生成式 AI 原理
及在金融領域中之應用與突破

資訊處

張浩祥

中華民國 114 年 6 月

作者任職於中央銀行資訊處，本文內容純屬個人意見，與服務單位無關，如有錯誤概由作者負責。

摘　要

生成式人工智慧（又稱生成式 AI，Generative AI）是一種能根據資料分佈生成新樣本的技術，涵蓋文本、圖像、音訊等多模態應用。本文介紹其核心原理與模型類型，包括自回歸模型（Autoregressive Model）、變分自編碼器（Variational Autoencoder）、生成對抗網路（Generative Adversarial Network）與 Transformer 架構，並延伸至參數高效微調、檢索增強生成及多模態模型等新興技術。進一步分析生成式 AI 在臺灣金融機構中的應用情境，如智慧客服、法遵審查、投資研究與流程自動化，並透過多項實務案例展現其實際成效。

本文同時亦探討生成式 AI 應用所帶來的風險與挑戰，包含資料隱私、模型偏見、可解釋性、技術穩定性及法律責任等問題。建議金融機構導入聯盟式學習（Federated Learning）、差分隱私（Differential Privacy）與可解釋技術，建立人機共治與審核機制，並強化法遵與內控架構。最終指出，唯有在創新與風險控管間取得平衡，方能實現生成式 AI 於金融領域之永續應用。

本文主要結論如下：

(一)金融機構可持續推動生成式 AI 技術深化並與業務融合

(二)金融機構宜導入以隱私與倫理為核心之 AI 應用準則

(三)金融機構宜建立 AI 法遵、內部控制及責任歸屬機制

目　次

一、序言

生成式人工智慧（又稱生成式 AI，Generative AI）係指一種能夠根據輸入資料之分佈，透過學習模型參數以獲取隱含特徵，進而生成全新樣本之技術。此種技術涵蓋文本、影像、音訊、程式碼等多種模態（Modality）之生成，已成為當前人工智慧研究與應用之重要領域。

自 2017 年 Vaswani[1]與其團隊提出 Transformer 架構，以自注意力（Self-Attention）機制徹底革新序列模型範式以來，生成式 AI 已有顯著進展與商業化應用。Transformer 之多重自注意力（Multi-Head Self-Attention）機制能夠並行學習序列中不同位置之關聯性，且能夠有效理解句子中與距離較遠詞彙之間的語意關係（語意長程依存關係）；模型也能從輸入資料中辨識出有意義的關鍵資訊或特徵（特徵抽取），進而大幅提升對語句整體語意與結構之理解能力。

隨著大型預訓練語言模型（又稱大型語言模型，Large Pre-trained Language Models, LLMs）問世（如 OpenAI 之 ChatGPT、Google 之 Gemini 等），採用自回歸（Autoregressive）解碼器架構以逐詞生成文本，並藉由最小化交叉熵損失（Cross-entropy Loss）進行參數優化，能夠理解高階語意結構及豐富上下文關係。大型語

[1] Ashish Vaswani 為知名人工智慧與機器學習領域之研究學者，其最廣為人知的貢獻為於 2017 年與其團隊在 Google Brain 共同發表論文〈Attention Is All You Need〉，正式提出 Transformer 架構。此一架構以「自注意力機制（Self-Attention）」為核心，突破過往序列模型如 RNN 之限制，成為現今生成式 AI（Generative AI）與大型語言模型（如 GPT、BERT 等）之基礎。

言模型經過微調（Fine-Tuning）後，能適用於指定或特定領域文字生成需求；此外，檢索增強生成（Retrieval-Augmented Generation, RAG）技術則能優化大型語言模型的內容生成，RAG 透過外部檢索模組，從預設的知識庫中擷取相關資訊，以補足模型因訓練時間限制所造成的知識不足。此機制不僅能有效降低模型因所知不足產生『幻覺』的風險，更能顯著提升生成結果的準確性與可追溯性。在運算資源較侷限，無法進行微調模型之情形下，亦常用 RAG 技術直接優化模型的輸出。

金融產業有感於生成式 AI 具備對結構化與非結構化資料之強大生成能力，遂成為該技術應用之先行領域之一。臺灣主要金融機構自 2024 年起，加速在內部流程自動化、員工培訓優化、智能客服、合規審閱、投資研究與精準行銷等多項場景部署生成式 AI，成為數位轉型之核心動力，金融監督管理委員會亦於同年發布《金融業人工智慧（AI）應用指引》[2]，明訂系統規劃、資料治理、模型驗證與持續監控四大生命週期階段，並要求金融機構依 AI 系統之風險等級採行相應控管措施，以維護客戶權益與金融穩定。該指引為金融界導入生成式 AI 提供法規遵循框架，並引導業者在技術、資安與治理三者間應取得平衡。

生成式 AI 雖推動金融創新，然亦伴隨倫理與風險挑戰，包括生成內容失真（幻覺）、模型偏見導致不公平決策，以及個資洩露

[2] 金融監督管理委員會於 2024 年 6 月 20 日正式發布《金融業運用人工智慧（AI）指引》，旨在引導金融機構於導入及使用 AI 技術時，能夠兼顧風險控管、消費者保護與資訊安全，並促進金融創新發展。

風險。因此，金融機構應強化資料治理、導入公平性與可解釋性機制，並建立審慎的人機共治架構，以確保 AI 應用符合法規、保障客戶權益並維護社會信任。

本文從序言開始進行生成式 AI 之相關介紹；第二章闡述生成式 AI 運作原理；第三章舉例說明生成式 AI 於金融領域之常見應用；第四章簡介生成式 AI 於金融領域常見應用之實際案例；第五章探討技術突破與未來發展；第六章提出生成式 AI 於金融領域之風險與挑戰；第七章為結論。

二、生成式 AI 原理簡介

生成式 AI 之核心在於「生成模型」（Generative Model），旨在學習資料之機率分佈，進而生成具備相似特徵之新樣本。依據模型設計與訓練方式，可區分為下列主要類型：

(一)自回歸模型（Autoregressive Model）

自回歸模型依序列先後順序逐步預測下一元素，典型代表為 GPT 系列。此類模型採用多層 Transformer 解碼器架構，以先前生成的詞彙作為上下文，透過 Query-Key-Value 注意力機制進行計算，逐詞選取機率最高的詞彙，並以交叉熵損失函數最小化訓練誤差。此機制允許模型捕捉語意依存關係，並兼顧語義與語法層面之整體一致性。

(二)變分自編碼模型（Variational Autoencoder, VAE）

VAE 屬於概率生成模型，透過編碼器（Encoder）將輸入數據

映射至潛在空間之概率分佈，再由解碼器（Decoder）從該分佈中重建樣本。VAE 適用於多模態資料之生成，如影像與音訊。此類模型設計能減少生成樣本之多樣性與解析度間之取捨，使得跨類別生成更具連貫性。

(三)生成對抗網路（Generative Adversarial Network, GAN）

GAN 由生成器（Generator）與鑑別器（Discriminator）構成對抗架構，生成器生成以假亂真之樣本欺騙鑑別器，鑑別器則學習分辨真偽，雙方藉博弈式訓練共同進化。GAN 在影像、音樂等領域展現優異生成品質，惟其訓練不穩定性與模式崩解問題，仍需透過架構改進與正則化（Regularization）技術加以緩解，生成對抗網路

資料來源：本文整理

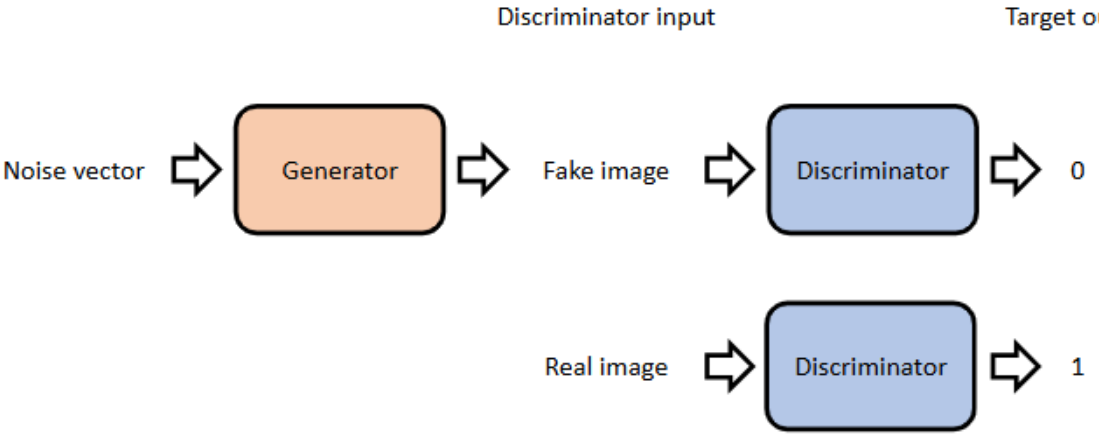

圖 1：生成對抗網路流程如圖

流程如圖 1。

(四)Transformer 與多頭自注意力機制（Multi-Head Self-Attention）

Transformer 架構的核心為自注意力機制，它允許模型評估序

列中每個詞彙與其他所有詞彙之間的語意相關性與重要性 。透過線性映射，每個詞彙會轉換為三種向量：查詢向量（Query, Q）代表當前詞彙欲尋求的資訊、鍵向量（Key, K）代表其他詞彙所能提供的資訊，以及值向量（Value, V）則包含實際的語意內容 。

模型透過 Query 與 Key 的內積計算，判斷詞彙間的語意匹配程度，再經由 Softmax 運算得到注意力權重 。這些權重會乘上 Value 向量並加總，最終形成一個融合上下文語意的豐富表徵。這種機制使得 Transformer 能夠有效捕捉語句中遙遠詞彙之間的語意長程依存關係，並從輸入資料中高效地抽取關鍵特徵，大幅提升對語句整體語意與結構的理解能力 。

(五)微調（Fine-Tuning）與檢索增強生成（Retrieval-Augmented Generation, RAG）

透過領域專屬微調，可使預訓練模型在少量標註資料上迅速獲取特定領域知識；然而，模型參數易受過度擬合影響。Low-Rank Adaptation（LoRA）等參數高效微調技術，僅需增加極少量可訓練參數，更有利於領域適配與資源有限場景部署。

RAG 則結合向量檢索與生成模型，先檢索相關文檔，再將其內容作為上下文送入生成模型，既確保生成結果之時效性，亦降低模型出現幻覺之機率，RAG 為現今常用的 LLM 查詢作法，流程如圖 2。

(六)多模態生成模型（Multimodal Generative Models）

資料來源：本文整理

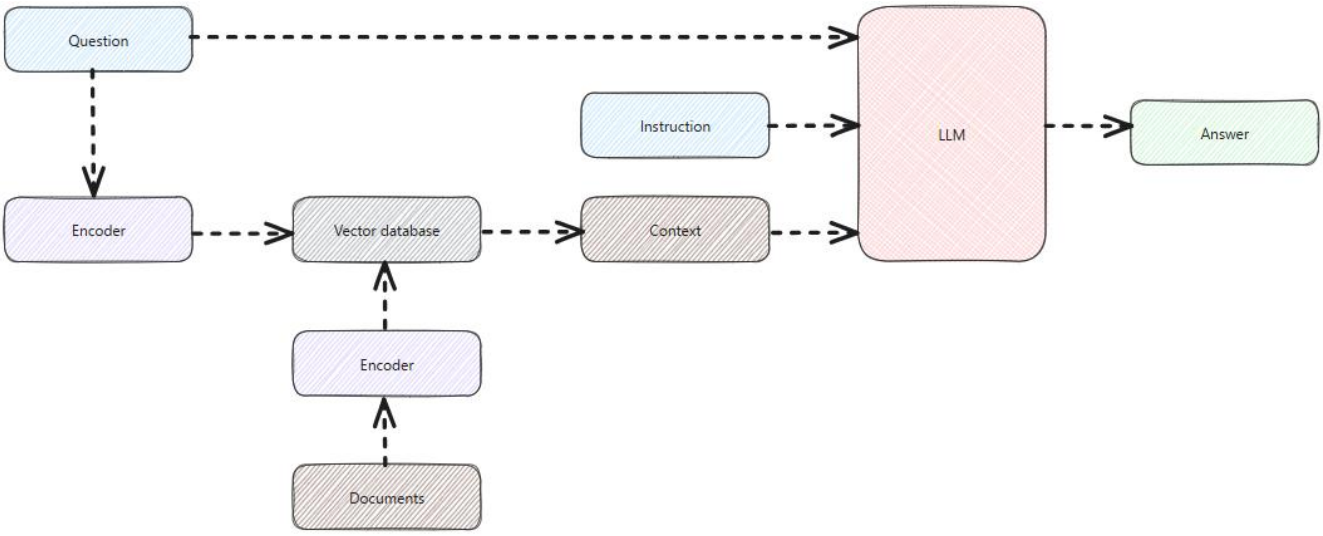

圖 2：RAG 查詢流程圖

隨著 CLIP[3](Contrastive Language–Image Pre-training)與 Vision Transformer（ViT）等多模態架構出現，生成式 AI 不再侷限於文字，亦可同時處理圖像、表格與語音。多模態模型透過共享潛在空間或對齊技術，將不同類型的數據映射至統一表徵，進而生成跨媒介內容，應用於報表自動化解析、智能行銷素材創作等領域。

(七)零樣本與少樣本學習（Zero-Shot & Few-Shot Learning）

受惠於大規模預訓練與自注意力架構，生成式 AI 發展出在無需或僅需極少標註樣本下，即可執行新任務之能力。Zero-Shot 學習利用模型內隱參數表徵之知識，直接對未知任務進行推理；Few-Shot 學習則透過提示設計（Prompt Engineering）加入少量樣本，即可顯著提升下游任務表現，降低資料標註成本。

[3] CLIP（Contrastive Language－Image Pre-training）係由 OpenAI 開發之多模態模型，能同時理解文字與圖像。其核心為透過對比學習，將文字描述與對應圖像映射至同一向量空間，使模型能執行圖文配對、分類與生成等任務，廣泛應用於多模態搜尋與生成式 AI 中。

三、生成式 AI 於金融領域常見應用

生成式 AI 憑藉其處理結構化與非結構化資料之卓越生成能力，正深刻變革傳統金融業多項核心業務。本文從智能客服與虛擬助理、合規審閱與法遵自動化、投資研究與決策支援、行銷與客戶洞察，以及內部流程自動化五大面向，闡述生成式 AI 於臺灣金融機構之實際應用與所創造之價值。

(一)智能客服與虛擬助理

於客戶服務領域，生成式 AI 已深度應用於智能客服與虛擬助理。諸多銀行將大型語言模型整合至線上客服及行動應用程式，藉由結合語音辨識、文字轉語音與人臉辨識等技術，得以提供近似人類對話品質的全天候互動體驗。這類系統能根據使用者性別、年齡等資訊，智能推薦信用卡及理財產品，並透過虛擬角色增強用戶互動。

此外，生成式 AI 智能客服系統結合大型語言模型與結構化 FAQ 資料，亦可有效提升話務替代率，顯著提高服務效率 。

(二)合規審查與法遵自動化

金融監理要求嚴格，法規內容龐雜，導致合規審查流程耗時且易出錯。生成式 AI 結合檢索增強生成技術，可自動擷取最新法規內容並產出摘要報告，應用 RAG 之法遵工具可提升審閱效率，且因資訊來源可追溯，有效減少因 AI 幻覺所導致的風險。

臺灣多家金控機構亦已於內部部署法遵自動化專案，如應用

生成式 AI 比對法條異動並進行風險初評，大幅降低人工作業時間，顯著提升作業效率與準確性。

(三)投資研究與決策支援

於金融投資研究領域，生成式 AI 可即時分析財報、新聞稿及市場數據，自動撰寫分析報告與策略建議，協助研究人員更聚焦在高層決策。如：整合生成式 AI 與圖表視覺化工具，生成個人化投資建議，協助投資人即時掌握市場動向。

此外，部分券商已於行動應用中導入 AI 選股工具，用戶僅需回答簡單問題，即可獲得依據其風險屬性推薦之投資組合，提升操作便利性與互動體驗。

(四)行銷與客戶洞察

金融行銷需兼顧合規性與個人化，生成式 AI 能根據客戶屬性、交易行為與社群互動紀錄，自動產生文案、廣告素材與社群內容。如：導入 AI 所設計之行銷活動提升點擊率；應用生成式 AI 自動產出季度宣傳素材，並透過測試優化文案，使產品推廣效率大幅提升。

(五)內部流程自動化

金融機構透過生成式 AI 進行內部流程自動化，主要應用於文件摘要、會議紀錄整理、合約審閱與知識問答等作業。系統可結合內部知識庫與 RAG 技術，快速擷取關鍵資訊並生成報告，減少人工查詢與處理時間，提升作業效率與準確性，並降低營運成本。

四、生成式 AI 於金融領域之應用實例

在臺灣，生成式 AI 已由概念逐步邁向實務應用，金融機構陸續建置專屬平臺與應用專案，結合自注意力模型、檢索增強生成技術與內部知識庫，實現智慧客服、決策支援、員工培訓與合規審閱等多元場景。以下擇要介紹五項具體實務案例：

(一)玉山銀行通用型生成式 AI 平臺「GENIE」

玉山銀行於 2024 年初對內推出通用型生成式 AI 平臺「GENIE」，整合集團內部累積的金融知識及法規資料，透過微服務架構與 RAG 機制實現之應用如下：

1. API 化與場景擴展：GENIE 採微服務設計，各內部系統可透過 API 快速整合生成功能，支援文件摘要、圖像生成、法規比對、會議記錄彙整等 11 項應用。

2. 檢索增強生成（RAG）：藉由內建的向量化金融知識庫，先行檢索關聯文檔段落，再作為提示詞輸入語言模型，以提升回應準確性與可追溯性。

3. 資訊安全強化：GENIE 採 Microsoft Azure OpenAI 封閉模型，並申請資料外洩防護（DLP）專利，確保客戶敏感資料安全。

(二)凱基人壽「智能業務對練 2.0」

凱基人壽於 2024 年首季正式推出「智能業務對練 2.0」，首

創將生成式 AI 應用於業務員訓練場域，成果如下：

1. 場景式模擬對話：可依目標客群設定模擬對話角色，如「退休族」、「年輕高資產族群」等，系統根據客戶輪廓生成對應提問，並給予反饋與評分。

2. 自動化教材產製：內建 RAG 模組整合法規、產品資訊與銷售指引，自動生成訓練教材與評量題目，輸出格式支援 PDF 與 Web 介面。

3. 訓練成效明顯：導入後，新進人員模擬測驗平均分數提升。

(三)台北富邦銀行：「Fubon＋」App 生成式 AI 助理

台北富邦銀行於 2024 年推出「Fubon＋」行動銀行 App 中導入生成式 AI 助理，整合客製化財務分析與產品推介功能，概述如下：

1. 個人化投資建議：系統分析客戶歷史交易與財務目標，主動生成投資組合建議，並透過自然語言說明風險與報酬模型。

2. 即時產品諮詢回應：對貸款或保險相關詢問即時解析條款與優惠資訊，並輔以對話介面強化互動體驗。

3. 全通道服務一致性：該 AI 助理亦部署於官網與客服平台，實現跨平台一致服務。

(四)富邦金控「首屆 GenAI 黑客松」

富邦金控於 2024 年舉辦首屆集團級生成式 AI 黑客松，結合內部數據與外部專家資源進行創新研發，具體作法如下：

1. 技術與數據開放：提供 Azure OpenAI API、GPU 資源與去識別化金融資料作為開發素材，並安排法遵專家輔導。
2. 跨部門參與：參賽隊伍涵蓋業務、工程、行銷與法遵等部門。
3. 實務落地原型：如「反詐監控助手」、「財報分析生成工具」等，將於 2025 年第一季逐步部署至內部系統。

(五)臺灣企銀「Hokii AI LAB」與樂齡行動銀行

臺灣企銀於 2024 年金融科技展中推出「Hokii AI LAB」，並展示結合生成式 AI 之銀髮族專屬行動銀行原型如下：

1. Dr. Hokii 智能助理：整合 GPT 模型與 Azure 認知服務，支援中、英、臺語語音互動，提供帳務查詢與操作指引。
2. 裸視 3D 虛擬形象：以 Hokii 虛擬人物進行互動，結合表情辨識協助長者溝通，增強使用者親和力。
3. 安全與照護融合：App 內建跌倒偵測與求助系統，可主動發出通知至家屬與客服中心。

五、生成式 AI 技術突破未來發展

得益於近期多項核心技術之突破，生成式 AI 可望在金融領域有更多的深度應用，本章節探討技術進展與未來可能的發展趨勢

如下：

(一)參數高效微調（PEFT）

Low-Rank Adaptation（LoRA）為近期備受關注的參數高效微調技術。其核心理念為凍結預訓練模型之權重，並於每層 Transformer 架構中注入可訓練的低秩矩陣(Low-Rank Matrix)，以大幅減少需調整之參數數量。相較於全面微調，LoRA 能將可訓練參數數量減少至原模型之約 0.01%至 0.1%，並降低 GPU 記憶體需求至原先之三分之一，顯著提升訓練效率與部署靈活性。

在金融應用方面，LoRA 可有效將通用大型語言模型（如 LLaMA）調適至特定任務專用，如合規審查、風險報告生成與自動化投資研究等。

(二)聯盟式學習與隱私保護

在金融業高度重視客戶隱私之背景下，聯盟式學習（Federated Learning）成為生成式 AI 技術合作之關鍵。此方法允許多家機構在不共享原始交易資料、用戶紀錄或契約文件等敏感資訊之情況下，共同訓練協同模型，以提升反洗錢、信用風險評估與詐欺檢測等任務之整體效能。

此外，聯盟式知識蒸餾（Federated Knowledge Distillation）[4]亦正加速於金融生成式 AI 中之應用。透過安全協議，各銀行可在不

[4] 聯盟式知識蒸餾（Federated Knowledge Distillation）係一種結合聯盟式學習與知識蒸餾技術的方法，各參與者不分享原始資料，而是交換模型預測結果（如標籤），再由中心端或成員端之模型進行學習，達到提升模型效能與保障資料隱私的雙重目的。

洩露私有模型參數的同時，共享預測結果並對生成模型輸出進行集體優化。

(三)多模態融合與跨媒介生成

隨著 CLIP、Vision Transformer 與多模態融合架構之出現，生成式 AI 已不再侷限於純文本輸入，亦能同時處理影像、表格、音頻與即時視訊等多樣化數據源。CLIP 模型透過對比學習（Contrastive Learning）[5]對圖像與文本進行對齊，為金融機構在 OCR 報表解析、交易圖表說明及營運儀表板生成等場景提供強大工具。

在信用評分與風險分析中，生成式 AI 可將掃描之紙本契約與客戶手寫簽名影像，與契約文本及系統記錄一併輸入多模態模型，生成包括風險摘要、契約條款異常提醒與法律建議之綜合報告，可有效提升審閱效率。

未來可以預期將有更多金融應用結合聲紋識別與語音生成技術之混合介面，結合多語預訓練模型，金融機構可同時支援中、英、日、韓等多種語言之即時多元化客戶服務。

(四)可解釋性與負責任人工智慧

隨著生成式 AI 於金融決策鏈中扮演越來越關鍵的角色，模型可解釋性（Explainable AI, XAI）[6]與負責任人工智慧（Responsible

[5] 對比學習（Contrastive Learning）係一種自監督學習方法，透過最大化相似樣本間的表徵相近程度、最小化不同樣本間的表徵相似度，來學習資料之潛在特徵結構。常應用於圖像、文本與多模態任務中，可有效提升模型在無標註資料下的表徵學習能力。

[6] 模型可解釋性（Explainable AI, XAI）係指讓 AI 模型的決策過程透明可理解，使用者能夠理解模型

AI）成為不可或缺之研究課題。XAI 技術可透過注意力可視化、特徵貢獻分析及反事實解釋等方法，揭示模型決策之內部邏輯；於信貸審批與投資建議中，能協助風控與合規團隊快速找到生成結果異常之原因。

此外，金融機構亦開始導入人工智慧影響評估（AI Impact Assessment, AIIA）流程，定期對生成式 AI 模型進行偏見檢測、隱私風險評估與效能回歸測試，並依據結果調整訓練數據及模型參數，以確保人工智慧應用之合規性與透明度。

六、生成式 AI 於金融領域之風險與挑戰

生成式 AI 於金融領域帶來諸多創新應用，然亦伴隨多重挑戰與潛在風險。以下章節將從資料隱私與安全、法規遵循與監理、模型偏見與公平性、可解釋性與透明度、技術與營運風險、對抗攻擊與網路安全，以及法律與責任歸屬七大面向，深入檢視關鍵議題並提出相應對策。

(一)資料隱私與安全

金融機構若要運用生成式 AI，就需處理大量敏感客戶數據，包括交易記錄、個人身分資訊、信用評分等。若無妥善保護，極易造成個資外洩或不當使用。研究指出，傳統之聯盟式學習雖可降低資料集中風險，然仍存在模型更新過程中可能洩露個人資訊之漏洞，因此尚需結合差分隱私（Differential Privacy）[7]等技術進行強

如何做出預測，進而提升信任、符合監管要求並便於錯誤診斷。

[7] 差分隱私（Differential Privacy）係一種保護個資的技術，透過在查詢結果中加入隨機雜訊，使外部

化。

此外，生成式 AI 模型常部署於雲端環境，若雲端供應商之安全防護不足，或 API 串接過程中憑證管理不當，皆可能成為攻擊者滲透之入口。美國國家標準與技術研究院（NIST）建議，金融機構應對人工智慧服務進行端到端（End to End）之加密、嚴格存取控制及常態化滲透測試，以防止中間人攻擊及模型竊取。

(二)法規遵循與監理

生成式 AI 因具自動生成文本之能力，幻覺現象易導致輸出錯誤或偽造資訊，若直接引用於法遵報告、合約審閱或投資建議，極易觸犯法令或引發監理爭議。監理機關對於人工智慧模型之合規性已有初步指引，但對於生成式 AI 之特性尚無專門規範，金融機構須在人工智慧影響評估中納入質量驗證與審核流程，以確保生成結果之合法性與準確性。

同時，責任歸屬不明亦為重大挑戰。當人工智慧模型建議錯誤投資決策或生成誤導性信貸評估報告，若造成客戶財務損失，究竟模型供應商、金融機構或使用者應承擔何種賠償責任，目前仍缺乏統一法規。建議業者於與供應商簽訂合約時，明確規範服務水準協議、責任限制條款，並購買專門之科技責任險，以降低潛在賠償風險。

無法判斷某筆資料是否存在於資料集中，保障個體隱私。

(三)模型偏見與公平性

生成式AI模型在訓練過程中可能繼承或放大訓練數據中之偏見，導致對特定族群之貸款申請或風險評估出現不公平待遇。根據Marsh McLennan 之報告，模型若未妥善監控與校正，便可能在信用評分、保險費率計算等環節造成系統性歧視。

為降低偏見風險，金融機構應實施以下措施：其一，於資料清理階段納入偏見檢測；其二，於模型訓練後執行公平性評估指標；其三，定期進行模型監控，確保模型表現未因市場或客群變化而產生新偏見。

(四)可解釋性與透明度

生成式 AI 之深度網絡結構本質上屬「黑盒」，用戶及內部審查單位難以理解輸出結果之內部推理邏輯。在信貸審核、投資建議等高風險決策場景中，缺乏可解釋性將對監理合規構成嚴重阻礙。人工智慧模型之可解釋性為金融監管核心需求，必須提供清晰、可追溯之決策依據，並能說明為何生成此內容。

建議應用注意力視覺化（Attention Visualization）[8]、特徵貢獻分析與反事實解釋（Counterfactual Explanations）[9]等 XAI 方法，並結合可解釋報告自動生成，協助法遵與風控人員快速找出模型可能出現之偏差或錯誤。

---

[8] 注意力視覺化（Attention Visualization）為透過圖像方式呈現模型在處理輸入資料時關注的部分，常用於解釋 Transformer 模型決策依據，提升 AI 模型的可解釋性與透明度。

[9] 反事實解釋（Counterfactual Explanations）為一種可解釋 AI 的方法，透過回答「若輸入改變某些特徵，模型結果會如何不同？」來說明模型決策邏輯。例如：若收入提高至某數值，貸款申請就會通過，藉此揭示模型的決策邊界與關鍵因素。

(五)技術與營運風險

生成式 AI 所依賴之大型語言模型通常參數龐大、算力需求高，導致推理服務成本居高不下，且運維複雜度高。銀行在大型語言模型之推理階段，若無有效之算力優化與資源分配策略，付出之成本可能遠超出預期。

此外，將人工智慧服務納入既有資訊科技生態後，須解決模型更新與持續整合/持續部署（CI/CD）流程整合、系統彈性擴展與服務可用性管理等問題，並須規劃完善的災難復原與業務持續性方案，以確保人工智慧服務在高峰交易時段或重大事件期間之穩定性與可靠性。

(六)對抗攻擊與網路安全

生成式 AI 易遭受對抗性攻擊（Adversarial Attacks），攻擊者可通過精心設計之輸入，如對客戶對話提示進行微小篡改，導致模型輸出錯誤或敏感資訊洩露。此外，模型竊取（Model Extraction）等攻擊手段，亦可能令有心者去重建模型或辨識訓練集中的個別樣本。

金融機構須在模型部署端採行對抗性防禦，如對抗訓練（Adversarial Training）、輸入驗證（Input Sanitization）、模型水印技術（Watermarking）、以及動態監控異常請求流量，以降低此類攻擊之風險。

(七)法律責任與倫理監管

當生成式 AI 自動生成投資策略或保險方案建議，並導致客戶依其操作蒙受損失時，金融機構將面臨誤導性陳述或詐欺的法律指控 。為此，金融機構應建立「人機共治」(Human-in-the-Loop)[10]與「人機最後決策」(Human-on-the-Loop)機制 ，確保所有關鍵決策皆有足夠的人工審核與最終責任歸屬 。

同時，須密切追蹤區域性與國際間關於人工智慧合規之新興法規，如美國之《人工智慧責任法》、歐盟之《人工智慧法規》等草案，並與內部法律、風控及合規部門協調，制定切實可行的合規框架與審核流程。

七、結論

本文經詳細探討及分析生成式 AI 於金融領用應用及突破，提出以下 3 點結論：

(一)持續深化技術並與業務融合

生成式 AI 於自然語言處理、多模態資料生成、語意理解等領域之展現已日趨成熟，對金融產業之價值創造已非單一場景導入，而是整體營運流程再造之關鍵驅動力。

以技術層面而言，未來發展重點應聚焦於三項關鍵：其一，強化模型推理效率與算力優化，以降低大型語言模型部署與維運成

[10] 人機共治(Human-in-the-Loop)係指 AI 系統運作過程中關鍵環節需由人類即時介入決策；人機最後決策(Human-on-the-Loop)則是 AI 自主運作，但人類保留最終審查與否決權，確保可控性與責任歸屬。

本，提升模型於邊緣設備或私有雲環境中之適應性；其二，發展專用的金融語料微調技術，提升生成內容之準確性及與產業語境之契合度，減少幻覺現象之發生；其三，導入多模態能力與多語言支援，強化模型處理財務報表、圖像與非結構資料之能力，提升 AI 應用廣度。

同時，金融機構宜建立跨部門 AI 研發團隊及產學合作機制，加速新技術導入與實驗性應用試點，促進模型研發成果快速轉化以具有商業價值。透過技術深化與場景融合雙軌並進，金融產業方能於全球 AI 浪潮中建立技術領先優勢與可持續競爭力。

(二)以隱私與倫理為核心導入 AI 準則

在生成式 AI 逐步發展至金融業之營運核心，如何於創新與隱私之間取得平衡，成為金融機構不可迴避之挑戰。金融服務涉及大量個人敏感資料，包括身分、財務狀況、交易行為與風險偏好等，若生成式 AI 之資料處理與建模過程未能妥善控管，易產生資訊濫用、演算法偏見與歧視性結果，進而侵害消費者權益與社會正義。

鑑此，各金融機構宜採行具前瞻性之資料治理策略，包括資料最小化、目的限制與使用可追溯性等原則，技術層面則須導入差分隱私等資料生成保護機制，以降低模型訓練過程中可能之資訊洩露風險，並宜推行模型偏誤偵測與公平性檢測，持續監控模型對不同群體所產生之影響差異，避免強化既有社會之不平等。

更重要者，金融機構宜設立 AI 倫理審議機制，由資訊、法律、

消費者保護與永續治理等跨領域專家組成，定期審查模型用途與資料來源是否合法、以及符合倫理，並確保所有生成內容符合透明性、可解釋性與人性尊重原則。倫理導向不僅維護企業聲譽，更可強化金融業與大眾間之長期信任關係，實現科技向善與社會共融之雙重目標。

(三)建立 AI 法遵、內控與責任歸屬機制

生成式 AI 技術於金融場域之快速發展，已促使各國監理機關積極研擬 AI 專屬之監管框架，AI 應用必須納入法遵架構，以確保其在合法性、透明性與風險可控性上具備高度可監管性。

因此，金融機構應主動建立 AI 法遵與內部控制機制。具體作法含括：其一，制定 AI 應用登錄與審核制度，涵蓋模型來源、訓練資料、適用場景與風險等級分類；其二，推動 AI 影響評估，針對每一項應用進行風險辨識與利害關係人評估，並設立審查、稽核與異常通報流程；其三：導入第三方合規審查，確保模型開發與部署流程符合內外部監理要求。

此外，應建立清晰的責任歸屬架構與法律應對機制，針對模型錯誤導致金融損失之情境，最終仍應交由人工決策，AI 模型僅是輔助工具，確保所有重大決策均經人為把關及可追溯處理。

唯有建立穩健的法遵架構及責任機制，金融機構於生成式 AI 應用中方能實踐企業責任、回應社會期待及樹立良好典範。

參考文獻

金融監督管理委員會 (2024) “金融業人工智慧（AI）應用指引”。

Izacard, G. et al. (2023), “Unifying Large Language Models and Retrieval for Knowledge-Intensive Tasks.”

Hu, E. J., Shen, Y., Wallis, P. et al. (2023), “LoRA: Low-Rank Adaptation of Large Language Models.”

Nguyen, D., Sun, Q., et al. (2023), “Financial Text Summarization with Retrieval-Augmented Transformers.”

Hu, E. J. et al. (2021), “LoRA: Low-Rank Adaptation of Large Language Models.”

Radford, A. et al. (2021), “Learning Transferable Visual Models From Natural Language Supervision (CLIP).”

Dosovitskiy, A. et al. (2021), “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (ViT).”

Li, X., Wang, S., Ding, Z. (2021), “Survey of Adversarial Attacks and Defenses for Deep Learning.”

Brown, T. et al. (2020), “Language Models are Few-Shot Learners.”

Lewis, P. et al. (2020), “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.”

Suresh, H., Guttag, J. (2019), “A Framework for Understanding Unintended Consequences of Machine Learning.”

Guidotti, R. et al. (2018), “A Survey of Methods for Explaining Black Box Models.”

Vaswani, A. et al. (2017), “Attention is All You Need.”

Goodfellow, I. et al. (2014), “Generative Adversarial Networks.”

Binns, R. (2018), “Fairness in Machine Learning: Lessons from Political Philosophy.”

永豐金控 AI 法遵智能平台 (鉅亨網)
https://news.cnyes.com/news/id/4743495

國泰金控 AI 智能法規解讀 (鉅亨網)

https://news.cnyes.com/news/id/4743495

玉山銀行通用型生成式 AI 平臺「GENIE」(商益)
https://www.businessyee.com/article/3929-GENIE

凱基人壽「智能業務對練 2.0」(凱基人壽官方網站)

https://www.kgilife.com.tw/zh-tw/articles?id=9914eab1d8334a5c9858aac8d8a4bce6&tid=bc1b63e6b0904af9965ac03e3cecb882

台北富邦銀行：「Fubon＋」App 生成式 AI 助理(富邦銀行官方網站)

https://www.fubon.com/financialholdings/news/news_1241101_460675.htm

富邦金控「首屆 GenAI 黑客松」(富邦銀行官方網站)

https://www.fubon.com/financialholdings/news/news_1241218_520977.htm

臺灣企銀「Hokii AI LAB」與樂齡行動銀行(鉅亨網)
https://news.cnyes.com/news/id/5760370?utm_source=chatgpt.com