數位金融實務規範工作圈
第二季報告

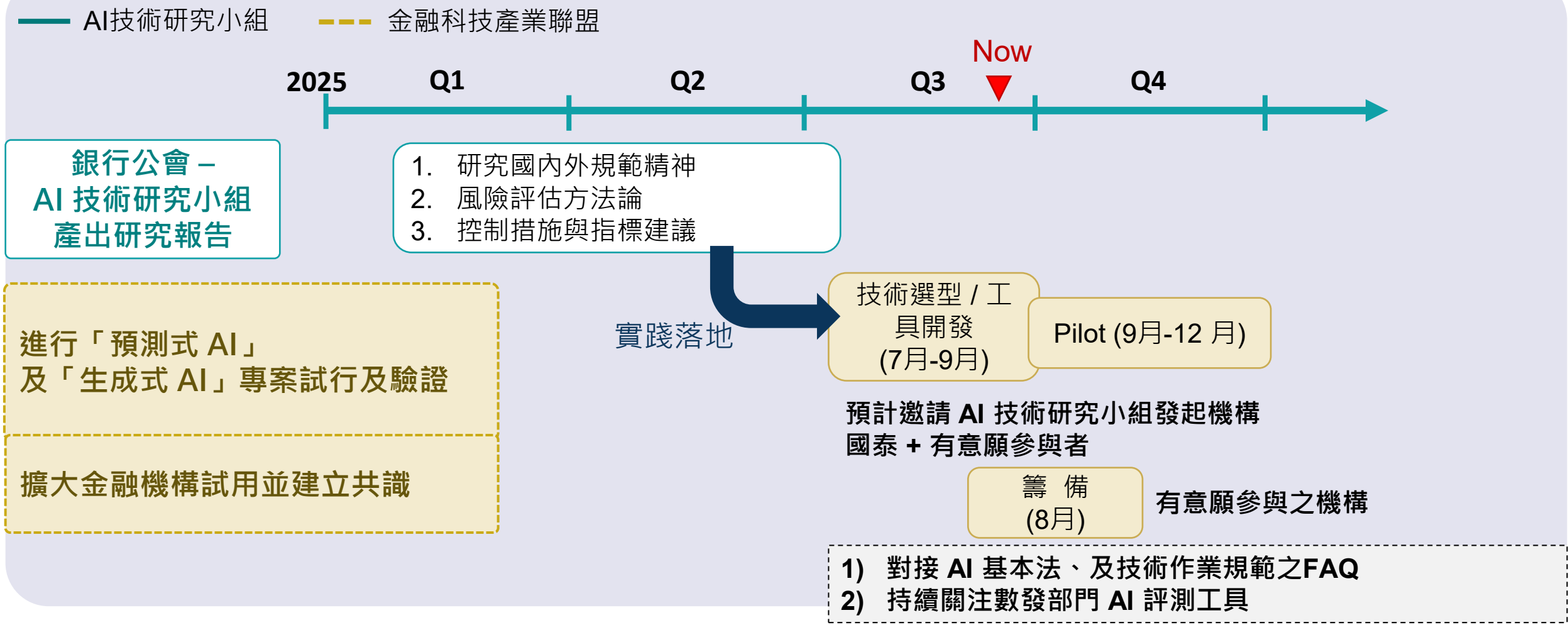
20250821

玉山金控 E.SUN FHC

B. 可程式化的 AI 治理

可程式化的 AI 治理 -時程

從理解精神走向落地實踐：實作可程式化AI的相關方法論和工具，確保金融機構在運用AI技術時能依循標準作業



AI 技術研究小組(讀書會) - 編制與進行方式

- **銀行公會金控業務委員會金融科技創新發展組轄下 - AI 技術研究小組**
- **小組成員**：共計 16 家金融機構，45+ 位成員

  發起機構 - 國泰金控、玉山銀行、中國信託銀行、

  參與機構 - 凱基金控、臺灣銀行、台新銀行、第一金控、第一銀行、新光金控、兆豐銀行、王道銀行、富邦金控、合作金庫、永豐銀行、華南金控、土地銀行
- **推動方式**：

  以讀書會分享形式舉辦，事前依照 AI 成熟度將各家機構區分三組。當月由主責之發起機構帶大家導讀與分享相關研究，研究主軸：

  (1) AI 風險評估分級制度與各級控管措施

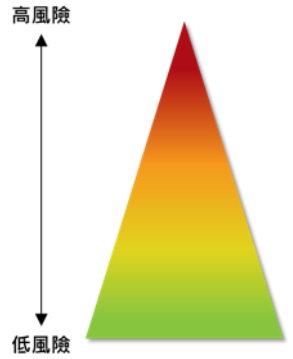
  (2) AI 六大核心原則的關鍵評測指標

讀書會各階段具體產出(1/3)

1 2月 研究國外重要規範與實務做法

研究對象	研究素材/機構	收斂台灣可借鏡發展項目
新加坡	1. MDDI(數位發展及新聞部) 2. CSA (網路安全局) 3. 金融管理局(MAS)	1. 制定執行指南：訂定治理架構、管理措施(包含指引表、風險分級之管制措施) 2. 發展檢測工具：由新加坡政府補助成立的非營利組織發展的AI檢測工具。發展可操作可執行達到可程式化的管理指標 3. 應用自評作業：新加坡政府機關提出治理架構後企業則依業務應用，自評執行狀況
香港	個人資料隱私公署(PCPD)	1. 區分為「自建模型」、「自第三方取得AI解決方案」(包含不限於生成式AI) 2. 隱含的風險分級(/類)概念、不可接受AI應用由組織自行決定 3. 提供不同規範情境建議的流程與實作案例，並對應至"現行適用法規"
歐盟	EU AI Act	1. 補充並定義GPAI ：通用人工智慧被廣泛運用於各場景並可以與下游系統或應用程式整合 2. 風險為導向的管理方法：正面表列不可接受風險，方向以禁止侵害基本人權出發 3. 補充鼓勵創新措施：AI監理沙盒應建立在國家層級、對中小企業或新窗公司可優先進入沙盒、使用訓練計畫等資源，並協助回覆對AI Act法案之疑問

2 3月 風險分級方法論



依據風險承受程度，區分適當的 AI 應用整體風險分級並由各機構自行給予分級定義。

交叉分析法

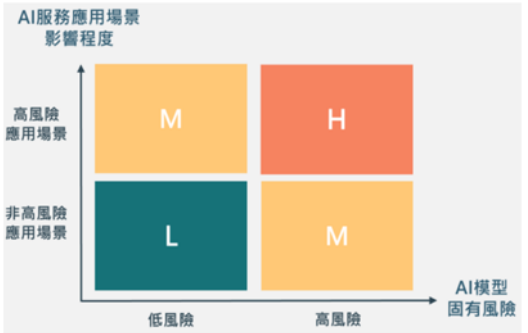
依據各評估面向分別作適當的判斷，求出AI 應用案件的風險等級。

影響對象		內部營運		外部用戶	
對客戶權益或企業之影響性		小	大	小	大
AI決策程度	AI直接判斷	中低	中	中低	高
	人機協作判斷	低	中低	低	中

內部營運大：提供監理規範、產品訂價、准駁決策參考
外部用戶大：提供金融商品交易、准駁、決策

條件矩陣法

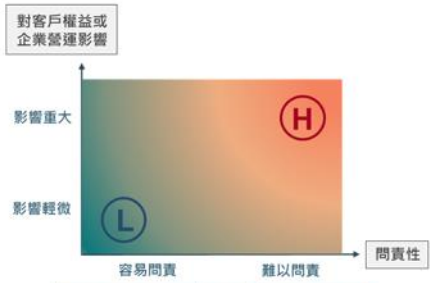
依據自定義條件，任一觸及即屬高風險，並以兩軸線交集切分三個等級。



AI模型固有風險：自主決策程度、產出不確定性程度
AI服務應用場景影響程度：對客戶、公司營運、員工權益

評分法

依據各評估因子給定分數，訂定整體風險總分所對應的風險分級



X軸(問責性)	權重	有無	評分
AI應用程度	2	0	0
模型可解釋程度	3	1	3
總評分			3

Y軸(客戶面)	權重	有無	評分
使用個資	3	1	3
影響顧客權益	2	0	0
服務對象(內/外)	1	1	1
總評分			4

讀書會各階段具體產出(2/3)

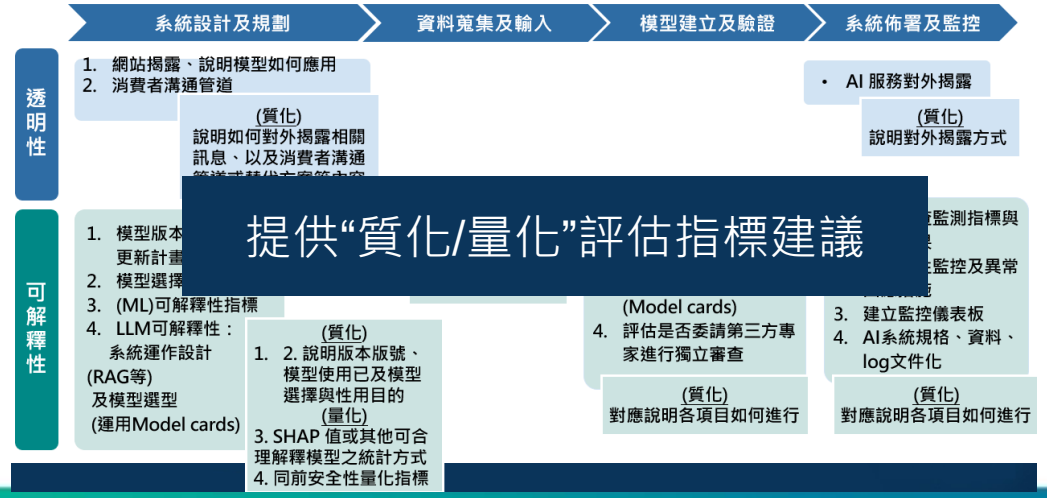
3 **4月** 分級制度對應控制措施建議

生命週期
① 系統規劃與設計
② 資料蒐集及輸入
③ 模型建立及驗證
④ 系統佈署及監控

✖

AI 指引核心原則	控制措施（可視機構內部措施增減）	風險評級 高	中	低	建議實施細項	建議二道防線單位
透明性與可解釋性	1. 模型版本控制、模型檢視更新計畫				既有AI模型如有重大變更(版本)更新，需重新進行AI服務應用的上線程序；或如採用 CI/CD 流程或其他標準作業流程，後續則依 CI/CD 流程或其他標準作業流程辦理。	
	2. 模型選擇說明				提案單位需說明模型類型、選型邏輯與版號。	
	3. 可解 depen					
	4. 網站 用於服				是否提供淺白扼要的	法遵、法務
	5. 消費者溝通管道(含意見提出請求申覆或提供替代方案)				如屬對外使用的AI服務應用，提案單位需說明是否提供替代方案(如真人服務)；或官網等對外通路設有意見回饋專區、使用告知聲明等內容。	法遵、法務
	6. 數據源評估及應用目標定義				評估有效數據源，並明確定義應用衡量目標，以提高模型的透明性及可解釋性	

原則與對應控制措施

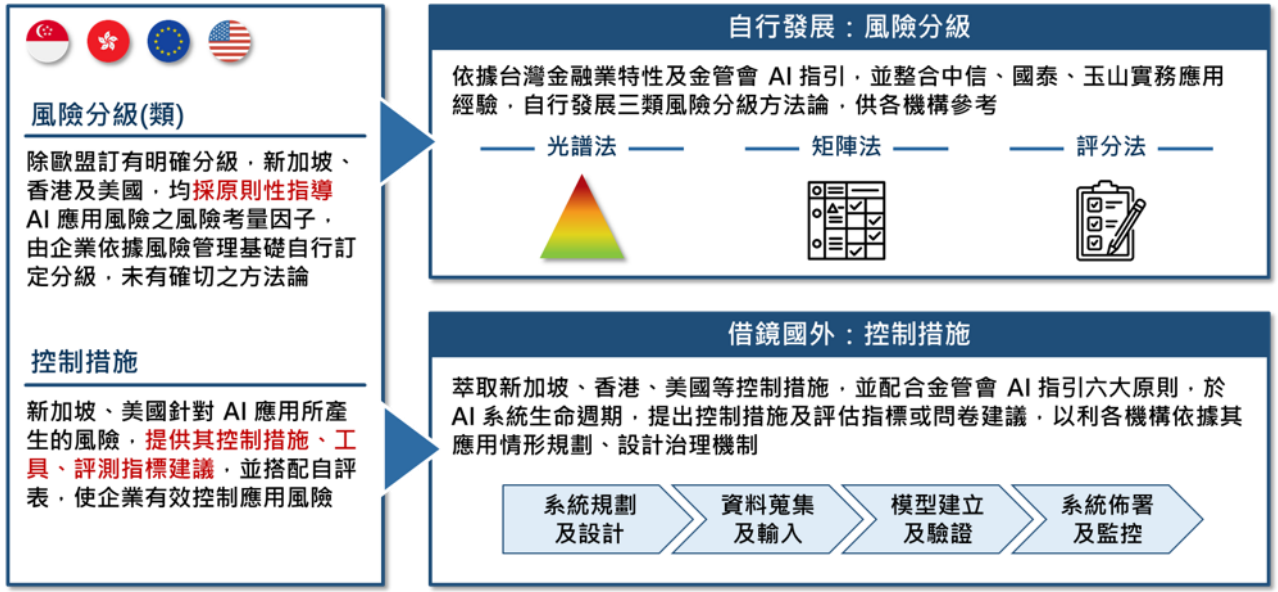


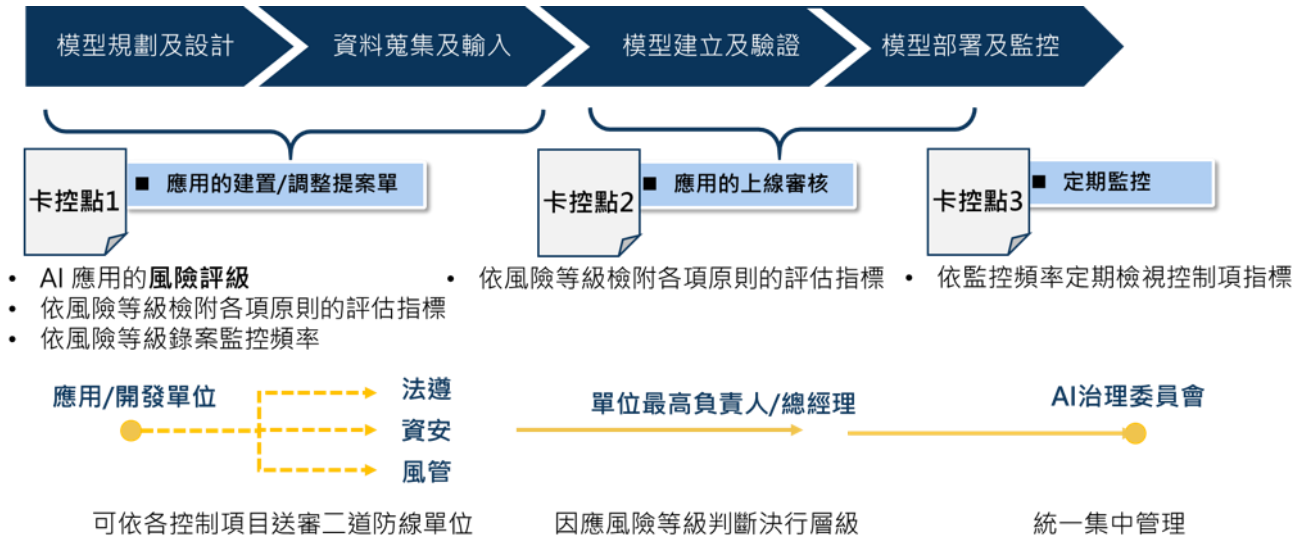
提供“質化/量化”評估指標建議

4 **5月** 各控制措施對應指標

讀書會各階段具體產出(3/3)

5 **7月** 最終報告





AI 創新加速前行，治理更需同步到位

監理趨勢

產業現況

核心痛點

- 金管會已於 2024 年 6 月正式發布「金融業運用人工智慧(AI)指引」，並預計將逐步完善 AI 基本法。
- 從監管單位釋出相關內容可見高度重視在導入 AI 時的風險控管、消費者保護與資訊安全

根據金管會公布金融業應用人工智慧(AI)調查結果，已有超過百家金融機構在不同業務中導入 AI，應用範圍涵蓋智能客服、風險管理到廣告行銷等，但治理仍在建構中

目前的 AI 指引多為「原則性」規範，例如公平性、問責性、透明度與可解釋性。這些原則立意良善，但在實務上缺乏客觀、量化的評估標準，導致：

- **合規舉證困難：** 如何向主管機關「證明」合乎原則的軌跡
- **內部管理不一：** 不同團隊、不同專案對「透明度」的解讀可能完全不同。

可程式化的 AI 治理 – 落地規劃與建議參與角色

Aug.
籌備

Sep.
招募與前置作業

Oct.
Pilot 與工具選型

Nov.
開發與實做
Pilot 報告解讀

Dec.
治理框架釋出

建議參與角色

資料科學家 / AI 工程師 **驗證實務 業務與技術可行性**

1. 實作組織 pilot 模型資訊卡 (Model Card)
2. 評估指標實用性
3. 分享「治理痛點」

風控 /法遵 /資安人員 **建立指標對應規範解讀性**

1. 校準合歸對照文件
2. 量化指標來實踐指引原則的可解讀性

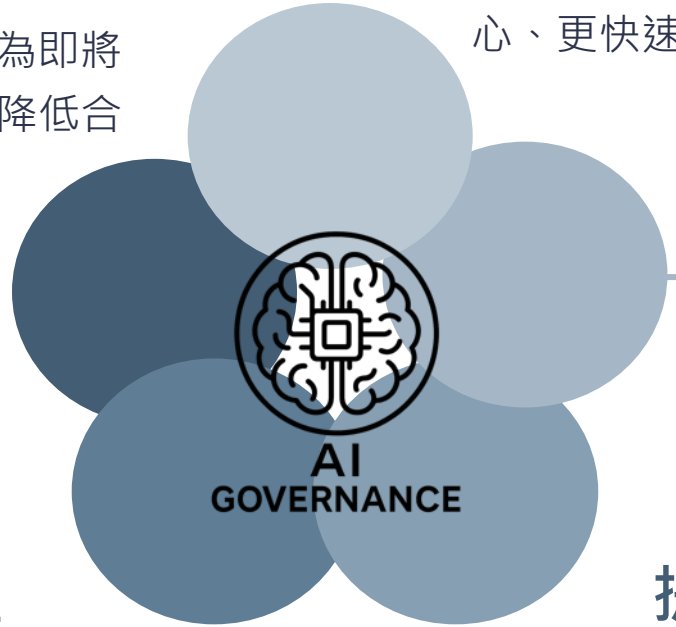
兼顧創新與風控，提升企業核心競爭力

對齊監管期望

具體落實金管會 AI 指引要求，為即將到來的 AI 法規做好準備，有效降低合規風險

加速業務創新

建立可靠的 AI 風險護欄，讓業務單位更安心、更快速地推動 AI 應用，搶佔市場先機



強化風險管理

將 AI 模型風險納入機構風險管理體系，實現標準化、自動化的監控與預警

建立內部標準

打造全機構一致的 AI 治理語言與標準，提升跨部門協作效率

提升客戶信任

透過可證明的公平與透明機制，贏得客戶與合作夥伴的信賴

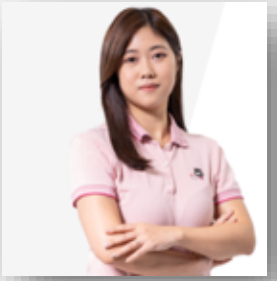
聯盟聯絡窗口

可聯絡聯盟、專案各種事宜

玉山工作圈主要窗口
智金處 林鉦育 資深副總工程師
ntuaha-13240@esunbank.com

玉山工作圈主要窗口
智金處 廖子慧 副總工程師
jcaliao-13701@esunbank.com

可聯絡專案相關事宜



金融無塵室 - PM
智金處 蘇麟雯 副主任工程師
sualice-21979@esunbank.com

可程式化的 AI 治理 - PM
智金處 林仙琪 主任工程師
kikilin-10281@esunbank.com

感謝聆聽

數位金融實務規範 – 目標 B：可程式化的 AI 治理

根據金管會的六大指引，整合銀行公會針對 AI 治理的研究報告，提供對應實務指南，
以建立適用於台灣金融機構的可程式化 AI 治理框架。

管理框架

1. **方法論與工具實作**
   實作可程式化 AI 的相關方法論和工具，確保金融機構在運用 AI 技術時能依循標準作業。
2. **檢測資料集**
   建立檢測資料集，幫助金融機構評估和驗證其 AI 應用的相關指標。
3. **監測與控制**
   制定監測指標，針對 AI 的運作進行觀察，確保符合規範與標準。
4. **風險管理**
   實施過程中，建立相應的風險控制措施，以降低潛在的 AI 風險。

參考資料：Veritas Initiative by MAS
to deal with AI's principles of FEAT
(fairness, ethics, accountability, transparency)