華新麗華股份有限公司
人工智慧應用暨資訊安全管理辦法

1. 權責

1.1. AI 系統負責單位

須依本規章辦法要求，對其管轄之 AI 系統負責採用原則的評估、申請與風險預估，並確保用途合乎規範。

1.2. 資訊單位

負責 AI 系統的日常管理包含平台、架構、整合、技術維運與身分存取之執行責任。

1.3. 資安單位

負責整體安全架構、AI 工具的安全性評估、權限控管及風險監控。

1.4. 法務/合規部門

負責提供法律、合規審查與諮詢、判定高風險使用之法遵要求，以確保符合《個資法》及相關法案。

1.5. 各業務單位

負責 AI 產出結果的最後核實（人為介入原則），以及用途合法性、輸入資料權限、人工覆核、決策/發布及異常通報責任。

2. 作業內容及管理重點

2.1. 名詞定義：

A. 生成式 AI（Generative AI）：指能根據使用者輸入之提示指令（Prompt）生成文字、圖像、程式碼或其他媒體內容的 AI 服務(如 ChatGPT, Claude, Copilot Chat)。

B. 代理式 AI (AI Agent)：具自主規劃、工具調用、資料存取或執行動作能力。

C. 提示詞注入 (Prompt Injection)：指攻擊者透過惡意或特定設計的輸入指令，意圖操控或繞過 AI 模型原有的安全限制，導致系統執行未授權操作或洩漏機敏資訊。

D. AI 系統衝擊評估 (AIA)：AI 系統導入或上線前，針對該系統可能對公司資訊安全、隱私保護、資料機敏性及營運持續性所造成之衝擊，進行系統化的評估。

E. 企業級 AI（Enterprise AI）：係指經公司正式評估、核准、採購或建置，並符

合公司資訊安全、法規遵循、身分驗證、權限控管、資料保護及營運管理要求之人工智慧服務、平台或系統。

F. 公開 / 非企業授權 AI（Public or Non-Enterprise AI）：係指未經公司正式評估、核准、採購或授權使用之公開人工智慧服務、網站、應用程式或模型。

G. 檢索增強生成（Retrieval-Augmented Generation, RAG）：係指人工智慧系統於產生回應前，先自內部或外部知識來源檢索相關資訊，再將檢索結果結合大型語言模型進行生成之技術架構。

H. AI 系統負責人（AI System Owner）：AI 系統負責人係指負責 AI 系統導入、維運、安全管理、權限控管、風險管理及法規遵循之業務或技術管理人員。

I. 高風險 AI（High-Risk AI）：係指其運作結果可能對個人權益、企業營運、資訊安全、法律責任、財務狀況、產品品質、職業安全或社會公共利益造成重大影響之人工智慧系統或使用情境。

2.2. 管理政策核心原則

本辦法依「負責任 AI」（Responsible AI）核心原則制定，包含：

2.2.1. 建立治理及問責機制（Governance and Accountability）

2.2.1.1. 制定 AI 政策與準則、設立專責監督機制、明確角色與責任分工，確保關鍵決策仍受人類監督。

2.2.1.2. 高階管理層應積極參與 AI 策略的制定，確保 AI 發展方向與企業價值觀及業務目標保持一致。

2.2.1.3. 公司必須對使用的系統承擔相應的內、外部責任。內部需指定高階主管負責監督，建立明確的治理架構與問責機制。

2.2.2. 落實可解釋性與透明度（Interpretability and Explainability）

2.2.2.1. 企業應能向利害關係人清楚說明 AI 系統如何運作、依據哪些因素做出決策，以及決策結果可能產生的影響。

2.2.2.2. 對於系統所做出的決策或產出，應對消費者落實適當的資訊揭露（如標記 AI 產出），並在可行範圍內提供合理的可解釋性，確保運作不流於黑箱。

2.2.2.3. 透明與可解釋性的 AI 不僅是建立信任的關鍵，更是符合監管要求的必要條件。

2.2.3. 重視公平性與消除偏見（Fairness and Bias）

2.2.3.1. 人工智慧系統應該公平對待每個人，避免偏見(Bias)與對相似群體產生不同的影響。

2.2.3.2. 本公司在 AI 開發的各個階段應主動識別與緩解偏見風險，實施偏見檢測程序，包括審視訓練資料的多元性與代表性、檢測模型輸出是否存在系統性差異、建立持續監控機制以及早發現並建立系統化的偏見緩解機制。

2.2.4. 確保可靠性與安全性（Reliability and Security）

2.2.4.1. 企業應確保 AI 模型在面對對抗性攻擊、資料品質問題或環境變化時，仍能維持穩定可靠的效能表現。

2.2.4.2. 落實資安防護措施，實施嚴格的資料驗證機制、建立模型效能監控體系、制定異常處理程序、定期進行安全測試與漏洞評估，避免受到惡意攻擊或發生重大系統潰散。

2.2.5. 保護隱私及客戶權益（Privacy）

2.2.5.1. 企業必須確保資料蒐集符合法規要求、實施適當的去識別化處理、建立嚴格的存取控管機制。

2.2.5.2. 在資料的整個生命週期中維護其安全性與保密性，嚴格遵守隱私權保護規範。

2.2.6. 促進永續發展與福祉

2.2.6.1. 管理政策應與環境、社會及公司治理（ESG）目標相結合，確保科技與業務在發展的同時，能降低數位落差並兼顧環境生態的永續。

2.3. 系統獲取與供應商評估

2.3.1. AI 工具採用基礎原則：

2.3.1.1. 使用企業級版本授權 AI 工具，且擁有企業管理後台。

2.3.1.2. 資料不回傳訓練聲明，官方承諾數據不會被用於訓練其基礎模型。

2.3.1.3. 可整合公司單一登入(SSO)整合與 MFA(多重要素認證)。

2.3.1.4. 具備資料傳輸與靜態儲存之加密機制。

2.3.1.5. 確認合約終止時，供應商具備資料完全銷毀且無法復原的機制。

2.3.1.6. 智財權侵權免責（供應商是否提供 Copyright Shield 保護）。

2.3.2. 導入申請：

2.3.2.1. 各單位因業務需求導入、採購或訂閱外部 AI 系統（含 SaaS 服務）

時，須提出申請並提交「AI 系統使用暨導入評估檢核表」。

2.3.2.2. 採購第三方 AI 軟體時，須確認供應商具備合法之資料授權，且合約或服務條款中必須符合「AI 工具採用基礎原則」。

2.3.3. 資安評估：

2.3.3.1. AI 工具分級、資料儲存與傳輸風險及合規審查。

2.3.3.2. 評估供應商是否取得國際資安驗證(ISO 27001)、或 AI 管理系統標準(ISO 42001)、資料儲存位置是否合法定落地要求，以及是否支援內部身分與存取控制（如 SSO 與 RBAC 權限控管）。

2.3.3.3. 建立資安管控與防護措施，納入風險評估與監控。

2.3.4. 法遵審查：

2.3.4.1. 供應商合約審查：採購第三方 AI 軟體時，須確認供應商具備合法之資料授權，且合約或服務條款中必須明訂禁止供應商將本公司輸入之業務數據與機敏資訊用於訓練其外部基礎模型。

2.3.4.2. 確認是否涉及著作權爭議、符合國內個資、隱私權法規。

2.3.5. 核准上線：

2.3.5.1. 核准上線登錄 AI 清冊。

2.3.5.2. 核准後提報評估納入「AI 工具准駁清單」。

2.4. AI 工具分類與准駁清單 (Whitelist)

公司應建立 AI 工具使用等級分類，並每季至少更新一次准駁清單，主要分為以下三類：

2.4.1. 第一類：禁止使用（如：含有重大安全性漏洞或資料外洩風險的非法工具）。

2.4.2. 第二類：受限使用（如：ChatGPT 公開版，僅限不涉及機密資料的日常作業）。

2.4.3. 第三類：企業級應用（如：Azure OpenAI、Copilot for Enterprise。資料受到企業合約保護，可用於處理營運數據）。

2.5. 資訊安全控管規範

2.5.1. 資料輸入限制（Data Leakage Prevention）：

2.5.1.1. 絕對禁止將公司營業秘密、未公開財務數據、客戶個資及產品原始

碼等輸入至「未經驗證或未承諾資料保密」的非企業授權之 AI 系統中。

2.5.1.2. 應優先使用具備企業帳戶、加密傳輸（TLS）及資料不留存（Zero Retention）聲明的工具。

2.5.2. 人工監督與 AI 決策控制（Human-in-the-loop）：

2.5.2.1. AI 產出之結果（如程式碼、分析報告、文案或設計圖）必須經由具專業知識的人員審查後方可發布。

2.5.2.2. 代理 AI 執行高風險工作流，在實際應用於公司業務前，使用者必須進行人工覆核，確保內容準確性、無資料或系統損害的疑慮。。

2.5.2.3. 嚴禁直接將 AI 產出結果或文件作為法律文件或合約之最終版本。

2.5.3. 防範 AI 投毒與對抗性攻擊：

針對內部自建模型，應防範 AI 投毒(訓練階段)、對抗性攻擊(推論階段)，並定期進行 Prompt Injection（提示詞攻擊）弱點掃描。

2.5.4. 導入 ISO/IEC 42001：

採用目前最新的 AI 管理系統國際標準，建議將其納入公司資安管理體系（ISMS）的一部分，與 ISO 27001 並行。

2.5.5. 自動化監控：

建議在公司網路閘道端部署 DLP（資料外洩防護）系統，自動識別並攔截上傳至公開 AI 網站的敏感關鍵字（如：CONFIDENTIAL, PASSWORD, API_KEY）。

2.5.6. 建立「AI 實驗沙盒」：

提供一個安全的、公司內部管控的 AI 環境（例如自建 LLM 或 Azure OpenAI 實例），引導員工從高風險工具遷移至受控工具。

2.5.7. 持續教育訓練：

2.5.7.1. 不同角色應採分眾化訓練(如一般使用者、開發者、代理程式建置者、管理者)，應每半年定期訓練且重大風險與政策變更時加訓。

2.5.7.2. 宣導與訓練內容包含資料分類、AI 幻覺、提示詞注入、深度偽造、智財與事件通報、AI 詐騙防範與「提示詞工程安全」進行宣導。

2.6. AI 應用管理

2.6.1. AI 產出物可理解性與可接管性

AI 產生或修改之程式碼、腳本、系統設定、基礎設施定義、工作流程、規則、技術文件、操作手冊或其他可能影響正式環境之產出物，在部署、發布或正式採用前，應由具相應專業能力且責任明確之人員審查，確認其目的、邏輯、影響範圍、依賴關係、測試方式、回復方式及後續維護方式均在可解釋性與透明度之範圍內。若無合格人員可說明、維護或接管，不得部署至正式環境。

2.6.2. AI 生成程式與正式部署權限分離

AI 或 AI 代理程式得協助產生、修改程式、設定與技術文件，但不得同時擁有未經人工核准即可完成「產生/修改 - >合併 - >部署至正式環境」之完整權限鏈。AI 生成或修改之程式與設定，應依公司既有 SDLC / CI/CD 流程完成具資格人員審查、必要測試、資安檢測及上線核准；高風險或不可逆變更應採職責分離或雙人核准。

2.6.3. AI 非預期自主行為 / 未授權操作事件

發生越權存取資料、未經核准執行外部或正式環境操作、自行繞過既定控制或嘗試擴權、無法依正常機制停止、連續執行與核准使用情境顯著偏離之操作、產生或部署無人可理解 / 可維護 / 可接管之程式或設定時，視為資安事件應立即停止相關代理程式或自動化流程、撤銷必要憑證 / 權限、隔離影響範圍、保存操作軌跡並啟動事件響應。

2.6.4. AGI / 高自主性新興 AI 保留條款

本辦法之適用範圍不限於已列舉之 AI 技術。未來若出現具顯著更高通用推理、自主規劃、工具調用、自我調整或長時間自主執行能力之新興 AI (包含未來可能之通用人工智慧 AGI)，於公司評估導入或使用前，應觸發特別 AIA 及本辦法適用性檢討；在完成風險評估、控制設計及必要核准前，不得直接導入正式環境。

2.6.5. AI 產出來源與人工覆核標示

建立「AI 產出來源與人工覆核標示」制度，適用於正式文件、程式碼、系統設定、工作流程、分析結果、圖像 / 影音及其他受控業務產出物。

2.7. 法律與合規要求

2.7.1. 隱私保護：

符合《個人資料保護法》，嚴禁上傳可直接或間接識別個人之資訊，除非經過去識別化處理。

2.7.2. 智財權保護：

員工應瞭解 AI 產出之內容目前在多數法規下不享有著作權，且需注意避免產出侵害第三方版權之內容。

2.7.3. 透明度原則：

若業務產出物為 AI 生成，且與外部客戶利益相關時，應適度標註「本內容由 AI 輔助生成」。

2.8. 系統上線、監控與汰除機制

2.8.1. 上線前掃描：AI 系統正式上線前(含雲端與地端版)，需整合 AI Security 安全管控平台系統，且須安全弱點掃描，確認無惡意軟體植入後門或「提示詞注入」等高風險。

2.8.2. 軌跡監控與異常處理：針對 AI 系統的存取紀錄進行備查，使用過程或前後，若發現 AI 模型產生嚴重偏差（模型漂移）或異常結果，須立即通報以進行查修、調整或重置。

2.8.3. 安全下架：不再使用的 AI 系統或模型，管理單位須依程序下架，並確保模型權重檔案與訓練資料被銷毀至無法還原的狀態。

2.9. 風險評估與事件響應

2.9.1. 風險評估：

引進大型 AI 專案前，必須進行 AI 影響評估，分析其對組織安全的衝擊。

2.9.2. AI 風險分類：

風險類型	說明
資料外洩	輸入敏感資料至公開 AI（如 ChatGPT）
模型幻覺	AI 產生錯誤資訊
法規違規	未符合個資或 AI 法規
智財風險	生成內容侵權
資安風險	Prompt Injection / Data Poisoning
決策風險	AI 輔助決策失準

2.9.3. 事件響應與重大變更：

當發生 AI 事件應變與模型/代理程式/供應商重大變更時，須再評估風險。

2.9.4. AI 使用分級管理

風險等級	AI 使用	管控方式
Level 1 (低風險)	• 查詢公開資訊。 • 文書輔助（不含敏感資料）	可直接使用（需教育訓練）。
Level 2 (中風險)	• 使用內部資料（非敏感）。 • AI 協助分析報告。	• 單位主管核准。 • 資料去識別化。
Level 3 (高風險)	• 涉及個資 / 財務 / 機密。 • AI 自動決策(如風控、法務)。	• 資安審查 + 法遵審查。 • 建立風險評估報告（AI RA）。

2.9.5. 違規與例外處理：

2.9.5.1. 因本公司所有同仁因工作需求或是特定目的，無法遵循本辦法或相關規章，需事先經例外管理簽核流程，取得申請單位處級主管授權，與「資訊安全推動組」主管同意，始得例外處理。

2.9.5.2. 違反本辦法導致公司機密資料外洩者，經調查屬實得依公司《員工獎懲辦法》及公司相關規範辦理。