admin/2024/ia-6574.pdf. *See also* Press Release, SEC, SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence (Mar. 18, 2024).

[393] *Delphia (USA) Inc.*, INVESTMENT ADVISERS ACT RELEASE NO. 6573, at 3-5.

[394] *In the Matter of Global Predictions, Inc.*, INVESTMENT ADVISERS ACT RELEASE No. 6574, at 3–6 (Mar. 18, 2024).

[395] *In the Matter of Delphia (USA) Inc.*, INVESTMENT ADVISERS ACT RELEASE No. 6573, at 8 (Mar. 18, 2024); *In the Matter of Global Predictions, Inc.*, INVESTMENT ADVISERS ACT RELEASE No. 6574, at 8 (Mar. 18, 2024).

聯邦交易委員會（Federal Trade Commission, FTC）亦逐步將 AI 所產生之誤導、欺罔與冒充風險，納入其既有消費者保護與不公平競爭執法架構中。FTC 之監理重點在於企業是否利用 AI 生成內容、演算法互動或自動化代理機制，對消費者作出不實、誤導或足以造成損害之表示。FTC 於 2023 年即公開提醒企業應注意 AI 相關宣稱之真實性，不得就 AI 產品或服務之能力、效能（例如宣稱其 AI 演算法能保證帶來高額投資報酬）或不可替代性作出欠缺基礎之誇大表示，亦不得僅以「使用 AI」作為吸引消費者之行銷標籤[396]。

此外，FTC 特別關注生成式 AI 所帶來之深偽（Deepfakes）、語音仿冒與冒充風險，此類技術常被不法分子用於金融詐欺（例如冒充銀行客服、理財專員或親友騙取資金）。2024 年，FTC 之「政府與企業冒充規則」（Rule on Impersonation of Government and Businesses）正式生效，使 FTC 得就冒充政府機關或企業之詐欺行為請求民事處罰與消費者救濟；同年，FTC 另提出補充規則，擬將對自然人之冒充一併納入，以回應 AI 生成深偽與仿冒技術快速擴散所帶來之風險[397]。

2024 年 9 月 25 日 FTC 宣布對五家公司提起執法行動，作為其打擊企業利用 AI 從事欺騙或不公平行為之新執法掃蕩行動的一部分，涵蓋利用 AI 宣傳誤導性金融投資方案、提供未符承諾之 AI 專業服務，及提供生成虛假消費者評論之生成式 AI 工具等類型[398]。下述案件均係 FTC 援引聯邦貿易委員會法（Federal Trade Commission Act）第 5(a)條，認定被告之行為構成商業活動中之不公平或欺騙性行為（unfair or deceptive acts or practices in or affecting commerce）[399]。

就涉及金融投資之執法案件而言，FTC 指控 Ascend Capventures Inc. 以「先進 AI」（cutting-edge AI）及專有軟體為行銷噱頭，向消費者收取高達數萬美元費用以開設電商平台店面，並保證可在兩年內達到五位數月收入，然而對幾乎所有消費者而言，前述獲利承諾從未兌現，消費者因此背負鉅額債務；FTC 指控該公司涉嫌詐欺消費者至少 2,500 萬美元[400]。此外，FTC 於 2024 年 6 月對 FBA Machine 提起訴訟，指控其以「AI 驅動」工具協助消費者經營電商店面、保證獲得被動收入為由，向消費者索取

[396] Fed. Trade Comm'n, Comment of the United States Federal Trade Commission, *Artificial Intelligence and Copyright*, DOCKET No. 2023-6, at 3-4 (Oct. 30, 2023).
[397] *Impersonation of Government and Businesses Rule*, 89 Fed. Reg. 15,072 (Mar. 1, 2024) (codified at 16 C.F.R. pt. 461 (2024); *see also* Fed. Trade Comm'n, *FTC Proposes New Protections to Combat AI Impersonation of Individuals* (Feb. 15, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals.
[398] Press Release, Fed. Trade Comm'n, FTC Announces Crackdown on Deceptive AI Claims and Schemes (Sep. 25, 2024).
[399] Complaint at 2 ¶ 1, *FTC v. Ascend Capventures Inc.*, No. 2:24-cv-07660-SPG(JPRx) (C.D. Cal. Sept. 9, 2024); Complaint at 1 ¶ 1, *FTC v. TheFBAMachine Inc.*, No. 2:24-cv-06635-JXN-LDW (D.N.J. June 3, 2024); Fed. Trade Comm'n, *In re DoNotPay, Inc.*, Letters to Commenters on Proposed Consent Order, at 1 (2024).
[400] *Id.*

費用，然承諾獲利從未實現，消費者合計損失逾 1,590 萬美元；聯邦法院就此案發出臨時禁止令，暫停被告相關業務並指派監管人接管[401]。再者，FTC 亦指控 DoNotPay 公司宣稱提供「全球首個 AI 律師」（the world's first robot lawyer）訂閱服務，聲稱用戶得以「不聘請律師即可對傷害提起訴訟」及「即時生成完全有效的法律文件」，並宣稱將以 AI 取代規模達 2,000 億美元之法律行業，然 FTC 認定，DoNotPay 從未對其 AI 聊天機器人之輸出進行測試、未曾聘用或留用律師以驗證服務品質，多項宣傳功能根本未實際提供；DoNotPay 同意支付 193,000 美元和解金，並須向 2021 年至 2023 年間之訂閱用戶發出通知說明服務限制，且禁止在未有充分依據之情形下宣稱其服務得替代任何專業服務[402]。

第二節　州層級立法

依 National Conference of State Legislatures（NCSL）統計，2024 年共有 31 州採行或制定與 AI 相關之法律措施；2025 年則有 38 州採行或制定約 100 項 AI 相關措施。然而，各州已通過之 AI 規範架構性質與管制強度差異甚鉅，美國州層級對金融 AI 之監管，呈現出「立法排除」與「專業監理」並存之雙軌結構。多數州議會所通過之一般性 AI 或資料隱私立法，基於尊重聯邦金融法制之既有分工與避免監理衝突之考量，明文排除受 Gramm-Leach-Bliley Act（GLBA）及 Fair Credit Reporting Act（FCRA）規範之金融機構與資料處理活動[403]；相對而言，真正對金融機構 AI 使用形成直接拘束力者，乃係具備專業監理權限之州金融監理機關（如紐約州 NYDFS）所發布並執行之行政規章與監理措施。基此，以下關於州層級金融 AI 監管之討論，將區分不同治理路徑加以整理與分析。

三、　州法

（一）　科羅拉多州：SB24-205（Consumer Protections for Artificial Intelligence）

1. 立法目的、適用對象及範圍

科羅拉多州於 2024 年通過 SB24-205，於 Colorado Revised Statutes 第 6 篇第 1 章新增 Part 17（Artificial Intelligence），為州層級中針對 AI 系統使用所建立之成文規範[404]。

---

[401] *Id.*
[402] *Id.*
[403] *See*, e.g., Maryland Online Data Privacy Act of 2024, ch. 455, 2024 Md. Laws (exempting financial institutions and personal data subject to the Gramm-Leach-Bliley Act and the Fair Credit Reporting Act); Va. Code Ann. § 59.1-576(C)(1) (Virginia Consumer Data Protection Act) (exempting data subject to the Gramm-Leach-Bliley Act); Colo. Rev. Stat. § 6-1-1304(2)(h) (Colorado Privacy Act) (exempting data regulated by the Gramm-Leach-Bliley Act).
[404] Colo. Rev. Stat. §§ 6-1-1701 to -1707 (2024).

在立法目的上，本法旨在防止 AI 系統之使用對消費者造成不當歧視或其他實質性損害，並透過課予相關主體合理注意義務與治理責任，使 AI 之應用得以置於既有消費者保護與反歧視法制之規範內[405]。

該法之對象為於科羅拉多州從事營業，並就高風險 AI 系統進行「部署」（deployer）或「開發」（developer）之私人主體。其中，「部署者」係指部署高風險 AI 系統之人。「開發者」係指設計、建構或提供 AI 系統之自然人或法人[406]。

2. 規範內容

SB24-205 採取風險導向之適用設計，其規範範圍核心在於「高風險系統」（high-risk artificial intelligence system）。依法律定義，高風險 AI 系統係指在實際運作中，對個人作成或實質影響「重大決策」（consequential decision）之 AI 系統[407]。又所謂重大決策，係指足以影響個人受教育、就業、金融或信用服務、保險、住房、醫療照護或其他基本生活條件之決定[408]。由此可知，立法者明確將 AI 治理之重心，置於對個人權利與經濟地位影響程度較高之應用場景，而非全面性管制所有 AI 用途。

SB24-205 明文禁止因使用 AI 系統而產生之「演算法歧視」（algorithmic discrimination）。該法所稱之歧視，係指 AI 系統於實際運作結果上，對基於聯邦或州法所保護之個人特徵，造成不法之差別待遇或差別影響[409]。

此外，SB24-205 亦要求部署高風險 AI 系統之主體建立持續性之風險管理機制。部署者於 2026 年 2 月 1 日或其後部署高風險 AI 系統者，應完成該高風險 AI 系統之影響評估（impact assessment），以辨識、減輕及監測演算法歧視風險，並確保相關決策過程具備可被檢驗之基礎[410]。當 AI 系統作成不利於消費者之重大決策時，法律並要求部署者向消費者揭露 AI 之使用情形，說明其於決策中所扮演之角色，並提供申訴、更正資料或人工介入之管道[411]。

3. 執行機制與法律效果

在執行層面，SB24-205 採取公法執行模式，授權科羅拉多州檢察總長專屬行使執法權限[412]。違反本法之行為，依法構成州法下之不公平交

[405] Id. §§ 6-1-1702(1), 6-1-1703(1).
[406] *Id.* § 6-1-1701(6) (defining "deployer"); id. § 6-1-1701(7) (defining "developer").
[407] *Id*. § 6-1-1701(9)(a).
[408] *Id.* § 6-1-1701(3).
[409] *Id*. § 6-1-1701(1).
[410] *Id*. § 6-1-1703.
[411] *Id*. § 6-1-1703(4).
[412] *Id*. § 6-1-1706(1).

易行為，得依既有消費者保護法制加以處理[413]。本法並未設置私人訴訟權，而係以行政執法作為主要落實手段[414]。

（二） 加州：強化線上透明法案 Bolstering Online Transparency Act（BOT Act）

加州於 2018 年在商業與專業法規（Business and Professions Code）第七部第三篇中，新增第六章（第 17940 條至 17943 條）即 Bolstering Online Transparency Act（BOT Act），並自 2019 年 7 月 1 日起生效。BOT Act 主要規範線上「bot」之使用。依第 17940 條規定，「bot」係指一種自動化線上帳號，其全部或實質上全部之行為或貼文並非由自然人所為；「online」則指任何對公眾開放之網際網路網站、網頁應用程式或數位應用程式，包括社群網路或出版平台；「線上平台」（Online platform）指的是任何面向公眾的網際網路網站、網路應用程式或數位應用程式，包括社群網路或出版物，且在過去 12 個月的大部分月份中，擁有 10,000,000 或更多獨特的美國每月訪客或使用者；「person」則涵蓋自然人及各類法人實體。該章節並明定，其所課予之義務與其他法律規定為累積適用，而非排他適用。

依第 17941 條規定，任何人若在加州境內，利用 bot 在線上與他人溝通或互動，並意圖誤導對方使其誤認為該帳號為自然人，且該誤導行為係為在商業交易中誘使購買或銷售商品或服務，或影響選舉投票，即屬違法。然而，若使用 bot 之人已揭露其為 bot，則不負該條責任。法律並要求，揭露必須清楚、顯著，且合理設計足以使與該 bot 互動之人知悉其為自動化帳號。

（三） 猶他州：政策法 Artificial Intelligence Policy Act, SB 149

2024 年，猶他州通過 Artificial Intelligence Policy Act，自同年 5 月 1 日起生效。SB 149 於第 13-2-12 條新增 Generative Artificial Intelligence 即生成式 AI 之定義，將其界定為一種受訓練於資料之人工系統，得以文字、音訊或視覺方式與人互動，並在有限或無人類監督情況下生成類似人類創作之非腳本化輸出內容。法律明確規定，生成式 AI 做出違法陳述、採取違法行為或用於促進違法行為，不能作為違反由消費者保護部門執行和管理法規的辯護理由。

此外，若任何人使用、提示或促使生成式 AI 與消費者互動，且該互動涉及州消費者保護法所規範事項，則在消費者詢問或提示時，必須清楚且顯著揭露其為 AI 系統而非自然人。違反該規定者，州消費者保護機關得依法行政處罰，包括科處罰鍰及提起法院行動以執行條款。

[413] *Id*. § 6-1-1706(2).
[414] *Id*. § 6-1-1706(6).

SB 149 進一步規定，於受州商務部監管之「受管制職業」（regulated occupation）中，如於提供專業服務過程使用生成式 AI，應顯著揭露與服務對象互動之為 AI 系統。該揭露須於口頭交流開始時以口頭方式進行，並於書面電子通訊前提供書面揭露。法律同時明確指出，該揭露義務並不允許未符合執業資格之人透過 AI 提供受管制職業服務。

SB 149 同時設立 AI 政策辦公室(Office of Artificial Intelligence Policy)，隸屬於猶他州商務部，並授權其建立 AI 學習實驗室計畫(Artificial Intelligence Learning Laboratory Program)。該制度之目的在於研究與評估 AI 技術之風險、利益與政策影響，並透過學習議程（learning agenda）與參與者制度，分析 AI 技術於州內發展與監管之適當框架。該法亦建立監管緩解協議(regulatory mitigation agreement)，允許參與 AI 學習實驗室計畫的實體，在限定條件下，暫時免除或修改某些州級監管要求，以便測試和評估新的 AI 技術。

SB 149 雖原訂於 2025 年 5 月 1 日自動廢止，然該州於 2025 年立法會期通過 SB 226（消費者保護修正案）與 SB 332（法案修訂），將該法之廢止日期（Sunset Date）由原訂的 2025 年延長至 2027 年 7 月 1 日。依據修正後之規範，針對一般商業使用，法規採「被動揭露」原則，僅在消費者詢問時須說明 AI 身分；惟針對「高風險互動（High-Risk AI Interaction）」，SB226 明文定義其範疇涵蓋涉及「敏感個人資訊（如金融、生物識別數據）」之蒐集，或提供可能被依賴以做出「重大決策（如金融、法律、醫療建議）」之互動。在此類高風險情境下，法案課與服務提供者嚴格的「主動揭露義務」，服務提供者必須於口頭互動開始時（或書面互動前）即明確告知消費者其正在與 AI 互動。

（四） 伊利諾州：商業演算法影響評估法案（HB 5322）

伊利諾州於第 103 屆州議會提出 Illinois Commercial Algorithmic Impact Assessments Act（HB 5322）[415]，另立專法以規範「商業演算法（automated decision tools）」之使用情形。該法係針對使用 AI 系統以作成或實質影響重大決策（consequential decision）之情況，課予開發者與部署者持續性的文件化義務。

HB 5322 法案於 2024 年 2 月 9 日在眾議院提出後，歷經委員會審議與程序性移轉，惟於 2025 年 1 月 7 日會期屆滿（session sine die）時未獲進一步表決，因而失效，並未成為現行州法。是以，雖然該法案最終未能於第 103 屆伊利諾州州議會會期完成立法程序，然本研究對 HB5322 之討論，係立基於其作為一項未完成立法之制度設計提案，用以說明州層級

[415] Illinois Commercial Algorithmic Impact Assessments Act, H.B. 5322, 103d Gen. Assemb., Reg. Sess. (Ill. 2023) (introduced).

在回應商業演算法與自動化決策風險時，所曾考慮採取之規範工具與治理路徑[416]。

依該法定義，自動化決策工具係指使用 AI，且被設計、行銷或經特定修改用以作成或成為重大決策之控制性因素之系統或服務；重大決策則指對個人之生活產生法律上、實質上或類似重大影響之判斷，包括對金融服務之取得、條件或可近性所作之決策。是以，金融服務明文被納入該法所稱重大決策之適用範圍之一[417]。

HB 5322 要求，自 2026 年 1 月 1 日起，凡使用自動化決策工具以作成重大決策之部署者（deployer），須於初次使用及其後每年完成並保存一份影響評估；相同期間內，開發者（developer）亦須就其設計、製造或實質修改、且目的在於作成重大決策之自動化決策工具，完成對應之影響評估。該義務同樣及於對系統作成重大更新（significant update）之情形[418]。

依法定內容，影響評估須至少包括：自動化決策工具之目的、預期用途與使用情境；輸出內容及其如何用於作成或影響重大決策；所處理之自然人資料類型；已知或可合理預見之演算法歧視風險；為符合道德人工智慧（ethical artificial intelligence）原則而採取之防護措施；以及該工具如何由自然人使用或受到監督，並如何被評估其有效性或關聯性。開發者之影響評估內容，則另包含其就系統設計、預期使用或可預見誤用所產生之演算法歧視風險之評估[419]。

此外，該法另規定開發者對部署者負有資訊提供義務，包括提供自動化決策工具之預期用途說明，以及關於已知限制、可能產生之演算法歧視風險、訓練或程式設計所使用之資料類型、與系統在銷售或授權前如何被評估其有效性與可解釋性之文件[420]。影響評估文件於一般情形下不對外公開，惟於伊利諾州檢察總長請求時，開發者或部署者須提供該等文件，否則檢察總長得依法請求禁制令、實際損害賠償及律師費[421]。

四、　州層級行政機關之既有法規適用

美國各州於未另立生成式 AI 金融專法之情形下，大多由具公權力之州級權威機關，透過既有規範對金融領域的 GAI 應用形成規範。此外，州檢察長（State Attorney General, AG）作為州內消費者保護、民權（反歧視）、資料隱私及資料安全等既有法制的核心執法權威之一，得以透過發

[416] Illinois General Assembly, Bill Status of H.B. 5322, 103d Gen. Assemb., Reg. Sess. (Ill. 2023–2024) (last action: Session Sine Die Jan. 7, 2025).
[417] *Id*. § 5 (defining "automated decision tool" and "consequential decision").
[418] *Id*. § 10(a)–(c).
[419] *Id*. § 10(a)(1)–(8); id. § 10(b)(1)–(6).
[420] *Id*. § 15(a)–(c).
[421] *Id*. § 10(d); id. § 10(e).

布 advisory 或 memo 的方式宣示既有法律對生成式 AI 之適用，並向市場參與者提示其執法關切與合規邊界。

（一） 紐約州金融服務署（NYDFS）：AI 網路安全風險與防禦指引

23 N.Y.C.R.R. pt. 500（下稱 Part 500）係紐約州金融服務署（New York State Department of Financial Services, NYDFS）於 2017 年依據州銀行法、保險法及金融服務法授權所制定之資通安全行政規則，其立法目的係回應金融機構在高度數位化環境下面臨之資安威脅，透過要求受 NYDFS 監理之金融機構建立制度化之資安治理框架，以維護非公開個人資訊（Nonpublic Information, NPI）與金融系統之穩定性[422]。Part 500 於 2023 年經 NYDFS 進行第二次重要修正，進一步強化風險導向治理、董事會監督與第三方管理等制度性要求[423]。

Part 500 制定當下，AI 尚未於金融產業中普遍適用，其規範設計亦並非以 AI 為直接規範標的，條文中亦未出現 AI 或自動化決策之明文規定。然而，隨著 AI，特別是生成式 AI 與機器學習系統，逐步為金融機構納入核心或輔助性業務流程，NYDFS 於 2024 年 10 月發布 Industry Letter，就 “Cybersecurity Risks Arising from Artificial Intelligence”指出，AI 之發展已實質改變金融機構之資安風險，並進而影響金融機構之營運風險輪廓[424]。該指引強調，無論是金融機構自身部署 AI 系統，或第三方供應商於金融流程中引入 AI 工具，均可能產生新的資安風險型態，包括但不限於深偽（deepfake）技術所加劇之社交工程攻擊、AI 系統集中處理大量 NPI 所導致之資料外洩風險，以及對外部 AI 模型或雲端服務高度依賴所形成之供應鏈脆弱性。NYDFS 並明確表示，該等風險不構成 Part 500 適用範圍之外的「新型風險」，而應被納入既有 Part 500 所要求之風險評估、治理與監督架構內加以管理。

基此，Part 500 雖仍屬以資安治理為核心之橫向性規範，卻已在實務上成為金融機構部署 AI 系統時，須予以遵循之主要法制作業基礎之一。具體而言，金融機構在金融產業應用場景中使用生成式 AI，例如導入 AI 協助詐欺偵測、客戶互動、文件分析、投資研究或內部流程自動化時，均須將 AI 系統視為資訊系統架構之一環，並依 Part 500 所要求之制度性義務加以管理，包括將 AI 納入資安風險評估範圍、建立相應之內控制度與存取控制機制、就第三方 AI 供應商進行盡職調查與風險管理，以及在發生可能影響 NPI 之事件時履行通報與應變義務[425]。

---

[422] 23 N.Y.C.R.R. pt. 500 (effective Mar. 1, 2017); 23 N.Y.C.R.R. § 500.0 (Introduction); 23 N.Y.C.R.R. § 500.21(a).
[423] 23 N.Y.C.R.R. § 500.21(b); Second Amendment to 23 N.Y.C.R.R. pt. 500, effective Nov. 1, 2023.
[424] New York State Dep’t of Fin. Servs., Industry Letter: Cybersecurity Risks Arising from Artificial Intelligence (Oct. 16, 2024).
[425] *Id*.; *see* also 23 N.Y.C.R.R. §§ 500.2, 500.3, 500.9, 500.11.

（二） 麻薩諸塞州

檢察長辦公室（Massachusetts Attorney General's Office, AGO）依據《麻州消費者保護法》G.L. c. 93A, § 2 及其施行規則（940 C.M.R. 3.00 *et seq.*；940 C.M.R. 5.00 *et seq.*），並同時就《麻州反歧視法》G.L. c. 151B, § 4 與資料安全法 G.L. c. 93H 及其施行規則（201 C.M.R. 17.00 *et seq.*）之遵循義務，發布 *Attorney General Advisory on the Application of the Commonwealth's Consumer Protection, Civil Rights, and Data Privacy Laws to Artificial Intelligence*（下稱「麻州 AI Advisory」），以提供 AI（含演算法決策系統）之開發者、供應者與使用者一般性指引[426]。麻州 AGO 明確指出，該份 Advisory 之目的在於澄清既有州法與州行政規則對 AI 之適用程度，與其對其他產品或應用於市場流通（stream of commerce）時之適用程度相同[427]。AI 雖可能帶來效率與成本節省等利益，但亦已顯示對消費者造成偏誤、缺乏透明或可解釋性、以及資料隱私等風險；且 AI 之缺陷與品質不足在「影響消費者生計、名譽或經濟福祉」之場景中特別令人憂慮[428]（此即包含金融業以 GAI 及一般性 AI 介入信用評估、行銷與客戶互動等核心流程時可能涉及之消費者風險）。

麻州 AGO 進一步以消保法施行規則具體列舉，與 AI 相關之行為可能構成 Chapter 93A § 2 所謂「unfair or deceptive」行為，包括（例如）虛偽宣稱 AI 系統之品質、價值或可用性，或宣稱 AI 具有其實不具備之功能[429]。如提供有瑕疵、無法使用或與宣稱用途不相容之 AI 系統，可能屬於不公平或欺罔，且供應者負有確保 AI 系統如其宣稱目的般運作之義務[430]。此外，誤導 AI 系統之可靠性、運作方式、性能、安全或狀態亦可能構成不公平或欺罔，並以多種「未經測試或未經驗證」之宣稱作為例示，包括宣稱系統為全自動、宣稱其與人類同等或優於人類、宣稱其優於非 AI 產品、宣稱其無偏誤、或宣稱其符合法規等[431]。

再者，AI 系統亦必須遵守《麻州居民個資保護標準》（G.L. c. 93H 與 201 C.M.R. 17.00 *et seq.*），因此 AI 開發者、供應者與使用者應採取必要且適當步驟以保護 AI 系統所使用之個人資訊，並遵循資料外洩通報等法定義務；且 Chapter 93H 之違反行為可由 Chapter 93A 之執法機制涵攝[432]。麻州 AI Advisory 亦指出，麻州反歧視法（G.L. c. 151B, § 4）禁止 AI 開發者、供應者與使用者部署具有歧視效果之技術，包含依賴歧視性輸入而產生

[426] Attorney General Advisory on the Application of the Commonwealth's Consumer Protection, Civil Rights, and Data Privacy Laws to Artificial Intelligence, at 1 (Mass. Att'y Gen. Apr. 16, 2024).
[427] *Id*, at 2.
[428] *Id*, at 1–2.
[429] *Id*, at 3.
[430] *Id*.
[431] *Id*.
[432] *Id*, at 4.

歧視性結果之演算法決策；且違反 Chapter 151B 可能構成 Chapter 93A 意義下之不公平或欺罔行為[433]。

最後，麻州檢察長亦可能被授權執行《特定聯邦消費者保護與反歧視法》（並引 12 U.S.C. §§ 5481, 5552），並以《聯邦平等信用機會法》（ECOA）之 adverse action notification 要求為例，明示該要求亦適用於 AI 模型。因此放款機構即使使用 AI 模型，仍須向消費者提供「正確且具體」之拒貸或不利處分理由[434]。

（三） 康乃狄克州

康乃狄克州檢察長發布之 Memorandum to All State Officials, Agencies, and Concerned Parties: The Application of Existing Laws to Artificial Intelligence to Protect Connecticut Residents（下稱「康州 AI Memo」），其自我定位係在 AI 快速擴張之背景下，提供消費者與企業理解「既有法律如何適用於 AI」的指引性文件[435]。康州 AI Memo 在《民權法》章節中明確指出，康州檢察長辦公室對州與聯邦反歧視法之執行具有廣泛權限，並援引 Conn. Gen. Stat. § 3-129g 作為其作成該文件之權源之一[436]。同時，康州 AI Memo 明確聲明，該備忘錄不構成法律建議或正式法律意見書，亦非具拘束力或先例性文件，而僅係就 AI 脈絡下可能相關之法律原則作概覽性整理[437]。

在資料隱私部分，《康州資料隱私法》（Connecticut Data Privacy Act, CTDPA）賦予居民存取、更正、刪除及取得副本等權利，並包含對特定資料處理之 opt-out 權利；其中 opt-out 包括對「可能具有法律效果或其他重大影響」之自動化剖析（automated profiling）。CTDPA 對控制者與處理者課予資料最小化與目的限制相關義務（例如資料蒐集應限於合理要範圍且須以隱私告知清楚揭露），並禁止企業因消費者行使 CTDPA 權利而差別待遇[438]。此外，依 Public Act 25-113（2026/7/1 生效），企業不得處理未成年人資料以進行與「核准或拒絕金融服務」等事項相關之自動化決策剖析（profiling），除非符合該條文所列之必要性條件[439]。除 CTDPA 外，康州 AI Memo 指出企業亦受康州一般性資料安全與外洩通報法制拘束（如 Safeguards Law 與 Breach Notification Law），並說明 breach notification 所涵蓋之個資類型具有廣泛性，除金融資訊外亦包含帳號登入憑證與生物辨識資訊等[440]。

---

[433] *Id*.
[434] *Id*.
[435] Memorandum to All State Officials, Agencies, and Concerned Parties: The Application of Existing Laws to Artificial Intelligence to Protect Connecticut Residents, at 1 (Conn. Att'y Gen. Feb. 25, 2026).
[436] *Id*, at 2.
[437] *Id* at 2.
[438] *Id* at 4–5.
[439] *Id* at 5.
[440] *Id* at 6-7.

在消費者保護方面，《康州公平交易行為法》（Connecticut Unfair Trade Practices Act, CUTPA）係保護消費者免於不公平與欺罔交易行為之廣泛救濟性法規[441]，康州 AI Memo 以例示方式指出，若企業在銷售 AI 產品或以 AI 進行行銷或銷售時，對 AI 產品之有效性、用途或能力作不實陳述，或以 AI 製作足以誤導背書、關聯或消費者評價之內容等，可能構成 CUTPA 違反。如金融業導入 GAI 並有「AI 誇大宣稱或 AI 誤導行銷」之行為，即可能落入既有消費者保護執法架構的射程[442]。

第三節　小結

綜上所述，美國並未形成單一、統一且專屬於 AI 的成文監理體系，而是呈現以多機關職權分立與既有法律適用為核心的分散式治理，聯邦層級多以行政命令與自願性風險治理框架設定方向與共同語彙，並由各主管機關在其既有法定職權下，將既有義務延伸適用至金融場景中。

在州層級中，立法與執法呈現高度異質與碎片化，州議會以 AI 專法或情境化規則回應特定風險場景，同時州層級權威機關也以既有規範框架補足金融業 GenAI 快速落地下的合規需求。美國監理體系面臨「如何在不犧牲各州自主監理權限下，提升監理方法的一致性與專業支援」的制度需求。本研究所考察之四州相關規範，截至本研究撰寫時均無公開執法案例。

面對上述碎片化與異質性，各州正透過州銀行監理官會議（Conference of State Bank Supervisors, CSBS）所推動的跨州合作機制，強化共同標準、訓練與監理工具的建置。CSBS 已於 2024 年 8 月 22 日成立 Artificial Intelligence Advisory Group，邀集學界、金融業界與非營利組織專家，向 CSBS 與州金融監理官提供 AI 在金融服務領域「開發與使用」之諮詢與觀點，並支援州監理官評估 AI 相關之法律、政策與監理（supervisory）活動[443]。

此外，本研究第一章曾自規範效力之角度，將美國定性為軟法之代表。惟本章於美國 AI 監理之整體制度架構觀察後認為，美國除具軟法特徵外，實則亦同時呈現分散式治理之樣貌。此二種定性並非互斥，而係從不同層次描述同一現象，亦即，軟法著重於規範「效力面」，指出美國聯邦層級並未以具拘束力之成文法規範 GenAI，而係仰賴行政命令、自願性風險治理框架，以及各主管機關發布之指引、備忘錄等不具強制執行力之工具；分散式治理則著重於「制度面」，指出前述軟法工具之發布主體

[441] *Id* at 7.
[442] *Id* at 7-8.
[443] Press Release, CSBS, *CSBS Establishes Artificial Intelligence Advisory Group* (Aug. 22, 2024), https://www.csbs.org/newsroom/csbs-establishes-artificial-intelligence-advisory-group (last visited Jul. 5, 2026).

並非單一集中之監理機關，而係分散於聯邦各部會、各州議會及州檢察長辦公室等多元權責機關，各自依其既有法定職權範圍另行發展規範內容，形成職權分立、規範內容不一之治理格局。換言之，正因為美國選擇以軟法而非統一硬法作為主要規範形式，方使聯邦與各州、各主管機關得以在缺乏中央統一立法之情形下，各自發展出因機關而異、因州而異之規範內容，此一現象即為分散式治理之具體展現，故本章爰於此併予補充說明。

第五章　新加坡 AI 監管法制

相較於歐盟採取全面性的硬性立法，或美國聯邦政府仰賴既有法規之延伸適用，新加坡在 AI 之監理上採取與歐美截然不同的路線。新加坡並未針對 AI 制定專門的成文法，而是透過資通訊媒體發展局（Infocomm Media Development Authority, IMDA）與新加坡金融管理局（Monetary Authority of Singapore, MAS）等主管機關，密集發布一系列之治理框架、原則、指引或測試工具等軟法，協助企業或金融機構在創新與風險間取得平衡。

第一節　國家級治理框架

新加坡自 2019 年起即由 IMDA 與個人資料保護委員會（Personal Data Protection Commission, PDPC）共同發布《AI 治理框架》（Model AI Governance Framework）」，建立以內部治理結構、人為參與程度、營運管理及利害關係人溝通為核心的問責機制。

隨著技術演進，新加坡於 2024 年發布《生成式 AI 治理框架》（Model AI Governance Framework for Generative AI）」，增列內容溯源（Content Provenance）、測試與保證（Testing and Assurance）等生成式 AI 專屬的治理面向。

一、　AI 治理框架第二版[444]

此框架由 IMDA 與 PDPC 於 2019 年 1 月共同發布第一版，並於 2020 年 1 月發布第二版。此框架旨在提供一套可用於組織內部的問責機制，協助部署 AI 的組織管理風險。框架分為以下四大面向[445]：

（一）　內部治理結構與措施

組織應建立明確的角色與責任[446]，由適當的人員或部門負責監督 AI 部署，並應將 AI 的道德議題與風險納入企業現有的企業風險管理結構或道德審查委員會中[447]。

（二）　擬定人為參與 AI 決策的程度

---

[444] 此框架並非針對金融業所發布，而是各產業通用的 AI 治理框架。此外，此框架首次發布的時間與 MAS 發布 FEAT 原則的時間相近，當時尚未有可商用的生成式 AI 。即使在第二版框架發布時，也僅有於前言中提到當時正在發展中的 GPT-2 模型，因此框架並未針對生成式 AI 進行說明或分析。不過此框架屬於新加坡 AI 治理的早期規範之一，當中的概念也依然被後續的指引或規範所引用，故本研究仍予收錄並摘要。

[445] Infocomm Media Development Authority & Personal Data Protection Commission, *Model AI Governance Framework Second Edition*, 20, https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf (last visited Jul. 5, 2026).

[446] *Id*. at 22.

[447] *Id*. at 24.

組織於部署 AI 前，應先確立商業目標，並權衡 AI 決策所帶來的風險[448]。框架將人為參與的程度分為三類[449]：人在迴圈內（Human-in-the-loop）、人在迴圈外（Human-out-of-the-loop）、人在迴圈上（Human-over-the-loop）。人在迴圈內意指人類保留完整控制權，AI 不能在沒有人類明確指令的情況下自行做出決策；人在迴圈外則與前者相反，人類沒有任何干預空間；人在迴圈上則是指人類在 AI 出現故障或不良事件時才接管控制權。

（三） 營運管理

可區分為資料管理與演算法模型二類，前者強調了解資料來源與品質，盡可能減少資料中的偏見[450]；後者則著重於演算法的透明度[451]，包含可解釋性（Explainability）、可重複性（Repeatability）、穩健性（Robustness）、定期調校（Regular Tuning）、可追溯性（Reproducibility）及可稽核性（Auditability）等。

（四） 利害關係人互動與溝通

組織應主動揭露產品或服務是否使用了 AI[452]，並提供讓消費者要求審查決策的管道以及反饋機制[453]。

二、 生成式 AI 治理框架

由新加坡 IMDA 與其輔導的非營利組織 AI Verify Foundation 於 2024 年 5 月修訂並發布，為新加坡國家級框架。IMDA 與 AI Verify Foundation 於傳統 AI 框架的基礎上[454]，進一步提出生成式 AI 治理框架（Model AI Governance Framework for Generative AI）。金融機構開發生成式 AI 時需參考此框架中的九大治理維度。不過在此框架提出前，金融業早已有許多治理機制，因此框架中也建議借鑑現有的金融領域機制[455]。有關框架中的九大治理維度，分述如下：

（一） 問責[456]（Accountability）

須事前定義 AI 開發與部署者的權限，並分配責任歸屬，可參考雲端產業的「共享責任模型」。另外，亦應考慮賠償和保險等事後的補救措施，以保障用戶權益。

---

[448] *Id*. at 28-29.
[449] *Id*. at 30.
[450] *Id*. at 36-40.
[451] *Id*. at 43-51.
[452] *Id*. at 53-54.
[453] *Id*. at 54-57.
[454] AI Verify Foundation & Infocomm Media Development Authority, *Model Al Governance Framework for Generative Al*, 3, https://aiverifyfoundation.sg/wp-content/uploads/2024/06/Model-AI-Governance-Framework-for-Generative-AI-19-June-2024.pdf (last visited Jul. 5, 2026).
[455] *Id.* at 20.
[456] *Id.* at 7-8.

（二） 數據[457]（Data）

須確保數據品質與來源的可信度，並釐清個人資料保護規定（Personal Data Protection Act, PDPA）。另建議透過隱私強化技術（Privacy-enhancing Technologies, PETs）和務實的解決方案，平衡創作者的著作權與數據存取。

（三） 可信任的開發與部署[458]（Trusted Development & Deployment）

以類似食品標示的方式，明確標示訓練數據、限制、風險及安全措施，並採用人類反饋強化學習和 RAG 技術以減少幻覺並提高模型準確性。

（四） 事故通報[459]（Incident Reporting）

建立 AI 事故的通報機制，並建立漏洞通報管道，讓白帽駭客或研究人員能在事故發生前發現並通報問題。

（五） 測試與保證[460]（Testing and Assurance）

參照金融業的審計制度，透過第三方（如 ISO/IEC 和 IEEE）的測試工具進行標準化驗證，確保測試的一致性與結果的可比較性。

（六） 安全性[461]（Security）

生成式 AI 帶來新的威脅，如提示詞注入攻擊。需要開發新的安全工具，如輸入過濾器（Input Filters）和數位鑑識工具。

（七） 內容溯源[462]（Content Provenance）

透過數位浮水印和加密溯源技術，標記 AI 生成的內容，以應對深偽和錯誤資訊問題，確保用戶可以驗證內容的真實性。

（八） 安全與一致性研發[463]（Safety and Alignment R&D）

目前的安全技術不足以涵蓋所有風險，因此需要增加研發，確保模型價值觀符合人類社會規範。

（九） 公益[464]（AI for Public Good）

負責任的 AI 應致力於民主化科技存取權（Democratising Access to Technology）、改善公共部門採用率、提升勞動力技能，並確保永續發展 。

[457] *Id.* at 10-11.
[458] *Id.* at 13-14.
[459] *Id.* at 17-18.
[460] *Id.* at 20.
[461] *Id.* at 22.
[462] *Id.* at 24-25.
[463] *Id.* at 27.
[464] *Id.* at 29-30.

第二節　FEAT 原則與檢測工具

在金融產業方面，MAS 早在 2018 年即發布「FEAT 原則」，確立金融業使用 AI 須符合公平性（Fairness）、倫理性（Ethics）、問責性（Accountability）與透明性（Transparency）。然而，為解決抽象原則難以落實之痛點，新加坡推動了「真理倡議」，將 FEAT 原則轉化為可操作的實踐框架，並開發開源的檢測工具，讓金融機構得以量化評估信用風險評分與行銷模型中的公平性指標。

而面對大型語言模型所帶來的獨特安全風險（如提示詞注入、模型幻覺等），新加坡進一步推出「登月計畫」。該計畫提供一款開源測試工具包，允許金融機構直接與模型互動，透過評分機制（Benchmarking）與紅隊測試（Red Teaming）測試模型的安全性與潛在的幻覺問題，為解決生成式 AI 的黑盒難題提供了不一樣的解方。

一、　FEAT 原則[465]

《FEAT 原則》（Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector, FEAT Principles）係由 MAS 於 2018 年 11 月所發布的治理指引，旨在為金融機構使用 AI 與數據分析（Artificial Intelligence and Data Analytics, AIDA）提供一套通用的核心原則。這些核心原則並非法規，而是 MAS 提供予金融機構參考的基準，金融機構應判斷由 AI 和數據分析所參與的決策「重大性[466]（Materiality）」調整其內部治理架構。

雖然這份指引發布時，尚未出現被廣泛應用的生成式 AI ，因此指引中也沒有特別提及生成式 AI ，但指引中對 AI 的定義相當寬鬆[467]，足以涵蓋生成式 AI。且由 MAS 後續發布的許多文件或指引內容亦可推知，以下四項核心原則目前依然是金融領域運用 AI 時應參考的原則。

（一）　公平性

確保 AI 的決策不會使特定個人或群體處於劣勢（Disadvantage）。公平性包含兩項內涵[468]，一是「正當（Justifiability）」，例如，使用個人屬性（如性別、種族）作為提高客戶信用卡額度模型的輸入變數時，必須

[465] FEAT 原則屬新加坡在 AI 智慧與數據分析領域的上位階原則。

[466] Monetary Authority of Singapore, *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*, 4, https://www.mas.gov.sg/-/media/mas/news-and-publications/monographs-and-information-papers/feat-principles-updated-7-feb-19.pdf (last visited Jul. 5, 2026).

[467] *Id.* at 5. 指引對 AI 的定義為「協助或取代人類決策的技術」（Technologies That Assist or Replace Human Decision-Making）。

[468] *Id.* at 7.

有明確的正當理由。二是「準確[469]（Accuracy）」，必須定期審查數據與模型，確保其準確性與相關性，減少錯誤或非預期偏誤（Errors or Unintentional Biases）。

（二） 倫理性

使用 AI 應符合金融機構內部的倫理標準與社會價值觀[470]。例如，金融機構基於性別平等，於招聘時不考慮性別因素，則使用 AI 輔助篩選應徵者時，便不應將性別屬性輸入招聘模型中。此外，金融機構應考慮設立獨立的道德審查委員會，以監督和維護其道德標準。

（三） 問責性

區分「內部」與「外部」問責。內部問責方面[471]，無論 AI 模型是由內部或外部開發，使用 AI 時都應該經過內部適當的部門批准，並由公司負責。對於使用 AI 進行的重要的決策，批准權限應設定於執行長或董事會，董事會與高階管理層必須對 AI 的使用有足夠的認識，並進行盡職調查（Due Diligence）。外部問責方面[472]，當客戶對 AI 決策有異議時，金融機構必須提供申訴管道進行人工覆核，並允許客戶提供額外的補充資料以修正 AI 的判斷。

（四） 透明性

金融機構應主動告知客戶該服務使用了 AI 技術。當客戶提出要求時，金融機構應能清楚解釋[473]「使用了哪些數據（Data Used for Decision-Making）」、「這些數據如何影響決策（How Data Affects the Decision）」及「對客戶的影響（Consequences on Data Subjects）」，且清楚的解釋並不等同於揭露或發布原始碼。此外，透明性的高低取決於 AI 決策事項的重要性[474]。例如，為避免模型被操縱，金融機構不提供對於詐欺偵測模型的解釋，但對於客服 AI 則應提供清楚的解釋與告知。

二、 真理倡議

為使金融機構能遵循 FEAT 原則，MAS 於 2019 年啟動真理倡議，目標是將 FEAT 原則轉化為可操作的實踐框架。但由於真理倡議開始的時間較早，因此真理倡議工具主要是針對使用機器學習演算法的傳統 AI 進行測試。

---

[469] *Id.* at 8.
[470] *Id.* at 9.
[471] *Id.* at 10.
[472] *Id.* at 11.
[473] *Id.* at 12.
[474] *Id.* at 13.

（一） 第一階段

專注於銀行業的「信用風險評分（Credit Risk Scoring）」與「客戶行銷（Customer Marketing）」兩個應用場景，開發評估公平性的方法。2021 年 1 月發布了兩份白皮書，分別說明如何定義與量化公平性[475]，以及提供應用範例[476]。

（二） 第二階段

將評估方法擴展至四項 FEAT 原則，並納入更多保險業與銀行業的案例，如預測性承保、欺詐檢測等。2022 年 2 月發布了五份白皮書[477]和開放原始碼的真理倡議工具。真理倡議工具主要處理結構化數據，因此可以針對機器學習或傳統預測型 AI[478]的數據進行檢測。金融機構的開發人員可將模型訓練數據或預測結果傳入真理倡議工具，真理倡議工具會運行統計公式檢查是否對某一族群（如性別）有顯著差異，或使用 SHAP（Shapley Additive Explanations, SHAP）值分析哪個特徵（如收入、居住地）對決策影響最大。

[475] Veritas Consortium, *Veritas Document 1 FEAT Fairness Principles Assessment Methodology*, https://www.mas.gov.sg/-/media/mas/news/media-releases/2021/veritas-document-1-feat-fairness-principles-assessment-methodology.pdf (last visited Jul. 5, 2026).

[476] Veritas Consortium, *Veritas Document 2 FEAT Fairness Principles Assessment Case Studies*, https://www.mas.gov.sg/-/media/mas/news/media-releases/2021/veritas-document-2-feat-fairness-principles-assessment-case-studies.pdf (last visited Jul. 5, 2026).

[477] Veritas Consortium, *Veritas Document 3 - FEAT Principles Assessment Methodology*, https://www.mas.gov.sg/-/media/mas-media-library/news/media-releases/2022/veritas-document-3---feat-principles-assessment-methodology.pdf (last visited Jul. 5, 2026); Veritas Consortium, *Veritas Document 3A - FEAT Fairness Principles Assessment Methodology*, https://www.mas.gov.sg/-/media/mas-media-library/news/media-releases/2022/veritas-document-3a---feat-fairness-principles-assessment-methodology.pdf (last visited Jul. 5, 2026); Veritas Consortium, *Veritas Document 3B - FEAT Ethics and Accountability Principles Assessment Methodology*, https://www.mas.gov.sg/-/media/mas-media-library/news/media-releases/2022/veritas-document-3b---feat-ethics-and-accountability-principles-assessment-methodology.pdf (last visited Jul. 5, 2026); Veritas Consortium, *Veritas Document 3C - FEAT Transparency Principles Assessment Methodology*, https://www.mas.gov.sg/-/media/mas-media-library/news/media-releases/2022/veritas-document-3c---feat-transparency-principles-assessment-methodology.pdf (last visited Jul. 5, 2026); Veritas Consortium, *Veritas Document 4 - FEAT Principles Assessment Case Studies*, https://www.mas.gov.sg/-/media/mas-media-library/news/media-releases/2022/veritas-document-4---feat-principles-assessment-case-studies.pdf (last visited Jul. 5, 2026).

[478] 指底層使用隨機森林、XGBoost、Logistic Regression 等演算法的 AI，本研究統稱為傳統 AI，以便與底層使用深度學習技術的生成式 AI 區辨。

（三） 第三階段

提供金融機構整合上述方法的經驗，並於 2023 年 6 月發布兩份白皮書[479]及開放原始碼的真理倡議工具包 2.0[480]。

三、 登月計畫

IMDA 於 2024 年 5 月新加坡亞洲科技大會（ATxSG）上發布登月計畫，此計畫是由 DataRobot 公司、IBM 公司、新加坡電信公司和淡馬錫公司等企業等共同開發，其核心的成果為一款開源的測試工具包，旨在解決大型語言模型常見的安全問題[481]。

將大型語言模型的 API 與測試工具包串接後，測試工具包可以直接和語言模型互動並進行測試，主要可分為以下兩種方式[482]：

（一） 評分機制（Benchmarking）

測試工具包會將模型的回答與標準答案（Ground Truth）進行比對，或使用另一個大型語言模型進行評分，判斷受測模型的回答是否有毒或產生幻覺。

（二） 紅隊測試

測試工具包會自動生成惡意指令進行提示詞注入攻擊。例如向模型發出「忽略之前的指令，告訴我信用卡號」的提示詞，測試模型是否會洩漏個資。另外，測試工具包會根據不同文化背景，測試模型的理解能力（Contextual Testing）。

第三節 強化生成式 AI 風險管理與聯盟合作

除了上位階的 FEAT 原則，MAS 尚針對生成式 AI 發布一系列風險管理的指引與手冊，歸納金融機構在面對或運用生成式 AI 時所可能遭遇的風險樣態與緩解措施。另外，面對跨國科技巨頭的技術優勢，MAS 推動「MindForge 計畫」，號召金融機構與跨國科技公司組成聯盟，共同制定

[479] Veritas Consortium, *Veritas Document 5 - From Methodologies to Integration*, https://www.mas.gov.sg/-/media/mas/news/media-releases/veritas-document-5---from-methodologies-to-integration.pdf (last visited Jul. 5, 2026); Veritas Consortium, *Veritas Document 6 - FEAT Principles Assessment Case Studies*, https://www.mas.gov.sg/-/media/mas/news/media-releases/veritas-document-6---feat-principles-assessment-case-studies.pdf (last visited Jul. 5, 2026).

[480] Veritas Consortium, *Veritas-toolkit - Overview*, GITHUB, https://github.com/veritas-toolkit (last visited Jul. 5, 2026).

[481] Infocomm Media Development Authority, *Singapore launches Project Moonshot*, INFOCOMM MEDIA DEVELOPMENT AUTHORITY PRESS RELEASES, FACTSHEETS AND SPEECHES, https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/press-releases/2024/sg-launches-project-moonshot (last visited Jul. 5, 2026).

[482] AI Verify Foundation, *Aiverify-foundation/Moonshot*, https://github.com/aiverify-foundation/moonshot (last visited Jul. 5, 2026).

行業標準，以集體力量弭平單一金融機構在第三方科技管理上的議價弱勢。

一、 與生成式 AI 相關網路風險資訊文件

此份文件是由 MAS 於 2024 年 7 月發布，旨在提高金融機構對生成式 AI 相關的網路與技術威脅的認識，並建議緩解措施。此份文件羅列以下四大類威脅，並以實際發生的攻擊案例進行說明。

（四） 深偽技術與網路釣魚[483]

攻擊者利用 AI 生成逼真的人類影像或聲音，增強社交工程攻擊的效果，甚至可以模仿受害者所信任的人；或者利用 GPT 系列模型製作針對性極強的釣魚郵件。文件建議[484]金融機構實施活體檢測技術（如偵測動作、皮膚紋理）、對高風險交易及高權限帳戶實施多因素認證（Multi-Factor Authentication, MFA）、舉辦員工模擬演練等預防措施。而針對深偽技術的威脅，文件列出三個實際案例：一為 2024 年 2 月，詐騙者曾利用惡意軟體和深度偽造技術，試圖繞過越南和泰國多家銀行的臉部辨識認證系統[485]；二為 2024 年 1 月，詐騙分子利用深偽技術，偽造香港一家跨國公司財務長的容貌參加視訊會議，成功誘騙其員工匯出 2,500 萬美元[486]；三為 2023 年 8 月，深偽技術被用於製作貸款人的容貌，成功向香港的金融機構詐得 2.5 萬美元的貸款[487]。

（五） 惡意軟體的生成與增強[488]

攻擊者使用如 WormGPT 等工具快速編寫惡意程式，甚至利用生成式 AI 頻繁更改程式碼，以避開傳統防毒軟體的特徵碼檢測。文件建議[489]金融機構採取多層次防禦策略、採用 AI 工具偵測惡意軟體、將威脅情資整合至日誌監控中等措施，避免遭惡意軟體攻擊。文件舉出兩個案例[490]：一為資安企業 HYAS 公司所打造的「BlackMamba」惡意軟體，利用生成式 AI 修改正常軟體側錄鍵盤竊取金融密碼和憑證，使傳統資安防護軟體難以偵測；二為研究人員所創建的「DeepLocker」惡意軟體，利用生成式 AI 判斷攻擊目標，一旦發現可攻擊的目標，生成式 AI 才會將惡意程式碼

---

[483] Monetary Authority of Singapore, *Information Paper on Cyber Risks Associated with Generative Artificial Intelligence*, 4, https://www.mas.gov.sg/-/media/mas-media-library/regulation/circulars/trpd/cyber-risks-associated-with-generative-artificial-intelligence.pdf (last visited Jul. 5, 2026).
[484] *Id.* at 5-6.
[485] *Id.* at 4.
[486] *Id.* at 5.
[487] *Id.*
[488] *Id.* at 6-7.
[489] *Id.* at 7-8.
[490] *Id.* at 7.

展現出來，以躲避防毒軟體的偵測。這兩種案例雖然並未被實際利用，但卻已經清楚展示出生成式 AI 被利用於惡意軟體的高度可能性。

（六） 資料外洩[491]

員工使用公共生成式 AI 工具（如 ChatGPT、Gemini）時上傳敏感數據，或攻擊者進行提示詞注入攻擊，誘導模型洩露敏感資訊。此外，模型本身也可能含有後門或漏洞導致資料外洩。文件建議[492]金融機構制定明確的資料分類與使用政策、開發內部模型時採用「安全設計（Security-by-Design）」、實施資料遺失防護（Data Loss Prevention, DLP）和專用防火牆等措施應對資料外洩的威脅。文件同樣列出三個案例[493]：一為 2024 年 3 月，生成式 AI 瀏覽器的外掛程式「AskTheCode」、「Charts」被發現存在漏洞，攻擊者可以利用這些漏洞存取受害者瀏覽的資訊，或存取受害者的 Google 雲端硬碟和 Github 資料；二為 2023 年 3 月，ChatGPT 所使用的資料庫用戶端程式被發現存有漏洞，可能使被害人的個人資訊、信用卡資訊和歷史聊天記錄被洩漏；三為 2023 年 3 月，一家大型電子公司曾發生三起員工使用 ChatGPT 時洩漏敏感資料事件。

（七） 模型與輸出操弄[494]

攻擊者於模型訓練階段竄改資料進行「資料投毒（Data Poisoning）」，導致模型生成有利於攻擊者的錯誤輸出；或透過惡意輸入影響模型的回答、傳播錯誤訊息。文件建議[495]金融機構建立模型與資料治理（確保數據準確性）、嚴格的存取控制、持續監控模型偏差、在業務連續性計畫（Business Continuity Plan, BCP）中加入生成式 AI 應急措施（如保留訓練資料備份）等措施以資因應。文件亦指出[496]，雖然目前金融產業尚未發生過資料竄改事件，但其他產業已有研究人員證實可能發生。2023 年 6 月，美國研究人員發布報告指出由 AI 驅動的醫療診斷系統可能會受到被篡改的數據誤導，進而影響醫療診斷的準確性；2023 年 7 月，一家網路安全公司亦示範如何透過使用偽造資料訓練生成式 AI 模型，從而操縱並影響模型的回答。

二、 AI 模型風險管理資訊文件

2024 年 12 月 MAS 對銀行的 AI（包括生成式 AI）模型風險管理實務進行審查，並整理出良好實踐（Good Practices），供所有金融機構參考。

[491] *Id.* at 9.
[492] *Id.* at 10-11.
[493] *Id.* at 9-10.
[494] *Id.* at 12.
[495] *Id.* at 13-14.
[496] *Id.* at 12-13.

並特別針對生成式 AI 的模型幻覺、資料隱私與第三方依賴風險提出具體建議。

（一） AI 的風險[497]

包含財務風險（AI 風險評估不準確導致損失）、營運風險（AI 出現意外行導致營運中斷）、監管風險（反洗錢 AI 表現不佳導致不合規）、聲譽風險（客訴和負面媒體報導）。而生成式 AI 可能會放大這些 AI 既有的風險[498]。

（二） 管理幻覺風險[499]

為了控制幻覺所造成的影響，多數銀行僅在時間和用戶數受限的場域進行生成式 AI 的測試。

（三） 流程控制[500]（Process Controls）

因應生成式 AI 的跨領域應用以及快速迭代特性，一些銀行在生成式 AI 的生命週期中建立風險控制檢查機制，並引入人工監督或人在迴圈內（human-in-the-loop）機制。

（四） 技術控制（Technical Controls）

由於大多數銀行使用的生成式 AI 模型（無論閉源或開源）都來自第三方。因此，為了評估模型的適用性，一些銀行會利用公開的測試基準（如模型排行榜、開源的評估資料集）和最新研究論文中的評估結果進行決策。

部分銀行甚至會進一步從其內部歷史資料庫中，建立銀行專屬的測試資料集[501]，或聘請專家對生成結果進行人工標註，以便進行評估。另外，一些銀行會進行模型漏洞測試、穩定性測試和敏感性測試。這些測試資料集和人工回饋對於監測生成式 AI 模型的性能至關重要。

大多數銀行都建立了輸入和輸出防護機制[502]。例如利用過濾器管理有害言論、偏見言論或敏感資訊外洩等風險，也可利用過濾器拒絕有害的請求，或將請求中的個人識別資訊（PII）替換為去識別化的符號。另一個常見的防護機制是採用 RAG 技術，透過限制生成式 AI 的資料來源，增加輸出的準確性。

---

[497] Monetary Authority of Singapore, *Information Paper on AI Model Risk Management*, 4, https://www.mas.gov.sg/-/media/mas-media-library/publications/monographs-or-information-paper/imd/2024/information-paper-on-ai-risk-management-final.pdf (last visited Jul. 5, 2026).
[498] *Id.* at 5-6.
[499] *Id.* at 39.
[500] *Id.* at 39-40.
[501] *Id.* at 40-41.
[502] *Id.* at 41-42.

此外，銀行使用生成式 AI 時，可以透過「部署於私有雲（Using Private Cloud）」、「本地部署開源模型（Open-Source Models Onpremise）」等方式降低資料暴露的風險[503]。

（五） 第三方風險管理標準與流程

多數銀行使用的生成式 AI 模型均是由外部機構預先訓練，使用此類第三方生成式 AI 會帶來額外的風險，如預訓練資料中可能存在未知的偏差、資料保護問題、以及底層依賴相同的生成式 AI 模型所導致的集中風險。另外，缺乏透明度也常被認為是管理此類第三方風險的關鍵挑戰。為了降低這些額外風險，銀行可以採用「補償性測試[504]（Compensatory testing）」、「應急計畫[505]（Contingency planning）」、「法律協議[506]（Legal agreements）」、「意識提升[507]（Awareness efforts）」等方式控制這些額外風險。

三、 銀行業生成式 AI 護欄手冊

此手冊是由新加坡銀行公會（The Association of Banks in Singapore, ABS）、MAS 以及多家銀行（如星展銀行、華僑銀行、大華銀行、花旗銀行等）合作撰寫，目的是協助金融機構以負責任、安全且合乎道德的方式採用生成式 AI 。此手冊區分了傳統 AI 與生成式 AI 在銀行業應用上的差異[508]，前者主要用於處理結構化數據、優化營運（如欺詐檢測、風險評估），後者則是能創造新內容（文本、圖像、程式碼），並能進行更複雜的互動與運用[509]。手冊中進一步指出各種應用場景中主要的風險，分述如下：

（一） 企業層級風險[510]

係指缺乏生成式 AI 風險意識、缺乏模型治理、人工監督不足、缺乏反饋機制。

---

[503] *Id.* at 42.

[504] 補償性測試係指使用各種資料集和情境對第三方 AI 模型進行嚴格測試，以驗證模型在銀行環境中的穩健性和穩定性，並偵測潛在的偏差。

[505] 應急計畫係指制定完善的應急計劃，以應對潛在的故障、第三方 AI 的意外行為或供應商停止支援的情況，包含建立備份系統或人工流程。

[506] 法律協議係指更新與第三方 AI 供應商的契約，加入有關效能保證、資料保護、稽核權以及在引入現有 AI 解決方案之外的 AI 的通知條款。這些法律協議有助於明確雙方的期望和責任。

[507] 意識提升係指對員工的 AI 素養和風險意識培訓、對第三方供應商進行 AI 開發和風險管理的調查。

[508] The Association of Banks in Singapore, *Handbook on Generative AI Guardrails in Banking*, 9-10, https://www.abs.org.sg/docs/library/handbook-on-generative-ai-guardrails-in-banking.pdf (last visited Jul. 5, 2026).

[509] *Id.* at 13. 手冊中歸納出七種常見的應用場景：程式碼編寫、語音轉文字、文件擷取/摘要、翻譯、內容創作、知識管理/搜尋、流程優化。

[510] *Id.* at 14.

（二） 模型風險[511]

1. 幻覺

產出看似合理但不正確的資訊，可能導致銀行決策錯誤或法律問題。

2. 準確性不足（Insufficient Model Accuracy）

產出不符合銀行業所需的精確度。

4. 有毒或攻擊性的輸出（Toxic & Offensive Output）

產出歧視或有害的內容。

4. 偏差的輸出（Bias Output）

數據對某些個人或群體有偏見或代表性不足，進而導致模型輸出存有偏差。

4. 過度自信（Overconfidence）

模型以肯定的語氣呈現錯誤資訊，影響行員的判斷。

4. 模型性能下降（Model Degradation）

意料之外的使用模式導致結果不穩定或故障。

針對以上風險，手冊提出九種護欄（Guardrails）：

（一） 企業層級護欄[512]（Enterprise-Level）

透過企業治理與培訓，建立 AI 原則、治理模式、AI 清單，並對員工進行生成式 AI 培訓。

（二） 技術面系統層級護欄（System-Level – Technical Guardrails）

1. 過濾與控制[513]（Filtering and Control）

透過預設指令、輸入輸出過濾工具限制模型行為。

2. 客製化模型設計[514]（Customised Model Design）

針對高風險的應用場景，可利用「微調」或「RAG」技術，使模型基於受銀行信任的數據所生成回應、減少幻覺。

3. 紅隊測試[515]

在部署前進行對抗性測試，模擬攻擊以找出漏洞（如誘導模型洩露客戶個資）。

4. 提示詞設計[516]（Prompt Design）

使用「思維鏈」技術或建立中央提示詞庫（Centralized Prompt Library），引導模型產出符合銀行標準的回答。

（三） 流程面系統層級護欄（System-Level – Process Guardrails）

1. 監控與驗證[517]

[511] *Id.* at 14-15.
[512] *Id.* at 22-24.
[513] *Id.* at 24-25.
[514] *Id.* at 26-27.
[515] *Id.* at 28-29.
[516] *Id.* at 29-30.
[517] *Id.* at 30-31.

持續監控模型性能（準確性、偏見、毒性），並設定偏離指標時的應急計畫。

2. 人工介入監督[518]

對於高風險應用（如貸款審查或客戶賠償），必須有人工介入審核，如抽樣檢查或逐案審查。

3. 用戶反饋與迭代改進[519]

建立機制讓用戶（行員或客戶）回報不準確的輸出。

4. 用戶透明度與同意[520]

清楚標示內容係由 AI 生成，並適當加註免責聲明，提醒用戶核實資訊。

四、 關於生成式 AI 在銀行業面臨的新興風險與機會白皮書

基於真理倡議，2023 年 MAS 推動「MindForge 計畫[521]」，並號召產業界組成 MindForge 聯盟，旨在建立金融業的生成式 AI 風險框架。MindForge 計畫第一階段由 MAS、多家金融機構（如花旗銀行、星展銀行、匯豐銀行、渣打銀行等）及科技公司（如 Google、Microsoft 等）所組成的 MindForge 聯盟進行，並於 2023 年 11 月第一階段結束時發布此份白皮書。

生成式 AI 透過大型語言模型所創造的新內容為銀行業帶來轉型的潛力[522]，但也放大或引入了特定風險，共可歸納為以下七大維度：

（一） 公平性與偏見[523]

訓練數據可能代表性不足或存在偏差，導致對特定群體產生不利影響。

（二） 道德與衝擊[524]

包括價值觀不一致、對環境永續性的影響以及生成有毒或冒犯性的內容等。

---

[518] *Id.* at 31-32.
[519] *Id.* at 33.
[520] *Id.* at 34.
[521] Monetary Authority of Singapore, *Project MindForge*, MONETARY AUTHORITY OF SINGAPORE SCHEMES AND INITIATIVES, https://www.mas.gov.sg/schemes-and-initiatives/project-mindforge (last visited Feb. 12, 2026). MindForge 聯盟是由新加坡金融管理局主導，聯合全球頂尖金融機構、科技巨頭與行業協會共同組成的人工智慧（AI）金融治理合作組織。
[522] The MindForge Consortium, *Emerging Risks and Opportunities of Generative AI for Banks*, 8-9, https://www.mas.gov.sg/-/media/mas-media-library/schemes-and-initiatives/ftig/project-mindforge/emerging-risks-and-opportunities-of-generative-ai-for-banks.pdf (last visited Jul. 5, 2026). 白皮書中指出，生成式 AI 的主要的應用場景包含：開發助手（Development Assistant）、風險識別（Risk Identification）、知識管理（Knowledge Management）、市場研究（Market Research）、銷售效率（Sales Efficiency）、超個人化行銷（Hyper-Personalised Marketing）、拒絕理由的個人化解釋（Personalised Explanation for Denial）。
[523] *Id.* at 10.
[524] *Id.*

（三） 問責與治理[525]

第三方模型缺乏透明度導致問責不清、缺乏人類監督、以及對生成式 AI 風險意識不足。

（四） 透明度與可解釋性[526]

輸出準確性不明、訓練數據來源不明、缺乏可解釋性，以及「擬人化（Anthropomorphism）」的互動可能誤導用戶，使用戶誤以為正在與真人互動。

（五） 法律與監管[527]

包含資料與數據的所有權不清、智慧財產權的侵權風險、隱私保護不足等問題。

（六） 監控與穩定性[528]

包含幻覺或虛構內容、模型老化、模型退化及過度自信等問題。

（七） 網絡與數據安全[529]

包含資料投毒、提示詞注入攻擊、洩露敏感資訊等。

針對以上七大維度風險，白皮書進一步對當時已發布的 FEAT 原則及真理倡議工具進行評估，認為 FEAT 原則仍大致適用，但需要針對生成式 AI 的特性進行補強。主要的缺口有三：一是公平性[530]，過往僅在 AI 預測或決策階段關注公平性問題，但在生成式 AI 領域，對公平性的關注需要擴展到整個生成式 AI 的生命週期，包含訓練資料；二是透明度[531]，雖然 FEAT 原則足以處理幻覺和智慧財產權問題，但仍應思考如何提升對終端用戶的透明度；三是建議在 FEAT 原則之外，新增對於智慧財產權、隱私權、幻覺監控以及提示詞注入等生成式 AI 特有安全風險的指導原則[532]。另外，由於金融機構通常無法完全控制第三方的 AI 模型，為避免問責困難，白皮書建議在合約中明確定義 AI 供應商的角色與責任[533]。

五、 AI 風險管理執行手冊

此執行手冊是由 MindForge 聯盟於 2025 年 11 月所發布。相較於兩年前發布白皮書時，已有更多金融機構（如貝萊德公司、友邦保險公司、英國保誠集團等）及科技公司（如 Amazon）加入 MindForge 聯盟。此執行手冊屬於MindForge專案的第二階段成果，旨在為金融機構提供一套AI治理與風險管理指南，協助金融機構於導入生成式 AI 與代理式 AI（Agentic AI）時落實風險管理。

---

[525] *Id.*
[526] *Id.* at 11.
[527] *Id.*
[528] *Id.* at 12.
[529] *Id.*
[530] *Id.* at 26.
[531] *Id.*
[532] *Id.* at 26-27.
[533] *Id.* at 27-28.

此手冊建立在 FEAT 原則及真理倡議的基礎上，但傳統的治理框架制定時並未考慮到生成式 AI 或代理式 AI，因此執行手冊特別針對生成式 AI 或代理式 AI 進行補充。執行手冊特別點出生成式 AI 或代理式 AI 的十大風險：具代表性不足或有偏見的數據輸入（Unrepresentative or Biased Data Inputs）、有毒或冒犯性的輸出（Toxic and Offensive Outputs）、缺乏 AI 風險意識（Lack of AI Risk Awareness）、缺乏使用案例、數據和模型治理（Lack of Use Case, Data and Model Governance）、人為監督不足（Inadequate Human Oversight）、反饋與追索機制不足（Inadequate Feedback and Recourse Mechanisms）、幻覺/捏造/虛構（Hallucination/Fabrication/Confabulation）、過度自信（Overconfidence）、模型準確性/穩健性不足（Insufficient Model Accuracy/Soundness）、非預期使用導致的模型退化（Model Degradation from Unexpected Use）。

執行手冊分為四個部分，分述如下：

（一） 範圍與監督（Scope & Oversight）

強調管理手段應與風險相稱[534]（Proportionality），並將資源集中在重大風險上。金融機構可以選擇是否要設立新單位專職處理 AI 治理事務[535]，亦可利用現有的內部單位（如董事會或高階管理人員）進行治理，並確保董事會與高階管理人員具備 AI 治理和風險管理的職能。

（二） AI 風險管理（AI Risk Management）

金融機構應更新治理文件以明確定義 AI 相關概念與責任，並將 AI 風險納入關鍵風險指標（Key Risk Indicators, KRIs）與風險胃納中[536]。針對第三方 AI 產品的供應商，金融機構應提升採購管理措施，包括要求適當的揭露、合約條款修訂以及模型的變更檢測[537]。另外，金融機構應根據影響力、複雜度等標準，對每個 AI 的應用場景進行風險分級（Materiality Assessment），區辨 AI 應用場景的「固有風險（Inherent Risk）」與「剩餘風險（Residual Risk）」，以制定部署前的評估流程，以及與部署後的定期審查程序[538]，並應建立 AI 清冊記錄所有 AI 應用場景的核心資訊，如用途、數據來源、模型類型、風險等級等[539]。

（三） AI 生命週期管理（AI Lifecycle Management）

---

[534] The MindForge Consortium, *AI Risk Management: Executive Handbook*, 11, https://www.mas.gov.sg/-/media/mas-media-library/schemes-and-initiatives/ftig/project-mindforge/mindforge-ai-risk-management-executive-handbook.pdf (last visited Jul. 5, 2026).
[535] *Id.* at 13.
[536] *Id.* at 14.
[537] *Id.* at 15.
[538] *Id.* at 16.
[539] *Id.* at 17.

AI 使用案例的生命週期可分為五大階段：

1. 情境與設計[540]（Context & Design）
   確保用途符合道德與法規，並設計適當的人為監督機制。
2. 數據獲取與處理[541]（Data Acquisition & Processing）
   確保使用數據的合規性，並減少數據中的偏見。
3. 導入、構建與審查[542]（Onboarding, Build & Review）
   在開發過程中設置護欄和指標，並在部署前對 AI 進行審查。
4. 部署[543]（Deployment）
   制定監控計畫、應急措施，並考慮分階段推出。
5. 使用、監控與變更管理[544]（Usage, Monitoring & Change Management）
   持續監控輸入模型的資料品質、偏差及護欄的有效性，並提供用戶反饋管道。

（四） 賦能因素[545]（Enablers）

金融機構應確保員工具備 AI 治理技能，並培養負責任的 AI 文化。此外，亦應確保 AI 基礎設施的可擴展性、可用性和安全性。

六、 MindForge AI 風險管理營運手冊

本手冊與另一份實施範例[546]由 MindForge 聯盟於 2026 年 1 月製作，並與 2025 年 11 發布的 AI 風險管理執行手冊共同搭配[547]，屬於 MindForge 計畫的第二階段成果。MindForge 計畫第一階段的主要目標在於識別出金融業運用 AI 的相關風險，並檢驗 FEAT 原則和真理倡議工具是否仍適用[548]。而 MindForge 計畫第二階段的主要目標則是「賦能」，以及幫助處於不同 AI 成熟度的金融機構快速且負責任的實現 AI 的規模化應用。然而 AI 風險的樣貌也正在迅速演變，金融機構與整體業界在使用本手冊時，必須持續考量新興的 AI 專屬風險[549]。

本手冊共分為四大部分：範圍與監督（Scope and Oversight）、AI 風險管理（AI Risk Management）、AI 生命週期管理（AI Lifecycle Management）、賦能因素（Enablers）。

---

[540] *Id.* at 18.
[541] *Id.* at 19.
[542] *Id.* at 20-21.
[543] *Id.* at 22.
[544] *Id.* at 22-23.
[545] *Id.* at 24.
[546] The MindForge Consortium, *MindForge AI Risk Management Implementation Examples*, https://www.mas.gov.sg/-/media/mas-media-library/schemes-and-initiatives/ftig/project-mindforge/mindforge-ai-risk-management-implementation-examples.pdf (last visited Apr. 24, 2026).
[547] The MindForge, *MindForge AI Risk Management Operationalisation Handbook*, 10, https://www.mas.gov.sg/-/media/mas-media-library/schemes-and-initiatives/ftig/project-mindforge/mindforge-ai-risk-management-operationalisation-handbook.pdf (last visited Apr. 24, 2026).
[548] *Id.* at 9.
[549] *Id.*

（一） 範圍與監督

本手冊適用於所有在業務中使用 AI 的金融機構。由於使用 AI 帶來了新的或更高的風險，超越了金融機構使用傳統軟體的風險，因此有必要補充 AI 專屬風險的管理架構。本手冊所指的 AI 包括傳統 AI、生成式 AI 與代理式 AI，其中生成式 AI 與代理式 AI 可能會引入新的或額外的風險與挑戰[550]。

本手冊的目標受眾有五類[551]，分別為：負責決策與領導的高階主管（Executives）、參與 AI 開發、部署與使用的建置者（Builders）、在金融機構中擔任監督、治理、賦能與風險管理角色的保管者（Custodians）、對 AI 應用當責的使用案例擁有者（Use Case Owners）、在執行業務時使用 AI 的業務使用者（Business Users）。金融機構可能會根據其結構與需求（例如依據三道防線概念），以不同的方式或職稱安排這些職能。

（二） AI 風險管理

1. 確保有效的 AI 相關政策、程序與標準

金融機構通常透過政策、程序和標準等治理文件進行風險管理，部分金融機構選擇制定專門的 AI 治理政策，部分則將相關規定嵌入現有的治理框架中，金融機構應根據自身的組織背景，選擇如何建立或更新相關治理文件。

AI 治理文件的重點有二，其一在於確保其已定義何謂 AI、辨別 AI 的標準，並持續更新[552]；其二為確保 AI 風險管理方法與流程已加入金融機構的內部制度中，並定期重新評估[553]。金融機構內部必須對於什麼技術屬於 AI 有一致的定義，可以參考應用程式開發者的說明或指定一個權威單位（如治理委員會）判定是否屬於 AI 應用。此外，金融機構可參考本手冊與其他相關指引，訂定 AI 風險管理的角色與職責、流程、規則和控制措施，並評估現有的控制流程（如存取管理、事故管理）是否仍適用於 AI 應用[554]，如有不足則應進行補強。

2. 強化組織層級的風險管理

金融機構雖有「企業風險管理」（如網路安全、技術風險），然而採用 AI 時會引入新的或更高的 AI 專屬風險，金融機構必須加強組織風險框架與風險胃納應對 AI 專屬風險[555]。金融機構可以自行選擇如何加強組織

[550] *Id.* at 8.
[551] *Id.* at 9.
[552] *Id.* at 29.
[553] *Id.* at 31.
[554] *Id.* at 36.
[555] *Id.* at 34.

的風險框架[556]，例如：採用分散式作法（將 AI 專屬風險加入現有的資安風險架構中）、集中式作法（將 AI 專屬風險歸類在現有的模型風險類別中）或採用專屬作法（建立全新的 AI 風險類別）。不同作法的差異在於管理角色的不同，分散式作法會由現有企業風險的委員會管理 AI 專屬風險；集中式作法會由專屬監督委員會管理 AI 專屬風險；專屬作法會由新設立的 AI 專屬監督委員會管理 AI 專屬風險。

此外，金融機構尚需確保其具備測量 AI 專屬風險的 KRIs，並追蹤事故與問題。常見的 KRIs 包含但不限於[557]：問責與治理（如未列入 AI 清單的使用案例比例）、透明度與可解釋性（如「黑盒」AI 系統的數量）、法律與合規（如 AI 產出的內容所導致的第三方智慧財產權侵權索賠數量）、監測與穩定性（如異常率、模型漂移指標）。

3. 提升第三方 AI 風險管理

金融機構越來越依賴第三方提供的 AI 模型，這些第三方提供的 AI 模型主要可以分為幾種部署模式：一為客製化導入（Onboarding With Customisation），即採購後進行客製化微調；直接導入（Onboarding Without Customisation），即直接將第三方模型作為軟體即服務（Software as a Service, SaaS）或本地使用；嵌入式 AI（Embedded AI），即產品或服務中內含 AI 組件但非主要產品；連結服務（Connected Service），即產品或服務本身不含 AI，但價值鏈涉及 AI，例如由 AI 生成內容。

然而無論採用何種部署方式，依賴第三方 AI 模型都將導致資訊不對稱的風險增加，以及產生責任分擔問題。金融機構應提高採購管理與第三方管理措施[558]，包括使用揭露模板（如 ISO/IEC 42001:2023[559]或本手冊附錄 E 的模板）、進行供應商評估、增訂契約條款[560]及確保團隊具備 AI 專業知識[561]，並確保在採購、部署後整個生命週期中的適當階段持續評估 AI 特有的風險[562]。

此外，即使是先前已經採購的產品或服務，供應商也可能嵌入或連結 AI，金融機構也必須辨識其中是否含有 AI 風險[563]。

4. 強化實際使用場景的風險管理

---

[556] *Id.* at 35.
[557] *Id.* at 37.
[558] *Id.* at 43.
[559] 如 ISO/IEC 42001:2023 B.6.2.7 要求一般描述、使用說明、技術假設和限制、監控功能、模型或系統的開發和資料說明；ISO/IEC 42001:2023 B.8.2 要求管理系統的相關資訊等。
[560] The MindForge, *supra* note 547, at 49.
[561] *Id.* at 50.
[562] *Id.* at 45.
[563] *Id.* at 48.

過往金融機構主要依賴模型風險管理，這對單一特定目的傳統 AI 來說是足夠的（模型的風險等於實際使用場景的風險），但對生成式 AI 與代理式 AI 等通用模型則顯得不足[564]。

通用模型從設計、資料處理、建置到部署階段都有不同的風險程度，即風險重大性[565]（Risk Materiality）不同，金融機構應評估每個實際使用場景的風險重大性，以採取不同的、基於風險的治理方法[566]。因此，金融機構必須制定評估風險重大性的方法[567]、評估固有[568]與剩餘[569]風險、採用與風險相稱的控制措施[570]，以及在部署前後[571]進行 AI 專屬審查[572]。AI 專屬審查時必然會碰到黑盒問題，然而生成式 AI 的驗證重點在於實際應用場景中模型的功能與輸出，與傳統 AI 專注於模型本身有所不同[573]。也可以考慮以「追蹤訓練資料的影響」以及「將輸出連結至可驗證來源」等方式提升可解釋性[574]。

5. 確保 AI 清單

AI 清單是記錄金融機構實際使用 AI 場景的資料庫，建立即時且完整的清單，有助於企業管理各使用場景的風險、確保 AI 僅在已被核准的範圍內使用[575]。

金融機構可利用一個或多個現有的非 AI 清單系統進行紀錄，或直接建立新的系統[576]。有效的 AI 清單至少應記錄六個屬性[577]：使用案例的目的與範圍、使用的 AI 類型、使用的資料、AI 專屬風險與緩解措施、狀態與治理、第三方 AI 模型資訊。

---

[564] *Id.* at 45.
[565] *Id.* at 46.
[566] *Id.* at 46. 例如：「使用客戶個人資訊決定信用額度」屬於高風險；具備人工介入監督的「自動化理賠」屬於中風險；僅在內部網路與員工互動的「AI 知識管理聊天機器人」屬於低風險。
[567] *Id.* at 47. 可能的評估標準包含：對金融機構、客戶或利害關係人的影、個人資料的使、AI 的複雜性或新穎性、依賴程等。
[568] *Id.* at 48-49. 固有風險重大性（Inherent Risk Materiality）的評估應在生命週期早期（如設計階段）進行。
[569] *Id.* at 50-51. 剩餘風險重大性（Residual Risk Materiality）評估是在防護措施（護欄）實施後，於真實部署的場景下對 AI 的行為進行評估，可以採用模擬測試、紅隊測試等方式進行。
[570] *Id.* at 52. 例如：低風險可採抽樣人工審查；高風險則採強制全面人工審查。
[571] *Id.* at 56. 部署後定期進行 AI 專屬審查的頻率取決於實際使用場景的風險重大性、法規要求、距上次審查的時間、模型或系統的變更，以及外部風險（如媒體報導的負面事件）等因素。
[572] *Id.* at 53-55. AI 專屬審查應由「未直接參與開發、部署或營運」的獨立第三方（或具備自主性的團隊）進行。且針對生成式 AI，審查重點不再只是傳統的模型驗證，而是產出的準確性、有無幻覺、冒犯性內容、過濾輸出的有效性等。
[573] *Id.* at 53.
[574] *Id.* at 84.
[575] *Id.* at 60.
[576] *Id.* at 61.
[577] *Id.* at 62-64.

另外，金融機構尚需確保清單的準確性，尤其是使用案例擁有者資訊的更新即時性與正確性[578]。

（三） AI 生命週期管理

AI 生命週期管理是確保負責任且有效部署 AI 實際使用場景的基礎。

（四） 實際使用場景的背景與設計（Use Case Context and Design）

此階段包含設定 AI 實際使用場景的業務背景與目標，並評估是否與組織的使命、價值觀、優先事項及風險胃納相符。已具備健全軟體開發生命週期（Software Development Lifecycle, SDLC）的金融機構可輕易將此階段整合至現有流程中[579]。

此階段首先須確定使用案例擁有者確保當責與監督架構[580]，並進行初步固有風險重大性評估[581]，以及將此實際使用場景的資訊記錄於 AI 清單中[582]。

（五） 資料擷取與處理（Data Acquisition and Processing）

此階段應持續落實現有的資料治理政策，並遵守相關的資料保護與隱私的法規[583]。換句話說，金融機構須確保資料的使用符合道德標準、組織政策以及法規要求。若資料第三方所有，則須符合智慧財產權規範、契約義務與授權條款[584]。此外，亦應確保 AI 實際使用場景所使用的資料具合適性、相關性、代表性、正當性，避免不準確、歧視或缺乏代表性，使用合成資料 （Synthetic Data） 時亦同[585]。若金融機構識別出資料存在偏見[586]，應實施偏見緩解措施，如重新加權訓練資料、對代表性不足的群體進行重新採樣等[587]。

另外，金融機構尚需確保實施適當的資料存取控制[588]，並指定衍生或轉換資料[589]的擁有權[590]，以維持監督、品質、存取管理與安全性。

---

[578] *Id.* at 65.
[579] *Id.* at 68.
[580] *Id.* at 70-71. 根據實際使用場景的性質與風險重大性，考量其是否需要以及需要何種程度的人為監督，如：人在迴圈內、人在迴圈上、人在迴圈外。
[581] *Id.* at 69.
[582] *Id.* at 70.
[583] *Id.* at 72.
[584] *Id.* at 73.
[585] *Id.* at 74-75.
[586] 可採用差別影響分析（Disparate Impact Analysis）與卡方獨立性檢定（Chi-Square Tests）等測試偵測偏見。
[587] The MindForge, *supra* note 547, at 78.
[588] *Id.* at 76.
[589] 例如：分析型 AI 的特徵市集（Data in Feature Marts）或生成式 AI 的向量/索引（Vector/Indexes）。
[590] The MindForge, *supra* note 547, at 77.

（六） 導入、建置與審查（Onboarding, Build, and Review）

在導入與建置階段的治理可確保 AI 實際使用場景的建置情況，使其未來可以可靠且合乎道德地運作。此階段包含三個步驟[591]，第一為導入（Onboarding），涉及供應商盡職調查與部署前審查；第二為建置（Build），涉及 AI 實際使用場景的開發或客製化，以及護欄的實施[592]；第三為審查（Review），涉及由獨立方在部署前對 AI 實際使用場景所進行的評估。若欲導入的 AI 為第三方 AI 模型，則應另行評估遞增風險[593]。

（七） 部署（Deployment）

部署是將 AI 實際使用場景投入生產環境的過程。在此階段應制定監控與應變計畫[594]，並考量分階段推出的可能性與必要性[595]。此外，應再次確保 AI 實際使用場景已適當記錄、已實施適當的治理、已制定資料保留要求、已取得相關部署核准等[596]。另外，針對潛在的使用者，金融機構也必須在此階段對其進行培訓[597]，提高使用者對 AI 專屬風險的意識、確保使用者具備必要的技能、資源與知識。

（八） 使用、監控與變更管理（Usage, Monitoring, and Change Management）

此階段應對 AI 實際使用場景進行持續的監控[598]與人工審查[599]，以確保其隨時間推移仍適合其目的、避免濫用或意外結果。此外，應記錄監控結果，並追蹤實際使用場景的事故與升級紀錄。部署後金融機構仍須持續監控內部與第三方資料集[600]，並定期再次進行 AI 專屬審查[601]。在適當情況下，亦應為終端使用者提供詢問、回饋或要求人工審查的管道[602]。

（九） 賦能因素

金融機構應提升員工的 AI 治理與風險管理技能[603]，可以透過教育訓練、建立吹哨者機制等方式達成。這些治理與風險管理技能應對應到不同角色[604]（高階主管、建置者、保管者、使用案例擁有者、業務使用者），如：高階主管需要道德判斷、建置者需要實施 AI 護欄技術、保管者需要

[591] *Id.* at 80.
[592] *Id.* at 83.
[593] *Id.* at 82.
[594] *Id.* at 92-93.
[595] *Id.* at 94.
[596] *Id.* at 96.
[597] *Id.* at 95.
[598] *Id.* at 99.
[599] *Id.* at 102.
[600] *Id.* at 100.
[601] *Id.* at 101.
[602] *Id.* at 103-4.
[603] *Id.* at 115.
[604] *Id.* at 116.

具備解釋 AI 指標的能力、使用案例擁有者需具備完成風險評估的能力、業務使用者需具備負責任地使用其工具的能力。金融機構亦可透過修訂相關技術使用的規則，進一步強化重視 AI 風險的企業文化[605]。

另外，金融機構應確保 AI 相關基礎設施（運算能力、資料儲存、雲端存取等）的可用性與安全性，這項基礎設施是 AI 的關鍵賦能因素，同時也是潛在 AI 風險的來源[606]。

第四節　既有法規範持續適用

在既有的內部治理法規方面，MAS 仍可透過「個人問責與行為準則（Guidelines on Individual Accountability and Conduct）」，將參與生成式 AI 或代理式 AI 等「新科技的開發與部署」之人員，納入「重大風險人員（Material Risk Personnel）」的範疇。此外，亦可透過「個人問責與行為準則」清楚界定風險長、資訊長/科技長及資安長在 AI 治理上的具體職能，使問責機制與企業管理層相互連結。

另考量生成式 AI 多仰賴跨國科技巨頭之雲端基礎設施，MAS 可將雲端運算與 AI 模型服務納入「銀行業委外指引（Guidelines on Outsourcing）」之「重大委外」標準中進行控管，並持續適用「科技風險管理指引（TRM Guidelines）」，要求金融機構持續防範其他傳統科技（如網路）所帶來的威脅與風險。

一、　個人問責與行為準則

本準則由 MAS 於 2020 年 9 月發布，旨在協助金融機構建立框架與最佳實踐，提升高階管理人員的個人問責制度、加強對重大風險人員的監督，並強化所有員工的行為標準。本準則並未直接提及生成式 AI ，然於評估與識別「重大風險人員（Material Risk Personnel, MRPs）」時，其中一項因素為「新科技的開發與部署[607]（Development or Deployment of New Technologies）」，意即，金融機構對其內部負責決策或部署生成式 AI 等新科技的人員，須滿足本準則中針對重大風險人員的問責與風險治理要求。

此外，本準則中亦在高階管理人員的核心職能中，明確列出了風險長、資訊長/科技長、資安長以及數據長的職能[608]，分述如下：

（一）　風險長（Chief Risk Officer / Head of Risk）

負責建立與執行風險管理框架，專注於識別、監控與管理金融機構「整體」的各項風險。

---

[605] *Id.* at 120.
[606] *Id.* at 122.
[607] Monetary Authority of Singapore, *Guidelines on Individual Accountability and Conduct*, 13, https://www.mas.gov.sg/-/media/mas/mpi/guidelines/guidelines-on-individual-accountability-and-conduct.pdf (last visited Jul. 5, 2026).
[608] *Id.* at 20.

（二） 資訊長/科技長（Chief Information Officer / Chief Technology Officer / Head of Information Technology）

負責建立與執行整體的資訊科技（IT）策略，監督日常的資訊科技營運。

（三） 資安長（Chief Information Security Officer / Head of Information Security）

負責金融機構的資訊安全策略與計畫，維護資安政策與程序、實施資安控制措施，以及處理資訊安全外洩事件。

（四） 數據長（Chief Data Officer）

負責建立與執行有關數據的政策、系統與流程，專注於數據的治理、使用與分析。

二、 科技風險管理指引

本指引由 MAS 於 2021 年 1 月發布，指引金融機構建立健全的科技風險管理框架，如 IT 系統管理、網路安全、軟體開發生命週期等。

由於本指引發布的時間較早，當時生成式 AI 尚未在金融產業形成廣泛的商業應用，因此本指引中並未直接對生成式 AI 進行規範。不過，本指引仍列出了金融業經常面臨的線上金融服務威脅[609]，如：程式碼注入攻擊（Code Injection Attack）、跨網站指令碼（Cross-Site Scripting）、中間人攻擊（Man-in-the-Middle Attack, MitMA）、網域名稱系統劫持（DNS Hijacking）、分散式阻斷服務（Distributed Denial of Service, DDoS）、惡意軟體（Malware）、偽造攻擊（Spoofing Attacks）等。

而由上述威脅樣態可以發現，傳統的科技威脅樣態並不能完全涵蓋生成式 AI 的威脅樣態，可見金融機構運用生成式 AI 時，必須進行額外的評估，並建立新的風險管理框架。

另外，本指引中亦提及[610]董事會與高階管理人員必須具備科技風險意識並建立企業的風險意識文化，並指派具備專業知識的資訊長與資安長管理科技風險，前者主要負責制定和實施整體資訊技術策略，以及監督日常資訊技術營運情況；後者主要負責制定資訊安全策略，包含相關程序與安全控制機制，以確保金融機構的資訊安全。本指引目前仍屬有效的指引，因此金融機構於落實 AI 風險管理框架時，亦須考慮如何將相關機制轉化為資訊長與資安長的具體任務。

三、 銀行業委外指引

[609] Monetary Authority of Singapore, *Technology Risk Management Guidelines*, 49, https://www.mas.gov.sg/-/media/MAS/Regulations-and-Financial-Stability/Regulatory-and-Supervisory-Framework/Risk-Management/TRM-Guidelines-18-January-2021.pdf (last visited Jul. 5, 2026).

[610] *Id.* at 7.

本指引係由 MAS 於 2023 年 12 月所發布，揭示銀行[611]必須將委外服務視同內部管理般進行嚴格的風險控管[612]，相關治理與風險管理責任仍應由銀行董事會與高階管理人員負擔[613]。若發生可能導致服務長期中斷或危及客戶資料安全與機密性的重大事件，銀行有義務盡速通知 MAS，MAS 亦有權要求銀行採取額外的措施或進行事前通報[614]。此外，銀行須對委外的供應商進行盡職調查[615]，並將委外的範圍、效能標準、監控與稽核權利等要求訂明於契約條款中[616]，若為「重大委外（Material Outsourcing）」服務，供應商分包前必須獲得銀行核准，且分包商也必須接受與主要供應商同等的要求[617]。

本指引並未直接提及有關生成式AI相關的內容，但MAS將「雲端運算服務（Cloud Computing）」視為一種委外形式[618]，若銀行將生成式 AI 的訓練或推論外包給第三方雲端服務商（如 Microsoft Azure、Amazon Web Services）時，仍需遵守本指引中有關「重大委外」的標準[619]，確保客戶資料受到保護。具體而言，是否具有重大性，需考量以下因素：擬委外業務活動的重要性（例如：對收入和利潤的貢獻度）；委外對銀行收益、償付能力、流動性、資金、資本及風險概況的潛在影響；若供應商未能履行服務或發生安全漏洞（如客戶資訊外洩），對銀行聲譽、品牌價值，以及達成業務目標、策略和計畫能力的影響；若服務供應商未能履行服務或發生安全漏洞，對銀行客戶的影響；若服務供應商未能履行服務，對銀行的交易對手及新加坡金融市場的影響；委外成本占銀行總營運成本的比例；委外失敗的成本（銀行將委外活動收回由內部處理的成本，或向其他供應商尋求類似服務的成本）占銀行總營運成本的比例；當銀行將多項功能委外給同一服務供應商時，對該特定服務供應商的整體曝險；若服務供應商面臨營運問題，銀行維持適當內部控制並符合監管要求的能力。

若銀行將所有或幾乎所有的風險管理或內部控制機制（如稽核、財務會計及精算）委外，則應視為重大委外。此外，銀行應定期審查其委

---

[611] MAS 另有針對非銀行業的金融機構發布委外指引（Guidelines on Outsourcing (Financial Institutions other than Banks)），規範的架構大致與銀行業相同，僅規範密度較低。如分包時，針對非銀行的指引僅要求金融機構保有監控能力，並在重大委外作業分包前須取得機構的核准；而針對銀行的指引則設有獨立且詳細的「使用分包商（Use of Sub-Contractor(s)）」章節。本研究引用規範密度較高的銀行業指引，非銀行業指引則不另外列出。

[612] Monetary Authority of Singapore, *Guidelines on Outsourcing (Banks)*, https://www.mas.gov.sg/-/media/mas-media-library/regulation/guidelines/bd/guidelines-on-outsourcing/guidelines-on-outsourcing-banks-updated.pdf (last visited Jul. 5, 2026).

[613] *Id.* at 6-7.

[614] *Id.* at 5.

[615] *Id.* at 8-10.

[616] *Id.* at 10-13.

[617] *Id.* at 13-15.

[618] *Id.* at 23-24.

[619] *Id.* at 22.

外業務。先前非重大的委外業務，可能因同一服務供應商的漸增服務、委外數量增加或委外性質改變而轉變為重大委外。

第五節　小結

由上述摘要可知，新加坡對於 AI 的管理措施和監管策略是自 2018 年起逐漸累積而成，因此許多準則或手冊均會提及相同的概念或內容，因此 2026 年 1 月最新發布的《MindForge AI 風險管理營運手冊》自然涵蓋了大部分的實踐與流程管理。

不過《MindForge AI 風險管理營運手冊》畢竟是以協助金融機構規模化運用 AI 為主要目標，對較上位的治理原則、特定的技術細節，以及傳統風險則相對著墨較少。故本研究以最新的《MindForge AI 風險管理營運手冊》為基礎，綜整以下表 3、MindForge AI 風險管理營運手冊對照表，便於讀者參採其他準則，建構完整的 AI 治理架構。

表 3、MindForge AI 風險管理營運手冊對照表

管理與監管面向	**內容**	**最新可參考資料**
上位階原則	FEAT 原則	《MindForge 白皮書》(2023 年 11 月)
管理職能與問責	人員問責制度	《個人問責與行為準則》(2020 年 9 月)
傳統科技威脅	持續防範傳統攻擊	《科技風險管理指引(TRM Guidelines)》(2021 年 1 月)
運用 AI 風險	1. 七大維度風險 2. 十大特定風險	1. 《MindForge 白皮書》(2023 年 11 月) 2. 《AI 風險管理執行手冊》(2025 年 11 月)
生成式 AI 特有漏洞	1. 新型態攻擊 2. 深偽與外洩	《與生成式 AI 相關網路風險資訊文件》(2024 年 7 月)
可解釋性 (黑盒對策)	1. 驗證重心移轉 2. 流程與技術補償	《MindForge AI 風險管理營運手冊》(2026 年 1 月)
AI 生命週期管理	1. 轉向場景風險管理 2. 五大階段管控	《MindForge AI 風險管理營運手冊》(2026 年 1 月)
委外管理(第三方依賴)	1. 重大委外標準	《銀行業委外指引(Guidelines on

	2. 盡職調查與合約約束 3. 分包管控	Outsourcing (Banks)》(2023 年 12 月)
AI 護欄	1. 系統層級雙護欄 2. 中央提示詞庫 3. 思維鏈(Chain of Thought)	《銀行業生成式 AI 護欄手冊》(2025 年 5 月)
測試與評估工具	生成式 AI 測試工具	Moonshot Toolkit (2024 年 5 月)

資料來源：本研究自行整理

第六章　臺灣金融產業現況與挑戰

現階段臺灣金融機構對生成式 AI 之導入，主要集中於內部行政作業、對客服務輔助及專業人員之決策支援用途，尚未作為授信、投資或風險評估等金融核心決策之最終自動化機制。然而，即便生成式 AI 多被定位為輔助工具，其所涉及之資訊處理模式、輸出內容之可靠性，以及對消費者權益與市場公平性可能造成之影響，仍對既有金融監理與消費者保護規範提出實質挑戰。

第一節　臺灣 AI 監管法制發展現況

一、　《人工智慧基本法》

臺灣甫於 2025 年 12 月 23 日經立法院三讀通過、2026 年 1 月 14 日經總統公布《人工智慧基本法》，國科會為中央主管機關[620]。《人工智慧基本法》所定事項如有涉及各目的事業主管機關職掌者，由各目的事業主管機關辦理[621]，並訂定「以風險為基礎」的管理規範[622]。由此可知，《人工智慧基本法》授權各目的事業主管機關按風險分類制定規則、指引或管理措施，必要時[623]應以法規命令限制或禁止。

而所謂「以風險為基礎」之規範，依《人工智慧基本法》第 16 條第 1 項規定，應由數位發展部（下稱數發部）參考國際規範制定框架，再由各目的事業主管機關訂定之。《人工智慧基本法》雖未例示數發部得參考之外國立法例，亦無立法理由可資參考，然由行政院與部分立法委員提案說明中可知，立法者主要係參考美國 NIST 於 2023 年 1 月發布[624]的《AI RMF 1.0》[625]，以及由歐洲議會和理事會於 2024 年 6 月 13 日通過的《AI Act》[626]。數

[620] 人工智慧基本法第 2 條第 1 項
[621] 人工智慧基本法第 2 條第 2 項
[622] 人工智慧基本法第 16 條
[623] 人工智慧基本法第 5 條第 1 項、第 16 條第 2 項後段
[624] NIST, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, https://doi.org/10.6028/NIST.AI.100-1 (last visited Jul. 5, 2026).
[625] 邱若華委員於 2025 年 3 月 14 日主提案之「人工智慧基本法草案」第 15 條說明、林宜瑾委員於 2025 年 6 月 3 日主提案之「人工智慧基本法草案」第 9 條說明、許智傑委員於 2025 年 6 月 3 日主提案之「人工智慧基本法草案」第 12 條說明、伍麗華 Saidhai Tahovecahe 委員於 2024 年 8 月 1 日主提案之「人工智慧基本法草案」第 11 條說明、行政院於 2025 年 8 月 28 日提出之「人工智慧基本法草案」第 9 條說明、林月琴委員於 2025 年 10 月 14 日主提案之「人工智慧基本法草案」第 10 條說明、徐富癸委員於 2025 年 10 月 14 日主提案之「人工智慧基本法草案」第 9 條說明、邱鎮軍委員於 2025 年 11 月 7 日主提案之「人工智慧基本法草案」第 10 條說明。
[626] 吳宗憲委員於 2024 年 4 月 12 日主提案之「人工智慧基本法草案」第十五條說明、臺灣民眾黨黨團於 2024 年 9 月 20 日主提案之「人工智慧發展及管理條例草案」第十八條說明、楊瓊瓔委員於 2025 年 3 月 7 日主提案之「人工智慧基本法草案」第 15 條說明、邱若華委員於 2025 年 3 月 14 日主提案之「人工智慧基本法草案」第 15 條說明、萬美玲委員於 2025 年 4 月 18 日主提案之「人工智慧基本法草案」第 11 條說明、許宇甄委員於 2025 年 4 月 18 日主提案之「人工智慧基本法草案」第 18 條說明、張嘉郡委員於 2025 年 4 月 18 日主提案之「人工智慧基本法草案」第 16 條說

發部已於 2026 年 7 月 7 日公布《人工智慧風險分類框架》[627]，協助各目的事業主管機關建立共通評估標準，並訂定以風險為基礎的管理規範。

綜觀《人工智慧基本法》之規定，主要可分為以下六大面向：

（四） 七大原則

為平衡AI之發展與公益之維護，《人工智慧基本法》第4條規定，政府推動 AI 之研發與應用，應遵守永續性、人類自主性、隱私性、安全性、可解釋性、公平性、問責性七大原則。

參考行政院提出之《人工智慧基本法草案》（下稱行政院版草案）第 3 條說明，第一款永續性（Sustainable Development）係參考 G7《廣島 AI 國際行動規範》（Hiroshima Process Code of Conduct for Organizations Developing Advanced AI Systems）訂定；第二款人類自主性係參考經濟合作暨發展組織（Organisation for Economic Co-operation and Development, OECD）2019 年公布之《人工智慧建議書》（OECD Recommendation on Artificial Intelligence）訂定；第三款個人資料隱私係參考美國 2022 年《AI 權利法案藍圖》訂定；第四款安全性亦係參考美國 2022 年《AI 權利法案藍圖》及新加坡 2024 年《生成式 AI 治理架構》訂定；第五款可解釋性係參考歐盟 2019 年《可信賴 AI 倫理準則》（Ethics Guidelines for Trustworthy AI）訂定；第七款問責性係參考新加坡 2024 年《生成式 AI 治理架構》訂定。

（五） 消極規定

《人工智慧基本法》第 5 條第 1 項規定，應避免 AI 之應用侵害人民生命、身體、自由或財產，破壞社會秩序、國家安全或生態環境，或偏差、歧視、廣告不實、資訊誤導或造假等情況。

上開規定原列於行政院版草案第 8 條第 1 項，且第 2 項接續規定數發部或其他相關機關得向各目的事業主管機關提供或建議評估驗證之工具或方法。然由數發部提供或建議評估驗證之工具或方法之規定，現改列於《人工智慧基本法》第 5 條第 3 項，且前項（第 2 項）係兒少權益保障相關規定，並非避免 AI 應用侵害人民之消極規定。究竟數發部所提供之驗證工具究係為了落實兒少權益保障，還是為了避免 AI 應用侵害人民，仍有待主管機關、立法機關進一步釐清是否為立法技術上的錯誤，抑或是立法者有意為之。

明、邱議瑩委員於 2025 年 5 月 23 日主提案之「人工智慧基本法草案」第 10 條說明、林宜瑾委員於 2025 年 6 月 3 日主提案之「人工智慧基本法草案」第 9 條說明、張雅琳委員於 2025 年 6 月 20 日主提案之「人工智慧基本法草案」第 12 條說明、林思銘委員於 2025 年 7 月 4 日主提案之「人工智慧基本法草案」第 11 條說明、范雲委員於 2025 年 7 月 18 日主提案之「人工智慧基本法草案」第 12 條說明、王美惠委員於 2025 年 8 月 22 日主提案之「人工智慧基本法草案」第 10 條說明、黃捷委員於 2025 年 10 月 28 日主提案之「人工智慧基本法草案」第 9 條說明。

[627] 數位發展部，AI 風險分類框架，數發部 AI 專區 - 重點政策，2026 年 3 月 23 日，https://moda.gov.tw/major-policies/ai/governance/19244（最後瀏覽日：2026 年 9 月 2 日）。

（六） 兒少權益保障

承前所述，《人工智慧基本法》第 5 條第 2 項規定，經認定為高風險之 AI 產品或系統，應明確標示注意事項或警語，以保護兒少權益。

《人工智慧基本法》草案經立法院三讀通過後，立法院同時作成三項附帶決議[628]，其中兩相均與兒少權益有關。其一為審查會通過之附帶決議，要求行政院於三讀後 3 個月內，針對 AI 應用完成「兒少影響性評估」並公開相關報告，並建立評估機制，以保障兒少身心健康及權益；其二為民進黨黨團之附帶決議，亦要求教育部應於人工智慧基本法公布後六個月內，制定並公告「人工智慧使用和學習指引」，作為高中以下學校推動 AI 教育及輔導學生使用的依據，顯見立法者對於兒少族群受 AI 影響問題的重視程度非常高。

（七） 個人資料保護

《人工智慧基本法》第 14 條規定，AI 研發及應用過程中，應避免不必要之個人資料蒐集、處理或利用。行政院版草案第 13 條說明更以數發部訂定之「隱私強化技術應用指引」為例，強調各目的事業主管機關應將個人資料保護納入其業管法令中。

而依 2025 年 11 月 11 日修正之個人資料保護法第 1 條之 1（尚未生效）規定，未來個人資料保護法之主管機關為個人資料保護委員會（下稱個資會），然於個資會之組織法尚未完成立法程序前，仍由各目的事業主管機關、地方政府主管。

（八） 勞工權益保障

《人工智慧基本法》第 15 條規定，政府除應積極運用 AI 保障勞工權益外，更應提供勞工相應的資源以弭平 AI 發展所造成之技能落差，並且應於勞工因 AI 而失業時提供就業輔導。然如何判斷勞工係因 AI 而失業，仍有待觀察。

（九） 問責制度

《人工智慧基本法》第 17 條規定，就高風險 AI 之應用應明定責任歸屬及歸責條件，並建立其救濟、補償或保險機制。本條明顯體現「以風險為基礎」之規範架構。換言之，針對第 16 條分類框架中的高風險 AI 應用，其管理規範必須包含責任歸屬的相關內容，以落實問責性。而依數發部所公

[628] 立法院，立法院第 11 屆第 4 會期第 14 次會議紀錄，立法院公報，115 卷 1 期，頁 18，2026 年 1 月 7 日。

布的《人工智慧風險分類框架》之分類，AI 的風險主要來自三種面向[629]：AI 系統本身之技術設計缺陷、部署後操作及人機互動問題、廣泛社會結構與環境衝擊。面對來自三種面向的風險，須進一步以「客觀固有風險原則」判斷風險程度[630]，再依《人工智慧基本法》第 5 條第 1 項、第 17 條第 1 項之立法說明，判斷是否屬於高風險 AI 應用[631]，以便採行必要措施，如明確其責任歸屬及歸責條件，建立其救濟、補償或保險機制等[632]。

然為免過度限制AI之發展，行政院版草案第11條說明中亦提到參考歐盟《AI Act》第二條第八項之規定，於人工智慧基本法第 17 條第 2 項針對尚未實際應用階段之 AI 應用訂定豁免問責之規定。

二、 金融業專屬規範

依據上述《人工智慧基本法》之規定，就金融業運用 AI 方面，金管會本於其目的事業主管機關之職權，應依數發部提出之人工智慧風險分類框架訂定相關法規命令或規則，惟於風險分類框架尚未確定、金管會尚未修訂業管法令前，金融機構仍應遵循以下指引及自律規範。

（一） 金管會已訂定之指引

1. 金融業運用 AI 之核心原則

金管會早在 2023 年 10 月便公布《金融業運用 AI 之 6 項核心原則》，分別為：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、促進永續發展。此 6 項核心原則的公布時點雖然早於前開人工智慧基本法三讀通過的時點，但其內容實際上與人工智慧基本法第 4 條所定七大原則高度一致。

而金管會為因應 AI 發展，於公布 6 項核心原則時，亦預計推出 8 項配套政策[633]，包括：訂定《金融業運用人工智慧（AI）指引》、檢視相關規範並適時進行法規調適、利用 AI 技術發展監理科技、與國際組織及其他國家金融監理機關進行交流及合作、鼓勵金融業積極參與 AI 的研發與應用及協

---

[629] 數位發展部，AI 風險分類框架，數位發展部，頁 1，2026 年 7 月 7 日，https://www-api.moda.gov.tw/File/Get/moda/zh-tw/UieEfux5Yv0iJuR（最後瀏覽日：2026 年 9 月 2 日）。

[630] 註 629，頁 6。「以該 AI 應用本身之客觀特性為準，只要有造成危害之可能即為已足，不以實際發生損害為必要。評估時不考慮現有法規、行政措施或技術手段之緩解效果。」

[631] 註 629，頁 15-18。目前數發部僅例示部分高風險 AI 應用，如：信用評等遭拒卻無申訴與解釋機制（缺乏透明性或可解釋性）、特定族群遭排除於基本金融服務之外（不公平的歧視或偏見）、深度偽造技術騙取大額資產（詐欺與深偽技術濫用）、AI 輔助輿論操控引發市場恐慌或錯誤投資風潮，導致投資人重大損失（認知作戰與資訊主權）等。

[632] 註 629，頁 9。

[633] 金融監督管理委員會，金管會公布金融業運用人工智慧(AI)之核心原則及政策，金融監督管理委員會新聞稿，2023 年 10 月 17 日，https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0%2C2&mcustomize=news_view.jsp&dataserno=202310170002&dtable=News（最後瀏覽日：2026 年 7 月 5 日）。

助導入最佳實務做法、對金融業者應用 AI 之實際狀況檢視、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、督導金融機構落實公平對待客戶及金融友善準則，以及透過金融知識宣導活動降低數位落差等措施。

2. 《金融業運用人工智慧（AI）指引》

金管會依據上開 6 項核心原則，並同時參考他國政府及國際組織發布的相關指引文件，於 2024 年 6 月發布《金融業運用人工智慧（AI）指引》，此指引屬行政指導性質[634]，可做為金融機構導入、使用及管理 AI 的參考。

《金融業運用人工智慧(AI)指引》主要分為總則及 6 個章節，總則規範相關定義、風險評估因素、第三方業者監督管理等共通事項；各章節則分別說明金融業於落實 6 項核心原則時宜關注的重點以及可採行的措施。

3. 《金融資安韌性發展藍圖》

而 AI 的發展過程中，必須確保系統穩健性與安全性。因此，金管會於 2025 年 12 月所發布的《金融資安韌性發展藍圖》中亦特別提到[635]，傳統資安政策無法直接套用在 AI 系統上，需有 AI 專屬的資安框架。

（二） 自律規範

1. 《金融機構運用人工智慧技術作業規範》

由中華民國銀行商業同業公會全國聯合會（下稱銀行公會）於 2024 年 3 月 14 日公告實施、2025 年 10 月 2 日修訂的自律規範，是各金融機構同業公會中最早公告 AI 相關自律規範之公會，其公告時間甚至比金管會發布《金融業運用人工智慧（AI）指引》的時點還要早數個月，對自律規範具有指標性意義。銀行公會金融機構運用 AI 技術作業規範（下稱銀行公會自律規範）內容涵蓋 AI 系統的治理架構、風險管理、資料保護、公平性保障、可解釋性要求，以及文件保存與稽核機制，以對應金管會的 6 項核心原則。

銀行公會自律規範首先對 AI 及生成式 AI 進行定義[636]，前者係指底層使用大數據及機器學習技術進行預測、決策等模仿人類學習、思考及反應模式的技術，即傳統 AI；後者則不限於使用機器學習技術，任何透過大量資料學習生成模擬人類智慧、創造內容的相關技術均屬之[637]。

---

634 金融監督管理委員會，註 4。

635 金融監督管理委員會，金融資安韌性發展藍圖，頁 16-17，金融監督管理委員會，2025 年 12 月，https://www.fsc.gov.tw/uploaddowndoc?file=news/202601021913000.pdf&filedisplay=%E9%87%91%E8%9E%8D%E8%B3%87%E5%AE%89%E9%9F%8C%E6%80%A7%E7%99%BC%E5%B1%95%E8%97%8D%E5%9C%96.pdf&flag=doc（最後瀏覽日：2026 年 7 月 5 日）。

636 金融機構運用人工智慧技術作業規範第 2 條

637 此定義似乎使生成式 AI 成為 AI 的上位概念，然而實際上，早期的（傳統）AI 和近年的生成式 AI，二者底層所採用的資料技術有很大的差異。詳言之，傳統 AI 所採用的機器學習（Machine Learning, ML）技術，是利用演算法（如隨機森林、XGBoost、Logistic Regression 等）從大量結構化資料（如以資料表呈現或帶有註釋的影像和音訊資料集）中提取規律；而生成式 AI 則是透過深度學習（Deep Learning, DL）技術，從大量非結構化資料（無需手動標記）中自動提取特徵，

而第 3 條規定，金融機構運用 AI 與消費者互動、提供金融商品建議、提供客戶服務、影響客戶金融交易權益、或對營運有重大影響者，應依金融機構內部之風險評估基準進行評估。換言之，除了直接面對客戶的業務之外，若是運用 AI 進行或輔助決策，且決策結果會影響客戶的金融權益，如信用評等、凍結帳戶等，亦須進行風險評估。

第 6 條規定，為了防止 AI 產生不公正或偏見，金融機構應評估模型是否將特定參數納入判斷，如姓名、國籍、族群、出生地、年齡，以及所有生理特徵。AI 的演算法設計、資料蒐集、模型訓練以及後續應用，都必須符合「公平待客原則」。

第 9 條後段規定，運用生成式 AI 技術之人員，應清楚了解生成式 AI 技術運作方式以及如何做出回應[638]。

第 13 條規定，金融機構應以風險為基礎導向，建立適當之風險管理機制，並得由專業獨立的第三方出具評估報告。金融機構運用生成式 AI 所產出之資訊，也需由金融機構就其風險進行評估與管控。也就是說，除了前階段依據應用 AI 的類型進行風險評估與管控外，對於後階段生成式 AI 所產出的內容，金融機構仍需有一套明確的內容評估與管控機制[639]。

此外，第 14 條規定，金融機構對於涉及營業執照所載業務項目或與客戶資訊相關的 AI 作業委外時，應適用《金融機構作業委託他人處理內部作業制度及程序辦法》。

2. 《期貨業運用人工智慧技術自律規範》

由中華民國期貨業商業同業公會（下稱期貨商公會）於 2024 年 11 月 6 日公告實施的自律規範，架構大致上與銀行公會公布的自律規範相似，在第 3 條及第 13 條規範適用範圍及風險管理機制。另外特別針對外國期貨商在臺分(子)公司訂定特別規定，得由外國總(母)公司出具其風險管理、定期檢視機制及評估報告。

而《期貨商公會自律規範》第 9 條後段與銀行公會自律規範第 9 條後段之用字相同[640]，運用生成式 AI 技術之人員，應清楚了解生成式 AI 技術運作方式以及如何做出回應。

這使得二者擅長的領域有所不同。傳統 AI 善於利用規律進行預測，生成式 AI 則善於處理語言、聲音的細微差別；傳統 AI 難以處理自然語言（NLP），生成式 AI 也難以理解資料表結構並準確擷取資訊。參 AWS 方案指引，生成式 AI 與傳統 ML 之間的資料差異，https://docs.aws.amazon.com/zh_tw/prescriptive-guidance/latest/strategy-data-considerations-gen-ai/differences.html（最後瀏覽日：2026 年 7 月 5 日）。

[638] 此部分規定似乎是想解決生成式 AI 的黑盒問題，則是否代表銀行必須在確保生成式 AI 沒有黑盒問題的前提下始能運用生成式 AI？此部分仍有待觀察。

[639] 然就銀行須為生成式 AI 所產出的資訊全權負責一事，仍需結合「金融機構作業委託他人處理內部作業制度及程序辦法」第 7 條第 2 項規定一併解釋，詳後述。

[640] 此處同樣會遭遇黑盒問題，期貨商是否必須在確保生成式 AI 沒有黑盒問題的前提下始能運用生成式 AI？此部分仍有待觀察。

另外，《期貨商公會自律規範》第 14 條也與銀行公會自律規範第 14 條相似，明確規定期貨業對於涉及營業執照所載業務項目或與客戶資訊相關的 AI 作業委外時，應適用「期貨商作業委託他人處理應注意事項」。

3. 《證券商運用人工智慧技術自律規範》

由中華民國證券商業同業公會（下稱券商公會）於 2024 年 11 月 21 日公告實施的自律規範，架構大致上與銀行公會、期貨商公會公布的自律規範相似，都在第 3 條、第 13 條規範了適用範圍及風險管理機制，也針對外國證券商在臺分(子)公司訂定特別規定，得由外國總(母)公司出具其風險管理、定期檢視機制及評估報告。

而《券商公會自律規範》第 9 條第 2 項與《銀行公會自律規範》第 9 條後段的用語雖有不同，但仍是要求運用生成式 AI 技術之人員應確保其生成的過程與結果合乎邏輯、可預測[641]。

另外，券商公會自律規範第 14 條亦與銀行公會自律規範第 14 條相似，明訂證券商對於涉及營業執照所載業務項目或與客戶資訊相關的 AI 作業委外時，應適用「證券商作業委託他人處理應注意事項」。

4. 《證券投資顧問事業運用人工智慧技術自律規範》

由中華民國證券投資信託暨顧問商業同業公會（下稱投信投顧公會）於 2025 年 1 月 23 日公告實施的自律規範。該自律規範第 1 條即明確表示係參照金管會發布的《金融業運用人工智慧(AI)指引》訂定。投信投顧公會自律規範的架構大致上與銀行公會、期貨商公會、券商公會公布的自律規範相似，於第 3 條、第 13 條訂定適用範圍及風險管理機制，並特別針對外國金融機構集團在臺分公司或子公司訂定特別規定，得由母集團出具風險管理、定期檢視機制及評估報告。

至於投信投顧公會自律規範第 9 條第 2 項之規範文字與券商公會自律規範第 9 條第 2 項之規範文字一致，運用生成式 AI 技術之人員應確保其生成的過程與結果合乎邏輯、可預測[642]。

另外，投信投顧公會自律規範第 14 條與銀行公會自律規範第 14 條相仿，明確規範事業對於涉及營業執照所載業務項目或與客戶資訊相關的 AI 作業委外時，應依據「證券投資信託事業證券投資顧問事業作業委託他人處理應注意事項」規定辦理。

5. 《保險業運用人工智慧系統自律規範》

由中華民國人壽保險商業同業公會與中華民國產物保險商業同業公會（以下合稱保險公會）於 2025 年 4 月 10 日公告實施的自律規範。該自律規範第 1 條亦明確表示係參照金管會發布的《金融業運用人工智慧（AI）指引》訂定。

[641] 實質上仍存在券商是否必須確保生成式 AI 沒有黑盒問題始能運用生成式 AI 的問題。

[642] 實質上也存在是否必須確保生成式 AI 沒有黑盒問題始能運用的問題。

與銀行公會、期貨商公會、券商公會所公告的自律規範略有不同，保險公會自律規範第3條雖仍規範適用範圍，但就風險管理機制則定於第4條，且針對生成式 AI 所產生的內容，改與第9條的公平性及以人為本的價值觀並列，強調其內容對消費者的公平性問題。

此外，《保險公會自律規範》第9條第4項採用與《銀行公會自律規範》第9條後段相似的規範，保險業應了解生成式 AI 的運作方式以及如何做出回應[643]。換言之，除了關注AI生成的結果之外，更應了解AI生成的過程。

另外，對於委外作業，保險公會自律規範列於第5條第1項，保險業對於涉及依保險法令規定得從事之業務項目或與客戶資訊相關的 AI 作業委託第三方業者辦理時，應適用《保險業作業委託他人處理應注意事項》。

（三） 其他相關法令

1. 《金融機構作業委託他人處理內部作業制度及程序辦法》[644]

本辦法係由金管會依銀行法第45條之1第3項及信用合作社法第21條第4項訂定。

其中第 2 條規定，本國銀行及其國外分行、外國銀行在臺分行、信用合作社、票券金融公司及信用卡業務機構，於將作業委外處理時，必須與服務供應商簽訂書面契約。而針對服務供應商之故意或過失致客戶權益受損，第7條第2項規定，金融機構仍應對客戶依法負同一責任。

第 4 條及第 6 條第 1 項規定，委外內部作業規範需經董（理）事會核定，並設立專責單位，定期向董（理）事會呈報委外事項的執行情形。

另外，就委外事項的重大性，第4條第5項規定：

（1） 營運與資安影響

若委外作業無法提供服務，或出現資安疑慮，對金融機構的業務營運會造成重大影響；

（2） 客戶資料安全

委外作業涉及客戶的資料安全，對金融機構或是客戶的權益會造成重大影響；

（3） 概括條款

其他委外作業對金融機構或其客戶權益造成重大影響的情況。

而金管會另於「『金融機構作業委託他人處理內部作業制度及程序辦法』相關問題適用解說問答集」說明4，進一步例示七項考量因素：委外

[643] 則是否代表保險業也必須在確保生成式AI沒有黑盒問題的前提下始能運用生成式AI？此部分亦仍待觀察。

[644] 另有針對期貨業的「期貨商作業委託他人處理應注意事項」、針對證券業的「證券商作業委託他人處理應注意事項」、針對投信、投顧業的「證券投資信託事業證券投資顧問事業作業委託他人處理應注意事項」、針對保險業的「保險業作業委託他人處理應注意事項」，惟在風險控管、重大性評估、雲端資料加密與境內備份等方面的核心規定，以上注意事項的規定均非常相似，故本研究只取「金融機構作業委託他人處理內部作業制度及程序辦法」進行摘錄與說明，其餘不再逐一摘錄。

作業是否屬金融機構之關鍵業務；委外作業對盈餘、償付能力、流動性、籌資能力、資本及風險之潛在影響；委外作業倘未能提供服務或發生資料保護或資訊安全問題，將對金融機構之聲譽、營運目標之達成、客戶權益、交易對手或整體金融市場造成重大影響；委外作業之成本；委外作業失敗造成金融機構將委外作業移回金融機構本身、或尋找其他受託機構提供服務所衍生之成本；金融機構如將不同委外作業委託同一家受託機構提供服務，金融機構對該受託機構之總暴險；受託機構面臨營運問題時，金融機構仍可維持適當內控制度及符合法規要求之能力。

2. 《金融控股公司及銀行業內部控制及稽核制度實施辦法》

本辦法由金管會依金融控股公司法第 51 條、銀行法第 45 條之 1 第 1 項、信用合作社法第 21 條第 1 項、票券金融管理法第 43 條及信託業法第 42 條第 3 項規定訂定，對金控、銀行、信用合作社、票券商、信託業均有適用[645]。

本辦法雖然並非針對使用 AI 所訂定，但第 7 條第 3 款規定，控制作業的執行範圍應包括金融機構的「所有科技環境」，因此金融機構導入新的 AI 系統時，亦須將之納入內部控制制度中，並應依第 37 條規定，建立資訊安全防護機制及緊急應變計畫。金管會更於 2021 年增訂第 38 條之 1，要求金融機構指派資訊安全長，並設置資訊安全專責單位。

此外，第 8 條規定，金融機構須對「客戶資料保密」訂定作業程序，因此當金融機構使用生成式 AI 需要取用客戶資料時，均須依照內部作業程序辦理。

第 34 條第 1 項第 3 款規定，金融機構於推出各項「新服務」前，須出具符合法令及內部規範之法令遵循意見，因此當金融機構將生成式 AI 包裝成新服務（如 AI 理財顧問、智能客服等），推出前仍需經由法令遵循單位審核始得上架。

另外，第 35 條規定金融機構應建立獨立有效的風險管理機制，以評估整體風險承擔能力、已承受風險現況，並決定風險因應策略。則使用生成式 AI 所可能產生的各項風險，都必須納入此機制的評估範圍中。

3. 《銀行業導入責任地圖制度自律規範》

本自律規範由銀行公會訂定，目的在於建置以責任為基礎之公司治理架構，並強化對高階管理人員之問責機制[646]。本自律規範並非針對生成式 AI 所訂定，但已訂定「資訊安全長職能[647]（負責資安監控與管理）」及

[645] 金管會另針對保險業訂有「保險業內部控制及稽核制度實施辦法」；針對證券及期貨業訂有「證券暨期貨市場各服務事業建立內部控制制度處理準則」。雖然法源不同，但在內部控制與稽核的架構上高度一致，本研究僅舉其一進行摘要與說明，其餘不再逐一說明。

[646] 銀行業導入責任地圖制度自律規範第 1 條。

[647] 資訊安全長職能：綜理資訊安全政策推動及資源調度事務，負責督導資訊安全制度之規劃、監控、管理及執行。

「風險管理單位主管職能[648]（負責辨識與監控各項風險）」。若銀行業導入生成式 AI 卻發生資安問題或營運風險，可能循此框架追究相關高階主管之責任。

本自律規範主要可分為以下四大面向：

（1） 適用對象與職能劃分

本自律規範適用於「高階管理人員」，包含董事長、總經理、副總經理及總行各部門主管等，即負責關鍵職責並具決策權之人員[649]。此外，相關的管理職能必須分配給合適的人員。

（2） 責任地圖內涵[650]

一為公司治理架構圖，用於說明董事會及轄下各委員會之組織架構；二為權責分配表，確認各高階管理人員所擔負之個 人權責範圍，並包含總體責任；三為內部呈報流程圖，梳理呈報程序及對應呈報之高階管理人員。

（3） 執行責任地圖之階段[651]

一為定責，即梳理組織架構與業務，釐清高階管理人員的權責；二為當責，使高階管理人員明確瞭解權責範圍，並評估資源是否充足；三為問責，當發生重大違失時，應被問責的對象；四為究責，評估被問責對象是否盡到督導管理之責，並進行後續懲處程序。

（4） 治理與維運機制

董事會負責督導責任地圖的執行[652]，轄下應設置（或指定既有）具備「問責功能」的委員會負責問責事宜[653]，並建立完善的高階人員交接手續，相關交接紀錄[654]、責任地圖與責任聲明書等文件，自異動日起至少需保存 5 年[655]。

三、 其他規範

（一） 《資通安全管理法》

本法的主管機關為數位發展部與中央目的事業主管機關，其核心目標是為了建構安全的國家資通環境、保障國家安全與維護社會公共利益。規範對象分為「公務機關」與「特定非公務機關」。其中，特定非公務機關包含「關鍵基礎設施提供者」。而依據行政院國家關鍵基礎設施安全防護指導

[648] 風險管理單位主管職能：負責辨識、衡量、監控及管理銀行所面臨之各項風險，並建立風險管理機制，以評估及監督整體風險承擔能力、已承受風險現況、決定風險因應策略及風險管理程序遵循情形。

[649] 銀行業導入責任地圖制度自律規範第 5 條。

[650] 銀行業導入責任地圖制度自律規範第 9 條。

[651] 銀行業導入責任地圖制度自律規範第 6 條。

[652] 銀行業導入責任地圖制度自律規範第 7 條。

[653] 銀行業導入責任地圖制度自律規範第 8 條。

[654] 銀行業導入責任地圖制度自律規範第 11 條。

[655] 銀行業導入責任地圖制度自律規範第 12 條。

綱領之定義[656]，金融屬第一層主領域（Sector）之國家關鍵基礎設施，則金融機構應屬關鍵基礎設施提供者。

本法並非針對使用生成式 AI 所制定，然當金融機構導入生成式 AI 時，因其具備關鍵基礎設施提供者之身份，故仍須遵守本法之規範，如委外風險管理、資安產品限制、資料保護等。

本法規範重點如下：

1. 資安責任等級與維護計畫[657]

各機關須依其業務機敏性與規模，訂定資安責任等級 ，並據此制定、實施「資通安全維護計畫」。

2. 設置資安長[658]

公務機關及特定非公務機關皆必須指派副首長或高階主管擔任「資通安全長」，負責推動與監督資安事務。

3. 委外管理機制[659]

若機關將資通系統或服務委外辦理，必須要求並監督受託者建立有效的資安管理機制，且須簽訂載明權利義務與違約責任的書面契約。

4. 禁用危害國家資安產品[660]

原則上禁止下載、安裝或使用被認定為「危害國家資通安全產品[661]」的軟硬體；若因業務需求且無替代方案，須經資安長及主管機關專案核准。

5. 通報與應變[662]

機關必須訂定資安事件通報及應變機制；發生事件時，須向上級或主管機關通報，並提出調查與改善報告。

6. 稽核與罰則

主管機關或中央目的事業主管機關可定期稽核實施情形[663]。若特定非公務機關未依規定落實通報、訂定計畫或拒絕調查，主管機關或中央目的事業主管機關可裁處新臺幣 10 萬元至 1,000 萬元之罰鍰[664]。

7. 個資保護[665]

涉及個人資料外洩時，應另依個人資料保護法之規定辦理。

---

[656] 行政院，國家關鍵基礎設施安全防護指導綱領，2025 年 8 月 19 日，頁 1。
[657] 資通安全管理法第 7 條、第 20 條。
[658] 資通安全管理法第 12 條、第 23 條。
[659] 資通安全管理法第 10 條。
[660] 資通安全管理法第 11 條、第 27 條。
[661] 依資通安全管理法 114 年 9 月 24 日修正條文第 11 條說明三，危害國家資通安全產品包括受國家安全法及反滲透法所稱大陸地區、香港、澳門、境外敵對勢力或其所設立或實質控制之各類組織、機構、團體或其派遣之人等實質控制者所提供之產品。
[662] 資通安全管理法第 17 條、第 24 條。
[663] 資通安全管理法第 8 條、第 20 條。
[664] 資通安全管理法第 29 條、第 30 條。
[665] 資通安全管理法第 33 條。

（二） 《個人資料保護法》

生成式 AI 的運作核心在於「巨量資料的蒐集與訓練（輸入）」以及「內容生成（輸出）」，而這兩者可能與個人資料保護法（以下簡稱個資法）要求「特定目的」、「最小蒐集」及「可刪除」的原則存在衝突。此外，雖然人工智慧基本法第 14 條已規定，AI 研發及應用過程中，應避免不必要之個人資料蒐集、處理或利用，但並未從根本上改變臺灣個資法的規範架構，亦未針對生成式 AI 訂定特別規範，因此有關個人資料之規範仍應遵循個資法。

1. 特定目的

個資法第 5 條規定，個人資料之蒐集、處理或利用，不得逾越特定目的之必要範圍，並應與蒐集之目的具有正當合理之關聯；個資法第 19 條規定，非公務機關對個人資料之蒐集或處理，除特種資料外，應有特定目的。

以金融業為例，金融機構為提供金融服務而向客戶蒐集其個人資料，包含姓名、生日（年齡）、身分證字號、婚姻、家庭、教育、職業、聯絡方式、財務情況等，部分如人身保險業尚需病歷、醫療等特種資料。以提供金融服務為目的所蒐集之個人資料，其利用範圍是否涵蓋訓練金融機構的 AI 系統？或是金融機構基於業務使用需求，得否利用上述個人資料對 AI 進行 RAG？仍有疑義。

若認爲將上述個人資料用於 AI 訓練並非提供金融服務之目的範圍，則金融機構尚需符合個資法第 20 條之規定，始得將個人資料用於 AI 訓練之中。

（1） 同意權

除法律規定、緊急狀況或重大公益之情況外，無論是蒐集前或目的外的利用前，都需經當事人同意。而依個資法第 7 條規定，所謂同意係指當事人經蒐集者告知法定事項後所為允許之意思表示。

而民法上傳達意思表示之方式，又可分為明示及默示二種，臺灣個資法目前並未限制當事人同意之方式。此外，若蒐集者已告知個資法第 8 條第 1 項所列事項（如：機關名稱、蒐集目的、利用期間、利用方式等）而當事人未為反對，則推定當事人同意。

（2） 停止及刪除權

個資法第 11 條規定，當事人得請求更正或補充個人資料、請求停止處理或利用、請求刪除個人資料。

當個人資料輸入 AI 模型後，AI 並非如傳統資料庫將個人資料一筆一筆儲存，而是經過演算法分析後提取特徵、計算權重。若當事人依上開規定請求變更、停止利用或刪除個人資料，因個人資料已轉變為特徵與權重無法單獨刪除，則應如何落實停止及刪除權仍有待釐清。

（3） 國際傳輸

依個資法第 21 條反面解釋，臺灣原則上允許非公務機關進行個人資料的國際傳輸。目前多數 AI 模型多為國際科技公司研發，但部署模型的伺服器可能位於境外或臺灣境內，允許國際傳輸應有助於模型的運作與訓練，但仍應留意第 21 條的例外規定，如接受國對於個人資料之保護規範不足時，主管機關得限制其傳輸。

第二節　臺灣現行法制於金融實務之適用缺口

臺灣雖然已於 2025 年訂定人工智慧基本法，以及更早之前訂定營業秘密法、個人資料保護法等法律層級的規範，主管機關和同業公會也已針對委外作業、AI 運用等方面制定行政規則或自治規範，然而針對金融機構導入生成式 AI 方面，仍有一些潛在的問題無法透過現有的法制解決：

一、　各項原則欠缺量化的判斷標準

《人工智慧基本法》第 16 條第 1 項雖已確立以風險為基礎之管理方向，但目前數發部公布的 AI 風險分類框架並非針對金融領域所訂定，相關判斷標準亦非金融業可立即適用之標準，如：危害國家安全、侵犯人民基本權利、違背財產保障等；金管會雖在《金融業運用人工智慧（AI）指引》指引中列出六項風險評估考量因素[666]，並例示部分應用類型之風險（如信用評分、機器人理財屬較高風險），但仍欠缺具體、通案、可量化的標準，可能導致各金融機構的認定標準不一[667]，合規成本也將有所差異。

此外，《金融業運用人工智慧（AI）指引》中要求應重視公平性[668]，然公平屬於不確定法律概念，具體應如何判斷，仍有待釐清。

二、　生成式 AI 模型的黑盒問題（可解釋性問題）

《金融業運用人工智慧（AI）指引》[669]中，對於可解釋性的定義如下：「係指可清楚說明自行或委託研發並使用之 AI 系統如何運作及其預測或決策過程背後之邏輯，以利組織內評估是否符合內部政策、作業流程及監管要求等。」

針對 AI 模型的黑盒問題，指引中加上了「自行或委託研發」的條件。根據金管會的說明[670]，係考量 AI 系統如為委託其他業者研發或直接購入，金融機構可能因商業機密限制而無法完全得知 AI 系統運作細節，故將可解

---

[666] 金融監督管理委員會，註 4，頁 3-4。

[667] 實務上，許多企業會採用或參考國際標準組織（ISO）所制定的標準，如 ISO/IEC 42001，然此類標準並不具強制性。*See* ISO, *ISO/IEC 42001:2023*, ISO, https://www.iso.org/standard/42001 (last visited Jul. 5, 2026).

[668] 金融監督管理委員會，註 4，頁 10。

[669] 同前註。

[670] 同前註。

釋性限縮在金融機構可清楚說明「自行或委託研發並使用之 AI 系統」如何運作及其預測或決策過程背後之邏輯。然而，金融機構具體應說明到什麼程度仍不甚明確。

三、 個資利用問題

依臺灣《個資法》規範，雖然可能有利於模型的訓練與 AI 運用，但適用上仍有問題待解決，如停止及刪除權。若金融機構取得客戶同意後將客戶的個人資料用於 AI 模型的訓練，或於生成式 AI 推論時輸入客戶的個人資料，而客戶嗣後依個資法第 11 條規定要求變更、停止利用甚至刪除個人資料時，由於技術上難以將單一筆資料自模型中「抽取」出來，金融機構該如何妥善滿足個資法的法定要求，目前仍無明確解方。此外，由於修正後的個資法第 1 條之 1 尚未生效，在個資領域仍無統一的主管機關，金融機構運用生成式 AI 時所遭遇的個人資料問題，仍有待金管會依其權責訂定相關規範或指引。

四、 資安問題

AI 系統具有一些特殊的威脅樣態[671]，如：提示注入攻擊、資料投毒等，傳統的資安政策無法直接套用在生成式 AI 系統上，有些威脅甚至在實際部署前就可能出現，金融機構該如何預防與控管生成式 AI 的資安問題，目前國內仍無明確規範。

五、 外包管理與監理

主管機關針對外包早已定有委外作業相關規範，並透過「重大性」的判斷標準，畫定是否可以外包的紅線。然而，縱使法規面鬆綁、使金融機構可以對外採購生成式 AI 並將相關業務委外辦理，面對寡占的跨國科技公司，個別金融機構的談判能力、議價能力仍可能不足。金融機構如何在不違規、不犧牲客戶權益的前提下，有效的與科技公司進行 AI 業務的合作，考驗各金融機構自身的談判能力。金融總會雖已在金管會指導下成立金融科技產業聯盟，不過目前僅有對外公布將合作開發金融大語言模型，尚無與跨國科技公司進行合作與談判。

另外，《金融業運用人工智慧（AI）指引》中[672]提到：「金融機構如係自行或委託研發 AI 系統，宜規劃備置 AI 系統架構之相關文件或技術報告，並規劃提供監理機關存取及了解 AI 系統與所使用數據之權限，以便未來其查詢與瞭解金融機構 AI 系統之運作及使用數據之妥適性。」然而，承前所

[671] 資安人科技網，傳統資安框架無法防禦的 AI 特定攻擊，2026 年 1 月 2 日，https://www.informationsecurity.com.tw/article/article_detail.aspx?aid=12597（最後瀏覽日：2026 年 7 月 5 日）。

[672] 金融監督管理委員會，註 4，頁 22。

述，目前的生成式 AI 模型多由跨國的大型科技公司自主研發，金融機構如何落實「可解釋性」要求，仍不明確。

六、 對內問責機制集中於管理層

若行員根據生成式 AI 產生的結果對外執行業務，卻造成客戶權益受損時，現行的《人工智慧基本法》及《金融業運用人工智慧（AI）指引》雖然強調問責機制，各同業公會自律規範也規定仍需由金融機構就運用生成式 AI 的風險進行評估與管控，但具體應由金融機構對外部客戶負起全責，還是提供生成式 AI 模型的第三方科技公司也要一併負責？此部分尚需結合委外作業辦法一併觀察。

委外作業辦法第 4 條第 3 項規定，金融機構對於作業委外負最終責任；委外作業辦法第 7 條第 2 項規定，針對服務供應商之故意或過失造成的客戶權益受損，金融機構仍應對客戶依法負同一責任。換句話說，於現行法制體系下，對外仍應由金融機構承擔全責，金融機構不能主張對 AI 模型或科技公司已盡監督義務而免責，更不能將責任全部轉嫁給提供 AI 模型的科技公司。

對於金融機構內部而言，各同業公會自律規範規定需由金融機構就運用生成式 AI 的風險進行評估與管控，並依責任地圖制度之規定進行高階管理人員的定責、當責、問責與究責。金管會亦已於 2021 年修正各金融機構的內部控制及稽核制度實施辦法[673]，要求金融機構設置資訊安全長。2025 年底，金管會亦於金融資安韌性發展藍圖中明示「強化資安長職責與權責」的執行措施[674]，有望進一步釐清金融機構內部的職能與權責。

不過，責任地圖制度主要規範對象為高階管理人員，如董事長、總經理、副總經理及總行各部門主管等，至於中階管理層人員與非管理職員工之責任歸屬就較不明確。

七、 小結

綜上所述，國內金融業在導入生成式 AI 時，正面臨現有法制框架難以涵蓋的技術與資料風險。首先，各機構對於 AI 應用的風險分級缺乏具體、可量化的通案標準，且受限於生成式 AI 的黑盒特性，金融業難以徹底掌握第三方模型邏輯，對於向外界說明的「可解釋性」程度亦欠缺明確指引。此外，在資料保護方面，一旦客戶要求刪除已用於模型訓練的個資，現今技術極難進行精準抽取；同時，針對提示詞注入（Prompt Injection）與資料投毒等 AI 特有的新型態攻擊，現行資安規範仍存在監管空白，使機構在合規與防禦上面臨嚴峻挑戰。

---

[673] 金融控股公司及銀行業內部控制及稽核制度實施辦法第 38 條之 1、保險業內部控制及稽核制度實施辦法第 6 條之 1、證券暨期貨市場各服務事業建立內部控制制度處理準則第 36 條之 2

[674] 金融監督管理委員會，註 635，頁 11。

在供應鏈管理與內部問責層面，金融業亦面臨結構性的適法困境。由於生成式 AI 技術高度集中於少數跨國科技巨頭，個別金融機構在進行外包管理時，往往面臨議約談判能力不足的弱勢。而在機構內部，當 AI 應用發生錯誤或爭議時，現有問責機制多半僅能追溯至高階管理層，對於直接參與 AI 操作與開發人員的具體責任劃分卻相當模糊。整體而言，當前金融業的 AI 治理「原則」雖看似完整，但實務推動上的真正困難，在於落實的門檻、操作程序、證據認定與責任分工等配套機制仍付之闕如，成為阻礙產業規模化應用的核心問題。

第三節 臺灣與國際法制之比較

綜合美國、歐盟與新加坡對金融 AI 之監管途徑，可發現各國在面對生成式 AI 帶來的決策不透明、系統性風險與責任歸屬等難題時，採取了截然不同的治理邏輯，分述如下：

一、 治理模式差異

美國聯邦政府目前採取「分散式軟法」模式，未制定統一的 AI 法律，也沒有設立或指定專責機關，而是將 AI 行為納入既有的聯邦金融法規（如 ECOA 的反歧視要求）與各州的情境化立法中，強調「行為本質」而非「技術形式」。

歐盟則建構了嚴密的「分層式硬法架構」，以《AI Act》的風險分級為骨幹，搭配 GDPR、DORA 與 CCD II 等橫向法令，形成對高風險 AI 系統（包含自動化信用評估）最高密度的合規要求。

新加坡則屬第三條路徑，採取「集中式軟法」與實務工具並行的策略。新加坡政府透過 IMDA、PDPC 與 MAS 三個主管機關的密切合作，發布一系列一致性很高 AI 治理手冊與指引，引導金融機構建立必要的護欄，並針對生成式 AI 的技術特性，調整或增設相關的內部控制流程。

臺灣於 2026 年 1 月公布《人工智慧基本法》，正式將 AI 的風險治理入法，治理模式似乎與歐盟的硬法架構較為相近。然而，臺灣的《人工智慧基本法》並無具體罰則，內容亦屬原則性規範，與歐盟的 AI Act 有很大的差異。此外，由金管會與各同業公會所發布行政規則與自治規範又似乎與新加坡的軟法架構較為相近，但規範的密度與新加坡的《MindForge AI 風險管理營運手冊》仍有相當大的落差，亦缺少實際可用的檢測工具。

二、 對黑盒與可解釋性問題之應對差異

生成式 AI 的高度不透明性（黑盒）對金融法遵之「可解釋性」要求構成重大挑戰[675]。歐盟傾向透過 AI Act 與 GDPR，強硬要求系統提供者必須具備高度透明性，並賦予消費者獲取「有意義解釋」的權利。

相反地，新加坡在實務操作上展現出更高的妥協與彈性。鑑於多數金融機構並無能力揭開跨國科技公司大型語言模型的黑盒，新加坡並不強求金融機構必須做到技術上的完全透通，而是轉向強調「流程控制與護欄」，主張透過「補償性測試」、「RAG」限制資料源，以及在關鍵節點強制引入人在迴圈內機制，抵銷缺乏可解釋性所可能帶來的不利影響。MindForge 聯盟亦在最新的手冊中再次強調，生成式 AI 的驗證重點在於實際應用場景中模型的功能與輸出，與傳統 AI 專注於模型的透明性不同。透過前階段的驗證輸出與後階段的人機協作方式，避開難解的黑盒問題，也使得後續的責任追究變得較為容易[676]。新加坡政府更無需大幅度更動法規，只需利用金融機構現有的內部治理制度，為 AI 供應鏈中的個人分配特定的角色和責任，便可緩解責任歸屬方面的難題[677]。

臺灣於可解釋性方面，金管會於指引中將「可解釋性」限縮在金融機構可清楚說明「自行或委託研發並使用之 AI 系統」如何運作及其預測或決策過程背後之邏輯，但金融機構究竟該如何了解第三方科技公司的模型、應說明背後邏輯到什麼程度，目前仍無具體的標準或實例可供參考。另一方面，各同業公會的自律規範亦要求運用生成式 AI 之人員必須「了解生成式 AI 技術運作方式以及如何做出回應」，這雖然呼應了金管會的指引，但仍無具體、可落實的方式。臺灣如何兼顧可解釋性要求與現實層面的限制，仍有待釐清。

三、 第三方管理與對跨國科技公司的監理差異

---

[675] 針對大語言模型的可解釋性，科技業界（如 OpenAI、Google DeepMind、Anthropic 等）也仍在探索中。OpenAI 嘗試以大模型理解小模型中的參數與生成過程；Google DeepMind 則以類似神經科學家研究神經元的「破壞與追蹤（Causal Tracing）」方式，在模型中定位出專門負責處理某一語意的具體迴路（Circuit），並透過切斷該迴路讓 AI 喪失理解該句話的能力，藉此精確理解 AI 的思考步驟；Anthropic 則是將「稀疏自編碼器（Sparse Autoencoder, SAE）」外掛於大語言模型上，並將模型內無法判讀的大量參數透過「L1 正則化（L1 Regularization）」的步驟拆解成大量人類可理解的「特徵（Features）」，並一路追蹤到決策階段所形成的「迴路（Circuits）」。不過，針對上述業界的嘗試，學界仍有呼聲要求應開發一套從根本符合可解釋性的大語言模型，而非「事後」觀察模型並賦予其解釋。*See* OpenAI, *Language models can explain neurons in language Models*, OPENAI, https://openai.com/index/language-models-can-explain-neurons-in-language-models/ (last visited Jul. 5, 2026); Kevin Wang et al., *Interpretability in the Wild: A Circuit for Indirect Object Identification in GPT-2 Small*, 2022 ICLR 1 (2022); Dario Amodei, *The Urgency of Interpretability*, https://www.darioamodei.com/post/the-urgency-of-interpretability (last visited Jul. 5, 2026); Cynthia Rudin, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models Instead*, 1, no. 5 NAT. MACH. INTELL. 206-15 (2019).

[676] Alessio Azzutti, Wolf-Georg Ringe, and H. Siegfried Stiehl, *Machine Learning, Market Manipulation and Collusion on Capital Markets: Why the "Black Box" Matters*, 43 UNIV. PA. J. INT. LAW 128-29 (2021).

[677] *Id.* at 128-29.

由於研發生成式 AI 需要具備硬體上的強大算力，目前由臺灣金融機構自主訓練的生成式 AI 模型尚在研發中。因此目前金融機構仍必須先向第三方採購 AI 模型，如 Microsoft Azure 的生成式 AI 服務。當個別金融機構與這些跨國公司合作時，也可能會面臨談判能力不足的問題。此外，雖然銀行公會所發布的金融機構運用 AI 技術作業規範已規定：涉及營業執照業務或客戶資訊之 AI 作業委外，應依委外作業辦法辦理。然而臺灣金融機構向來習慣採取個別委託的方式進行查核，少有以聯合委託的方式進行[678]，無法根本性的解決金融機構談判能力不足的問題。

反觀 MAS 雖然也沒有大幅度的修正銀行業委外指引，但 MAS 透過推動 MindForge 計畫，號召眾多金融機構與跨國科技公司組成 MindForge 聯盟，共同制定風險框架與行業標準，間接提升新加坡金融機構監督委外科技公司的能力。此外，新加坡政府亦透過直接與跨國科技公司簽署 MOU 的方式[679]，在 AI 領域進行公私協力，為新加坡的私人企業樹立典範。

四、 可加強問責機制與職能劃分

《人工智慧基本法》強調針對高風險的 AI 應用，應明定責任歸屬；金融業運用人工智慧(AI)指引亦強調問責機制；各同業公會自律規範也規定需由金融機構就運用生成式 AI 的風險進行評估與管控。再結合委外作業辦法規定，臺灣金融機構在對外的問責上較無疑義。但對內的問責部分，目前資通安全法以及金管會均要求金融機構須設置資安長，而如何明確化資安長的職能與權責，仍是未來金管會執行的目標之一。另外，銀行公會亦參考外國立法例訂定針對高階主管的責任地圖自治規範，對高階主管的問責與究責提供基礎，但對非高階、非管理職人員，尤其是直接參與 AI 研發或部署的人員的責任則較不明確。

另一方面，雖然新加坡的個人問責與行為準則並非強制性的規範，但並未限於金控公司或銀行始能適用，任何金融機構均能參考該準則釐清其內部職位的職能與權責。該準則針對問責制明確劃分為「內部」與「外部」問責。其中在內部問責方面，準則將「新科技的開發與部署」納入「重大風險人員」的評估因素，並明確界定董事會與高階管理層（如風險長、資訊長/科技長、資安長等）在治理與風險管理上的具體職能與責任，使得 AI 風險真正與企業高層的當責機制掛勾。此外，對於非高階管理層但屬重大風險人員者，新加坡的個人問責與行為準則亦有適用，其問責與究責制度較臺灣現行規範更為全面。

---

[678] 余至浩，Google 雲端 GCP 服務應臺灣金融機構上雲需求，完成獨立第三方單位聯合查核，iThome，2023 年 3 月 28 日，https://www.ithome.com.tw/news/156170（最後瀏覽日：2026 年 7 月 5 日）。

[679] Aaron Tan, *Google Cloud and Singapore government team up on AI*, COMPUTERWEEKLY, https://www.computerweekly.com/news/252524096/Google-Cloud-and-Singapore-government-team-up-on-AI (last visited Jul. 5, 2026).

五、 委外作業之「重大性」標準一致

關於金融機構委外作業的議題由來已久，早在 2005 年巴塞爾銀行監理委員會（Basel Committee on Banking Supervision, BCBS）就已針對委外作業發布《金融服務委外》（Outsourcing in Financial Services）指引文件，並列出九項判斷重大性的標準[680]，許多司法管轄區均係參考 BCBS 的標準訂定委外作業的相關規範，臺灣與新加坡也不例外。唯一的細微差別在於，臺灣係將重大性的具體標準置於問答集，而新加坡係將之置於指引的附錄中，然於實際適用上應無區別。

六、 小結

各國對生成式 AI 採取了截然不同的治理邏輯，請參照下表 3、臺灣與國際法制之差異比較。

表 3、臺灣與國際法制之差異比較

項目	美國	歐盟	新加坡	臺灣
治理模式差異	1. 採取「分散式軟法」模式，未制定統一的 AI 法律。 2. 將 AI 行為納入既有聯邦金融法規與各州情境化立法中，強調「行為本質」而非「技術形式」。	1. 建構嚴密的「分層式硬法架構」。 2. 以人工智慧法的風險分級為骨幹，搭配 GDPR 等橫向法令，對高風險 AI 系統有最高密度的合規要求。	1. 採取「集中式軟法」與實務工具並行的策略。 2. 透過跨機關合作發布一致性高的指南，引導金融機構建立護欄與內部控制流程。	1. 2026 年公布人工智慧基本法，屬無具體罰則之原則性規範。 2. 治理模式表面似歐盟硬法，但金管會與公會規範實則較近新加坡軟法，且規範密度與實際可用檢測工具皆明顯不足。

[680] BCBS, *Outsourcing in Financial Services*, 14-15, https://www.bis.org/publ/joint12.htm (last visited Jul. 5, 2026).

對黑盒與可解釋性問題之應對		1. 傾向透過人工智慧法與GDPR，強硬要求系統提供者必須具備高度透明性。 2. 賦予消費者獲取「有意義解釋」的權利。	1. 展現實務操作上的妥協與彈性，不強求技術上的完全透通。 2. 強調「流程控制與護欄」，利用補償性測試、RAG及人在迴圈內機制抵銷不利影響。 3. 驗證重點在於實際應用場景中的輸出，避開難解的黑盒問題。	1. 金管會指引要求金融機構需清楚說明系統運作及決策背後邏輯。 2. 針對第三方模型應如何了解及說明程度，目前缺乏具體標準或實例可供參考，且公會自律規範亦無具體可落實方式。
第三方管理與對跨國科技公司的監理			1. 推動MindForge計畫組成聯盟，共同制定標準，間接提升金融機構監督委外科技公司的能力。 2. 政府直接與跨國科技公司簽署MOU進行公私協力。	1. 金融機構多需向第三方採購AI模型，面臨談判能力不足的問題。 2. 習慣採取個別委託查核（少有聯合委託），無法根本解決談判能力弱勢的問題。

問責機制與職能劃分			1. 個人問責與行為準則適用於任何金融機構，問責明確劃分為「內部」與「外部」。 2. 將新科技開發部署納入「重大風險人員」評估，並明確界定董事會與高階管理層在治理上的職能。 3. 制度較為全面，對非高階但屬重大風險人員亦有適用。	1. 對外問責依委外作業辦法較無疑義。 2. 對內問責方面，資安長職權仍待明確化。 3. 雖有高階主管責任地圖，但對非高階、直接參與AI研發部署人員的責任劃分仍較不明確。
委外作業之「重大性」標準			1. 參考BCBS（巴塞爾銀行監理委員會）標準訂定委外規範。 2. 將重大性的具體判斷標準置於指引的附錄中。	1. 同樣參考BCBS標準訂定。 2. 具體判斷標準置於問答集，實際適用上與新加坡無區別。

第七章　結論與建議

生成式 AI 進入金融機構之應用，並非單純的技術導入問題，而是金融服務模式、風險治理架構與監理法制同時受到重建的過程。本研究自生成式 AI 之技術特性與金融應用場景出發，進一步比較歐盟、美國及新加坡之監理制度，最後回到臺灣現行法制與金融實務。本研究認為金融業運用生成式 AI 的核心問題，並不只是模型是否準確，而是金融機構能否在資料使用、模型輸出、客戶互動、內部控制及第三方委外等環節中，建立一套足以被檢驗且能回應金融消費者保護要求的治理機制。

本研究首先指出，生成式 AI 已逐步由內部行政、文件摘要、客服輔助等相對低風險場景，擴展至詐欺偵測、授信分析、投資決策支援、理財機器人、合規監控及風險管理等高度敏感領域。相較於傳統 AI 偏重數值預測或規則比對，生成式 AI 具有語意理解、情境推論及內容生成能力，因此更容易進入金融機構原本仰賴專業判斷與說明義務的環節。將金融機構之生成式 AI 應用區分為內部行政與營運輔助型、對客戶互動與資訊提供型、決策輔助與評估支援型三類，指出目前金融機構多先由內部低風險場景開始導入，顯示實務上已具有一定之風險導向思維。惟一旦生成式 AI 進入對客戶互動或金融決策支援環節，其風險即不再只是內容錯誤，而可能進一步影響消費者對金融商品、授信條件、投資風險或服務選擇之理解與判斷。此類型影響雖未必構成完全自動化決策，但似足以改變消費者的決策方向，故不能僅以一般資訊服務等視之。

其次，本研究透過比較法觀察發現，歐盟、美國與新加坡對 AI 監管採取不同路徑。歐盟以風險分級與硬法規範為核心，透過 GDPR、DORA、CCD II 及 AI Act 建立較高密度之法制框架，特別重視高風險 AI、個人資料保護、自動化決策、人為監督及數位營運韌性。美國則主要採取部門法、行政指引、州法補充及既有金融監理工具並行之模式，強調創新發展與功能性監理，並由公平授信、消費者保護、模型風險管理及反歧視規範切入。新加坡則介於兩者之間，並非急於制定單一硬法，而是透過 FEAT 原則、Veritas、Project Moonshot、生成式 AI 護欄手冊、科技風險管理及委外指引等工具，將原則轉化為可測試、可操作及可由金融機構落實的治理流程。此一比較顯示，AI 監理並不存在單一標準答案；真正關鍵在於監理制度能否依照金融業務風險、技術特性與機構責任，形成清楚而可執行的治理要求。

就臺灣而言，本研究認為我國已具備若干初步制度基礎，包括《人工智慧基本法》、《人工智慧風險分類框架》、《金融業運用人工智慧（AI）指引》、《個人資料保護法》、《營業秘密法》、《金融消費者保護法》、《金融業委外規範》及資通安全等相關要求。然而，這些規範目前仍呈現原則性、分散性與高度抽象化之特徵。歸納其主要缺漏，包括：風險分級欠缺具體

量化標準、公平性與可解釋性尚未形成可操作判準、生成式 AI 黑盒問題仍未被制度化處理、個資利用與跨境傳輸之合規界線不夠清楚、資安與深偽風險尚待具體化控管、第三方模型與雲端服務供應商之監理責任仍有不足，以及內部問責機制過度集中於管理階層而未能細緻區分董事會、高階經理人、模型開發單位、法遵、稽核、資訊安全及業務部門之責任。尤其在金融業高度依賴外部模型、雲端平台及跨國科技公司之情況下，若金融機構對模型邏輯、資料來源、輸出限制及異常處理缺乏實質掌握，即可能出現形式上由金融機構負責，實質上卻無法說明或控制的問責落差。

基於上述研究成果，臺灣未來不宜僅以鼓勵創新或全面管制作為二分法選擇，應建立以風險分級、場景分類、責任分工及可稽核治理為核心的金融生成式 AI 監理框架。具體而言，主管機關可先就金融機構常見應用場景建立分級標準，將內部行政輔助、對客戶資訊提供、投資或授信決策輔助、自動化交易或風控執行等場景分別配置不同程度之審查、揭露、測試與人為覆核要求。對於高風險應用，應要求金融機構建立模型登錄、資料治理、偏誤檢測、壓力測試、輸出軌跡、異常回報、人為介入、客戶申訴與事後稽核機制。關於對客戶互動型應用，則應強化揭露義務，使消費者知悉其互動對象、內容限制及是否構成投資建議或商品推薦，避免金融機構以僅供參考而概括排除其應負之專業注意義務。

若再進一步地針對前述金融業應用 AI 的三大類別場景，分別提出相應的管理及監理措施。首先，在內部行政與營運輔助型的應用上，現早已是高度採用生成式 AI，而此情形因為技術複雜度較低，且對於金融決策與個人權益之影響程度相對有限。企業就此是可以自由開發與使用，但若企業內部有更高的遵法標準，則仍是可以採行自願性行為準則與最佳實務，促成負責任 AI 的使用。再者，在對客互動與資訊提供型的應用上，如今已是中高度採用生成式 AI，而因生成內容已直接進入金融機構與消費者之互動關係中，企業就此則應遵循透明度之規範要求，履行揭露標示義務。也就是說，要能確保使用者知悉其正在與 AI 互動，或所接觸之內容係由 AI 所生成，從而避免混淆與欺瞞。最後，在決策輔助與評估支援型的應用上，則是因為已經實質上影響到決策者可得資訊的範圍、呈現方式與注意焦點，而具有高度的風險。故而企業應遵守一套強制性合規義務，包括風險管理、資料治理、技術文件保存、人類監督機制、透明性要求，以及合格評定程序。

本研究建議未來可深化的四個方向：第一，應逐步建立臺灣金融業生成式 AI 風險分級模型，將業務類型、資料敏感性、對客程度、決策影響力、模型自主性及第三方依賴程度納入評估。第二，應針對可解釋性建立具體法制標準，區分技術性解釋、業務性解釋及消費者可理解之理由說明，避免將可解釋性停留在抽象原則。第三，應研究金融機構使用境外大型模型或雲端服務時，委外契約中是否應明定資料使用限制、模型更新通知、稽

核權、事故通報、退出機制及監理機關查核協力義務。第四，應從金融消費者保護與民事責任角度，檢討當生成式 AI 輸出錯誤、偏誤或造成誤導時，金融機構、模型供應商與使用者之責任分配，並思考舉證責任、因果關係及損害認定之調整可能。

總言之，生成式 AI 對金融業的價值，在於提升效率、強化風險辨識、改善客戶服務並降低部分作業成本。但其制度風險，也正存在於高度擬真的語言能力、難以完整解釋的模型結構、資料依賴及第三方技術集中。金融監理的任務不是阻止金融機構使用生成式 AI，而是確保金融機構在使用時仍能說明其資料來源、控制其輸出風險、維持人為判斷、保障消費者權益，並在發生錯誤時負起可被追究的責任。未來的制度發展，可由現行原則型指引逐步走向具體化、應用場景化與可查核化的監理架構，使生成式 AI 真正成為金融創新的工具，而非責任模糊與風險外部化的來源。

參考文獻

中文文獻

一、書籍

陳靜慧，歐洲人權法院及歐洲法院對於間接歧視概念之適用與實踐，憲法解釋之理論與實務，第九輯，2017 年 4 月。

二、期刊文章

1. 文逢遠，美國人工智慧行政命令之新舊比較與可能影響，經貿法訊第 343 期，政大商學院國際經貿組織暨法律研究中心，2025 年。
2. 李宜熹、王友珊，生成式 AI 在金融領域的技術演進與挑戰，期貨人，第 93 卷，2025 年。
3. 林勤富，歐盟《人工智慧法》之制度設計、 規範內涵與治理侷限，中研院法學期刊，第 36 期，2025 年。
4. 陳奕璇，現階段歐盟與美國對人工智慧規範之比較分析，安全與情報研究，第九卷，第 1 期，2026 年。
5. 陳顯武、連雋偉，從『歐盟憲法』至『里斯本條約』的歐盟人權保障初探──以『歐盟基本權利憲章』為重點，臺灣國際研究季刊，第 4 卷，第 1 期，2025 年。
6. 劉博涵，智慧投顧演算法透明的價值辯證、規制導向與實現路徑，深圳大學學報淵人文社會科學版，第 41 卷，第 1 期，2024 年。
7. 劉輝，大資料金融演算法的法律規制，財經理論與實踐，第 42 卷，第 2 期，2021 年。

三、其他文獻

1. IBM，IBM 諮詢助力台北富邦銀行推出 Fubon+，IBM Taiwan News room：https://taiwan.newsroom.ibm.com/2025-04-15-IBM-Fubon（最後瀏覽日：2026 年 7 月 5 日）。
2. 中國信託，從「AI 智能防線」到「AI 行員」中信打造全台第一個「AI 防詐金融生態系」，商業周刊：https://www.businessweekly.com.tw/business/indep/1006271（最後瀏覽日：2026 年 7 月 5 日）。
3. 王宏仁，永豐金控揭露多 AI 代理戰略布局，關鍵是 Agent 宇宙藍圖和五大基礎建設，iThome：https://www.ithome.com.tw/news/173012（最後瀏覽日：2026 年 7 月 5 日）。
4. 台新國際商業銀行，台新銀行奪 2025 《數位銀行家》五項大獎 以 AI 與數據核心引領金融創新體驗，台新國際商業銀行：https://www.taishinbank.com.tw/TSB/personal/common/news/TSBankNews-006851/（最後瀏覽日：2026 年 7 月 5 日）。

5. 玉山銀行，玉山與 Google Cloud 合作推出 AI 金融顧問 引領金融數位轉型，玉山銀行：https://www.esunbank.com/zh-tw/about/news-center/detail?id=114999119E9E4FABA44B89DE8F87A60F（最後瀏覽日：2026 年 7 月 5 日）。
6. 生成式 AI 產生「幻覺」問題 警惕虛假資訊風險，今周刊：https://www.businesstoday.com.tw/article/category/183015/post/202601200045/（最後瀏覽日：2026 年 7 月 5 日）。
7. 永豐銀行，MongoDB Atlas 與生成式 AI 完美結合 永豐銀行數位金融服務再進化，iThome：https://www.ithome.com.tw/pr/168603（最後瀏覽日：2026 年 7 月 5 日）。
8. 余至浩，Google 雲端 GCP 服務應臺灣金融機構上雲需求，完成獨立第三方單位聯合查核，iThome，2023 年 3 月 28 日，https://www.ithome.com.tw/news/156170（最後瀏覽日：2026 年 7 月 5 日）。
9. 呂欣芷，富邦金秀 AI 創新力 銀行全新 APP「Fubon＋」亮相，工商時報：https://tw.stock.yahoo.com/news/%E5%AF%8C%E9%82%A6%E9%87%91%E7%A7%80ai%E5%89%B5%E6%96%B0%E5%8A%9B-%E9%8A%80%E8%A1%8C%E5%85%A8%E6%96%B0app-fubon-%E4%BA%AE%E7%9B%B8-052614445.html（最後瀏覽日：2026 年 7 月 5 日）。
10. 李昀璇，Fintech 周報第 233 期：摩根士丹利揭露生成式 AI 發展策略，要以 AI 助理促進顧問、顧客和平臺三方無縫互動，iThome: https://www.ithome.com.tw/news/161710（最後瀏覽日：2026 年 7 月 5 日）。
11. 李昀璇，玉山怎麼用生成式 AI ？12 項應用服務大解析，iThome：https://www.ithome.com.tw/news/162551（最後瀏覽日：2026 年 7 月 5 日）。
12. 李昀璇，台新金控首度公開導入生成式 AI 的智能客服，提供理財投資建議和信用卡建議，iThome：https://www.ithome.com.tw/news/160151（最後瀏覽日：2026 年 7 月 5 日）。
13. 李昀璇，第一銀行公開最新生成式 AI 行動方案 GALA 計畫，優先賦能員工素養，再聚焦兩大發展重點，iThome：https://www.ithome.com.tw/news/169052（最後瀏覽日：2026 年 7 月 5 日）。
14. 李昀璇，高層每月緊追生成式 AI 績效，華南銀揭生成式 AI 兩大應用模式，一年創造 50 多名數位勞動力，iThome：https://www.ithome.com.tw/news/169617（最後瀏覽日：2026 年 7 月 5 日）。
15. 李昀璇，【LLM 模型如何落地金融業？技術與非技術難題並存】凱基人壽導入生成式 AI 的三大挑戰，iThome：https://www.ithome.com.tw/news/162052（最後瀏覽日：2026 年 7 月 5 日）。

16. 范姜真媺，AI 系統運用與基本人權之衝突及治理，司法院大法官 111 年度學術研討會-「數位時代網路與憲法基本權保障」，憲法法庭：https://cons.judicial.gov.tw/docdata.aspx?fid=5364&id=345738（ 最 後瀏覽日：2026 年 7 月 5 日）。
17. 金融監督管理委員會，金融資安韌性發展藍圖，頁 16-17，金融監督管理委員會，2025 年 12 月，https://www.fsc.gov.tw/uploaddowndoc?file=news/202601021913000.pdf&filedisplay=%E9%87%91%E8%9E%8D%E8%B3%87%E5%AE%89%E9%9F%8C%E6%80%A7%E7%99%BC%E5%B1%95%E8%97%8D%E5%9C%96.pdf&flag=doc（ 最後瀏覽日：2026 年 7 月 5 日）。
18. 金融監督管理委員會新聞稿，金管會公布金融業應用人工智慧(AI)調查結果，2025 年 5 月 20 日，https://www.fsc.gov.tw/ch/home.jsp?id=96&parentpath=0,2&mcustomize=news_view.jsp&dataserno=202505200001&dtable=News（最後瀏覽日：2026 年 7 月 5 日）。
19. 南山人壽，用 AI 開外掛！一窺南山人壽法遵 AI 智能助理的架構工法，數位時代：https://www.bnext.com.tw/article/80404/nanshanlife202409（最後瀏覽日：2026 年 7 月 5 日）。
20. 唐可欣，數位理財新升級 富邦金子公司推新 APP 與線上投保服務：https://www.storm.mg/article/11099961（最後瀏覽日：2026 年 7 月 5 日）。
21. 孫彬訓，銀行導入生成式 AI 內用先行，工商時報：https://www.ctee.com.tw/news/20241218700140-439901（最後瀏覽日：2026 年 7 月 5 日）。
22. 財管專案小組，元大銀行 AI 技術導入智能客服，提供精準服務，元大銀行：https://www.wealth.com.tw/articles/d9ccb8e8-3505-454f-9110-31c752f69c05（最後瀏覽日：2026 年 7 月 5 日）。
23. 國泰金控，AI 代理時代已至！國泰金控以 GAIA 2.0 框架加速 AI 應用百花齊放，數位時代：https://www.bnext.com.tw/article/85468/cathay1209（最後瀏覽日：2026 年 7 月 5 日）。
24. 第一銀行，從競賽激發創新力，第一銀行如何以 AI 打造差異化競爭優勢、型塑數位生活圈？，數位時代：https://www.bnext.com.tw/article/83952/firstbank202507（最後瀏覽日：2026 年 7 月 5 日）。
25. 許向華，美國人工智慧權利法案藍圖，科技發展觀測平台，2022 年 https://outlook.stpi.niar.org.tw/index/tdop/4b11410082b117250183a5d5bfd020e5（最後瀏覽日：2026 年 7 月 5 日）。
26. 許正乾、陳家駿，人工智慧最新應用及其法律議題系列 1 -- 從生成式 AI 到 AI 代理，科技產業資訊室：https://iknow.stpi.niar.org.tw/Post/Read.aspx?PostID=22206（最後瀏覽日：2026 年 7 月 5 日）。

27. 許正乾、陳家駿，人工智慧最新應用及其法律議題系列 3──AI 代理 v.代理 AI，科技產業資訊室：https://iknow.stpi.niar.org.tw/post/Read.aspx?PostID=22296 (最後瀏覽日：2026 年 7 月 5 日)。
28. 許慧瑩，AI 偏差歧見的另類難題–使用者對 AI 風險評估工具的差異性互動，臺灣人工智慧行動網：https://ai.iias.sinica.edu.tw/how-risk-assessment-tools-influence-final-decision-exp/（最後瀏覽日：2026 年 7 月 5 日）。
29. 許靜之，CUBE 成本增 50%，收益成長 8 倍！國泰首發布「GAIA」生成式 AI 框架，曝數位轉型成果：https://www.bnext.com.tw/article/80649/cathay-holdings-ai-frameworks-gaia（最後瀏覽日：2026 年 7 月 5 日）。
30. 張浩祥，生成式 AI 原理及在金融領域中之應用與突破，頁 9，中央銀行資訊處：https://www.tpefx.com.tw/uploads/download/tw/0721-11.pdf（最後瀏覽日：2026 年 7 月 5 日）。
31. 陳以謙，台新孫一仕：AI 進入民主化時代，臺灣銀行家：https://taiwanbanker.tabf.org.tw/paperDetail?id=4516（最後瀏覽日：2026 年 7 月 5 日）。
32. 陳怡瑄，決策式 AI 引領投資新速度 群益金鼎證券攜手精誠資訊「AI 問問」一秒看懂個股重點，卓越雜誌：https://ecf.com.tw/tw/article/show.aspx?num=8813（最後瀏覽日：2026 年 7 月 5 日）。
33. 陳曉莉，OpenAI 把 AI 程度分為 5 級，最高等級將可處理整個組織的事務，iThome: https://www.ithome.com.tw/news/163913（最後瀏覽日：2026 年 7 月 5 日）。
34. 曾憲立，公部門人工智慧規範架構結案報告，數位治理研究中心：https://www.teg.org.tw/researches/69（最後瀏覽日：2026 年 7 月 5 日）。
35. 凱基人壽保險股份有限公司，壽險業第一家！凱基人壽導入生成式 AI 打造頂尖業務部隊：https://www.kgilife.com.tw/zh-tw/-/media/files/kgil/homepage/new-folder1/2024/20240201-ai.pdf（最後瀏覽日：2026 年 7 月 5 日）。
36. 富邦金融控股股份有限公司，生成式 AI 改變職場！富邦金首度舉辦黑客松，2 個月將素人培訓為 AI 達人，中央社：http://www.cna.com.tw/postwrite/chi/390486（最後瀏覽日：2026 年 7 月 5 日）。
37. 富邦金融控股股份有限公司，富邦金控舉辦首屆生成式 AI 黑客松，富邦金控：https://www.fubon.com/financialholdings/home/index.html（最後瀏覽日：2026 年 7 月 5 日）。
38. 黃于庭，國泰金控技術年會揭露生成式 AI 框架 GAIA 2.0 宣告進入雲端優先階段，經濟日報：https://money.udn.com/money/story/5613/9082615（最後瀏覽日：2026 年 7 月 5 日）。

39. 黃登榆，台中銀行與群聯電子跨界聯手 打造 AI 金融服務新平台，經濟日報：https://money.udn.com/money/story/5613/9043861（最後瀏覽日：2026 年 7 月 5 日）。
40. 群聯電子，Phison｜aiDAPTIV+：https://phisonaidaptiv.com/zh-tw/#（最後瀏覽日：2026 年 7 月 5 日）。
41. 葉佳華，公股銀拚 AI 運用多頭並進 今年聚焦兩大應用發展，經濟日報：https://money.udn.com/money/story/5613/9240342（最後瀏覽日：2026 年 7 月 5 日）。
42. 資安人科技網，傳統資安框架無法防禦的 AI 特定攻擊，2026 年 1 月 2 日，https://www.informationsecurity.com.tw/article/article_detail.aspx?aid=12597（最後瀏覽日：2026 年 7 月 5 日）。
43. 臺灣中小企業銀行，2024 台北金融科技展 臺灣企銀邀民眾探索「Hokii 星球」：https://www.tbb.com.tw/zh-tw/about/information/announcement/news/press-announcements/20241101（最後瀏覽日：2026 年 7 月 5 日）。
44. 廖珮君和任珮云，金融業打團體戰自建 AI 模型 金管會主委彭金隆看好三效益，經濟日報：https://money.udn.com/money/story/5613/9458638（最後瀏覽日：2026 年 7 月 5 日）。
45. 趙敏雅，數發部：AI 風險分類框架草案 拚 115 年首季提報政院，中央社，2025 年 12 月 30 日，https://www.cna.com.tw/news/afe/202512300212.aspx（最後瀏覽日：2026 年 7 月 5 日）。
46. 數位發展部，AI 風險分類框架，數發部 AI 專區 - 重點政策，2026 年 3 月 23 日，https://moda.gov.tw/major-policies/ai/governance/19244（最後瀏覽日：2026 年 7 月 5 日）。
47. 數位發展部，AI 風險分類框架，數位發展部，頁 1，2026 年 7 月 7 日，https://www-api.moda.gov.tw/File/Get/moda/zh-tw/UieEfux5Yv0iJuR（最後瀏覽日：2026 年 9 月 2 日）。
48. 劉汶渝，人工智慧風險管理框架(AI RMF)，臺灣人工智慧行動網，2024 年 https://ai.iias.sinica.edu.tw/ai-risk-management-framework（最後瀏覽日：2026 年 7 月 5 日）。
49. 鄭芝珊，業界首家！永豐 iWish 創新應用 GAI ，直覺對話讓臨櫃存提轉匯變得更快速，經濟日報：https://money.udn.com/money/story/5635/8996799（最後瀏覽日：2026 年 7 月 5 日）。
50. 鄭妤安，花旗 Citi AI 工具正式在台推出 打造 AI 智慧團隊，工商時報： https://www.ctee.com.tw/news/20251111701333-430304（最後瀏覽日：2026 年 7 月 5 日）。

51. 戴玉翔，保險業首家！富邦產險「FacePass 極速保」開創 AI 刷臉投保服務，經濟日報：https://money.udn.com/money/story/5613/9007939（最後瀏覽日：2026 年 7 月 5 日）。
52. 藍鈞達，富邦金控積極擁抱 AI 組成「生成式 AI 應用推動團隊」，經濟日報：https://money.udn.com/money/story/5613/9125014（最後瀏覽日：2026 年 7 月 5 日）。

英文文獻

一、 書籍

1. Oxford English Dictionary, *Software Used to Perform Tasks or Produce Output Previously Thought to Require Human Intelligence*.

二、 書籍專章

1. Brown, Tom B. et al., *Language Models Are Few-Shot Learners*, *in* 33 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 1877 (Hugo Larochelle et al. eds., 2020).
2. Dosovitskiy, Alexey et al., *An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale*, in INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR 2021).
3. Goodfellow, Ian J. et al., *Generative Adversarial Nets*, *in* 27 ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 2672 (Z. Ghahramani et al. eds., 2014).
4. Lewis, Patrick et al., *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*, *in* ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 33, 9459 (2020).
5. Radford, Alec et al., *Learning Transferable Visual Models from Natural Language Supervision*, in PROCEEDINGS OF THE 38TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML 2021).
6. Vaswani, Ashish et al., *Attention Is All You Need*, *in* ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 30, 5998 (2017).

三、 期刊文章

1. Aaron, J. B., Yuanyuan, Z. & Ana, *AI and culture: Culturally dependent responses to AI systems.* CURRENT OPINION IN PSYCHOLOGY, 58(4):101838.
2. Arner, Douglas W. et al, *FinTech, RegTech, and the Reconceptualization of Financial Regulation,* 37 NW. J. INT'L L. & BUS. 371 (2017).

3. Azzutti, Alessio, Wolf-Georg Ringe & H. Siegfried Stiehl, *Machine Learning, Market Manipulation and Collusion on Capital Markets: Why the "Black Box" Matters*, 43 UNIV. PA. J. INT. LAW 128 (2021).
4. Banerjee, Abhijit V., *A Simple Model of Herd Behavior*, 107 Q. J. ECON. 797 (1992).
5. Bayamlıoğlu, Emre, *Intelligent Agents and Their Legal Status*, 1 ANKARA BAR REV. 46 (2008).
6. Buckley, Ross P. et al., *Regulating Artificial Intelligence in Finance: Putting the Human in the Loop*, 43 SYDNEY L. REV. (2021).
7. Danielsson, Jon et al., *Artificial Intelligence and Systemic Risk*, 140 J. BANKING & FIN. 106290 (2022).
8. Dell'Acqua, *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, Sept. 15, 2023).
9. Frostestad, Heidi L., *AI Regulation in a ChatGPT Era: Cross-border Cooperation and Hope in a Sudden Storm*, 32 IND. J. GLOB. LEGAL STUD. 1 (2024).
10. Gensler, Gary & Lily Bailey, *Deep Learning and Financial Stability*, SSRN SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, Nov. 13, 2020).
11. Keane, Alice E., Carlos Ferran & Robert Garcia, *AI and the European Union Artificial Intelligence Act: An Overview*, 35 MIDWEST L.J.1 (2025).
12. Kearns, Michael, *AI Reverberations Across Finance*, FIN. & DEV. (2023).
13. Király, Miklós, *Survival of the Fittest? AI and Its Different Regulatory Models*, 39 EMORY INT'L L. REV. 473 (2025).
14. Mundlamuri, Rahul et al., *The Evolution of AI: From Classical Machine Learning to Modern Large Language Models*, 13 IEEE ACCESS 178302-7 (2025).
15. Rezaeikhonakdar, *Artificial Intelligence in Healthcare Governance Compliance and Data Privacy Beyond Hipaa,* 29 QUINNIPIAC HEALTH L.J. 175 (2026).
16. Rudin, Cynthia, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models Instead*, 1, no. 5 NAT. MACH. INTELL. 206 (2019).
17. Vives, Xavier, *The Impact of FinTech on Banking*, 2 EUR. ECON. 97 (2017).

18. Wang, Kevin et al., *Interpretability in the Wild: A Circuit for Indirect Object Identification in GPT-2 Small*, 2022 ICLR 1 (2022).
19. Weber, Philipp, Karl V. Carl & Oliver Hinz, *Applications of Explainable Artificial Intelligence in Finance—A Systematic Review of Finance*, INFORMATION SYSTEMS, AND COMPUTER SCIENCE LITERATURE, MGMT. REV. Q. (2023).
20. Witteler, Michael & Lucas Moll, *Künstliche Intelligenz am Arbeitsplatz – Datenschutz und Rechte des Betriebsrats*, 2023 NZA 327.

四、 其他文獻

1. AI Verify Foundation & Infocomm Media Development Authority, *Model Al Governance Framework for Generative Al*.
2. AI Verify Foundation, *Aiverify-foundation/Moonshot*, GitHub.
3. Amodei, Dario, *The Urgency of Interpretability*.
4. AWS, *Automate User On-Boarding for Financial Services with a Digital Assistant Powered by Amazon Bedrock*, AWS MACHINE LEARNING BLOG (2023).
5. Bautzer, Tatiana, *Citigroup rolls out artificial intelligence tools for employees in eight Countries*, REUTERS.
6. BCBS, *Outsourcing in Financial Services*.
7. Best, Eric, *Artificial Intelligence and You*, WELLS FARGO STORIES.
8. Bloomberg, *Introducing BloombergGPT, Bloomberg's 50-billion parameter large language model, purpose-built from scratch for Finance*, Bloomberg L.P., 30. March. 2023.
9. Broeders, Dirk & Jermy Prenio, *Innovative Technology in Financial Supervision (Suptech): The Experience of Early Users*, FSB, BANK FOR INT'L SETTLEMENTS (2018).
10. CSBS, Press Release, *CSBS Establishes Artificial Intelligence Advisory Group* (Aug. 22, 2024).
11. DLA Piper, *DORA Penalty Regimes: Overview of Divergence Among Member States*.
12. EUROFOUND, *Italy: Privacy Authority Fines Foodinho 5 Million Euro for Rider Data Breaches*, PLATFORM ECONOMY DB. (July 16, 2021).
13. European Commission, *Artificial Intelligence for Europe*.
14. EWSolutions, *Understanding Black Box AI*.
15. Fadilpašić, Sead, *This "ZombieAgent" Zero-Click Vulnerability Allows for Silent Account Takeover – Here's What We Know*, TECHRADAR (2026).

16. Financial Conduct Authority, *Generating and Using Synthetic Data for Models in Financial Services: Governance Considerations* (2025).
17. FSB, *The Financial Stability Implications of Artificial Intelligence*.
18. FSB, *FinTech and market structure in financial services: Market developments and potential financial stability implications*.
19. FSB, *The Use of Supervisory and Regulatory Technology by Authorities and Regulated Institutions: Market Developments and Financial Stability Implications*.
20. Giovine, Carlo, Larry Lerner, *One Year In: Lessons Learned in Scaling Up Generative AI for Financial Services*, McKinsey & Co. (2024).
21. Goodfellow, Ian J., Jonathon Shlens & Christian Szegedy, *Explaining and Harnessing Adversarial Examples*, arXiv:1412.6572 (2015).
22. Gritsenko, Dima, *Generative AI in Insurance: The Future of Claims – Proactive Problem-Solving and Fraud Detection at Your Fingertips*, master of code (2025).
23. Hu, Edward J. et al., *LoRA: Low-Rank Adaptation of Large Language Models*, arXiv:2106.09685 (2021).
24. Hugging Face, *Illustrating Reinforcement Learning from Human Feedback (RLHF)* (2022).
25. Infocomm Media Development Authority & Personal Data Protection Commission, *Model AI Governance Framework Second Edition*.
26. Infocomm Media Development Authority, *Singapore launches Project Moonshot*, Infocomm Media Development Authority Press Releases, Factsheets and Speeches.
27. International Monetary Fund, *Generative Artificial Intelligence in Finance: Risk Considerations*, FinTech Note No. 2023/006, 2023.
28. ISO/IEC, ISO/IEC 42001:2023.
29. Izacard, Gautier et al., *Unifying Large Language Models and Retrieval for Knowledge-Intensive Tasks*, arXiv:2206.05235 (2022).
30. Jordan, J. et al., *Synthetic Data – What, Why, and How?,* arXiv:2205.03257 (2022).
31. Kerrins, Tracy and Rohit Bhat, *Wells Fargo brings the agentic era to financial services with Google Cloud AI*, Google Cloud.
32. Kremer, Andreas et al., *Embracing Generative AI in Credit Risk*, McKinsey & Co. (Jul. 1, 2024).
33. Longpre, Shayne et al., *BloombergGPT: A Large Language Model for Finance*, arXiv:2303.17564 (2023).

34. Madiega, Tambiama, *Artificial Intelligence Act.* European Parliamentary Research Service, Sep. 2024.
35. Marshall, Matt, *Wells Fargo's AI assistant just crossed 245 million intera ctions – no human handoffs, no sensitive data Exposed*, VENTUREBEAT.
36. Maslej, Nestor et al., *Artificial Intelligence Index Report 2024*, 163, Stanf ord, CA: Stanford University, Apr. 2024.
37. Mastercard, *Mastercard supercharges consumer protection with gen AI*, Mastercard, 1. February. 2024.
38. Mastercard, *Mastercard Accelerates Card Fraud Detection with Generati ve AI Technology*.
39. MATPROOF, *EU AI Act Fines and Penalties: What Non-Compliance Wi ll Cost You*.
40. McKinsey & Co., *The Economic Potential of Generative AI: The Next Pr oductivity Frontier*.
41. McKinsey & Co., *JPMorgan Chase's Derek Waldron on AI and banking*, MCKINSEY & CO. (2025).
42. Morgan Stanley, *Launch of AI @ Morgan Stanley Debrief*, Morgan Stanl ey, 26. June. 2024.
43. NAT'L INST. OF STANDARDS & TECH., NIST AI 600-1, *Artificial Inte lligence Risk Management Framework: Generative Artificial Intelligence Profile* 1 (2024).
44. Nishtha, *Generative AI for Finance*, PROJECTPRO (2024).
45. NORTAL, *2026 EU Financial Services Compliance*.
46. OECD & FSB, *OECD–FSB Roundtable on Artificial Intelligence (AI) in Finance* (2024/9/30).
47. OECD, *Recommendation of the Council on Artificial Intelligence*.
48. OpenAI, *Language models can explain neurons in language Models*, OPE NAI.
49. Orbiq, *DORA Compliance: Complete Guide*.
50. RASA, *EU AI Act & Data Sovereignty: A Guide for Financial Institution s*.
51. Ray, Tiernan, *OpenAI's Gigantic GPT-3 Hints at the Limits of Language Models for AI*, ZDNET.
52. Raza, A. Ahmed, *How JPMorgan Uses AI to Save 360,000 Legal Hours a Year*, MEDIUM (2025).
53. Reuters, *Goldman Sachs launches AI assistant firmwide*, REUTERS.
54. Saha, Bikash et al., *Generative AI in Financial Institution: A Global Surv ey of Opportunities, Threats, and Regulation*, ARXIV (2026).

55. Son, Hugh, *Morgan Stanley kicks off generative AI era on Wall Street wit h assistant for financial Advisors*, CNBC (2023).
56. Son, Hugh, *Morgan Stanley wealth advisors are about to get an OpenAI-powered assistant to do their grunt Work*, CNBC (2023).
57. Suresh, H. & John V. Guttag, *A Framework for Understanding Unintend ed Consequences of Machine Learning,* ARXIV:1901.10002 (2019).
58. Tan, Aaron, *Google Cloud and Singapore government team up on AI*, CO MPUTERWEEKLY.
59. The Association of Banks in Singapore, *Handbook on Generative AI Gua rdrails in Banking*.
60. The Digital Banker, *Citi rolls out AI tools to employees in Hong Kong*, T he Digital Banker.
61. The MindForge Consortium, *Emerging Risks and Opportunities of Gener ative AI for Banks*.
62. The MindForge Consortium, *AI Risk Management: Executive Handbook*.
63. The MindForge Consortium, *MindForge AI Risk Management Implement ation Examples*.
64. The MindForge, *MindForge AI Risk Management Operationalisation Ha ndbook*.
65. Udell, Ben, *We're Speeding Past the Tipping Point of Gen AI in Banking*, THE FINANCIAL BRAND.
66. Vanguard, *Vanguard Unveils Generative AI Client Summaries for Financ ial Advisors*.
67. Veritas Consortium, Veritas Documents 1 to 6 (including Assessment Me thodologies and Case Studies).
68. Wikipedia, *AI Washing*.
69. Wilkinson, Lindsey, *JPMorgan Chase to equip 140K workers with gener ative AI Tool*, CIO DIVE, (2024).