統一超商（UNI-PCSC）負責任人工智慧政策

(UNI-PCSC Responsible AI Policy)

- **版本號：** V2.45
- **核准單位：** 董事會 / 永續發展委員會

1. 承諾與願景（Commitment & Vision)

統一超商（UNI-PCSC）致力於透過數位創新提升零售服務體驗。我們承諾在人工智慧（AI）的開發、採購與應用過程中，秉持「真誠、創新、共享」的企業核心價值，確保 AI 系統旨在增進人類福祉、促進社會公平與環境永續。

2. 治理架構與問責（Governance & Accountability)

- **2.1 董事會監督：** 由「永續發展委員會」及「AI 數位轉型高階主管」負責最高監督，指定人員負責人工智慧的成果、風險決策及合規性。董事會成員承諾定期接受相關培訓，確保具備足夠專業知識監督 AI 策略。
- **2.2 責任歸屬：** 參考 ISO/IEC 42001 標準，每一項 AI 系統均定義明確的擁有者（System Owner)，設計、開發及營運人工智慧系統，以在人工智慧開發與發布的全生命週期中保護個人資料。此舉措包含遵守適

用的資料保護法規，確保在系統發生偏差或損害時，具備清晰的問責路徑，並對受人工智慧影響的決策負起法律責任。

3. AI 授權用途與行為邊界 (Authorized Use & Boundaries)

為防止 AI 系統超出預期範圍運作並確保安全，本公司實施以下措施：

- **3.1 授權範疇（Authorized Scope）：** 明確定義 AI 之功能與局限性。僅限於授權之業務範疇，嚴禁任何導致 「任務膨脹（Mission Creep）」 或非預期應用之功能擴展。
- **3.2 允許使用政策（Allowed Use Policies）：** 建立 AI 應用規範準則，透過政策與技術手段明確規範 AI 之「能做」與「不能做」事項，並轉換為具體技術限制。
- **3.3 技術護欄（Guardrails）：** 於系統層級配置技術防護層（Guardrails）與能力限制，強制過濾不當內容，防止 AI 操作超出其預定範疇。
- **3.4 系統網路安全(System Network Security)：**採取強健的安全措施以防止網路攻擊。包括滲透測試、事件應對、加密和安全編碼實踐。也可能包含用以防止不當內容串改的機制。

- **3.5 升級控制點（Control points for escalation）：** 在關鍵決策中維持人工核准並允許人類介入
- 人機協作（Human-in-the-loop, HITL）確保受 AI 影響的關鍵決策納入有效且經書面記錄的人為監督，並賦予介入、推翻或升級處理之權限與能力。
- 監督模式涵蓋多種層級：
- (1)人為主導（human-in-command，擁有最終決定權）
- (2)人為監控（human-on-the-loop，負責監控運作）
- (3)人為介入（human-in-the-loop，須經核可方可執行）。
- 系統必須確保 AI 在缺乏人為監督及人為審查或推翻機制之情況下，無法做出不可逆或高風險之決策。

4. 環境永續與資源績效 (Environmental Sustainability)

本公司認可 AI 對環境之潛在影響，採取以下 強制性措施（Mandatory Measures）降低生態足跡：

- **4.1 供應鏈強制要求（Supply Chain Mandate）： 強制要求 (Mandates)** 自有與第三方 AI 資料中心供應商，必須衡量並提供 **能源消耗（Energy use）、水資源使用（Water use）及 碳排放強度**

(Carbon intensity) 報告，並強制要求使用節能基礎設施及鼓勵使用再生能源。

- **4.2 資源優化：** 強制要求實施節能基礎設施準則，透過演算法優化降低運算負載和能源浪費。

5. 禁止事項 (Prohibited AI Practices)

本公司嚴格遵守個人資料保護法及國際隱私標準，嚴格禁止開發或使用以下具有人權風險之 AI 系統：

- **操縱性行為：** 禁止使用潛意識技術或欺騙手段扭曲人類行為。
- **弱點剝削：** 禁止利用特定群體（如兒童、身心障礙者）之弱點進行誘導或剝削。
- **社會評分 (Social Scoring)：** 禁止基於社會行為對個人進行信用或道德評分。
- **生物辨識識別：** 禁止在公共場域進行未經授權的即時遠端生物辨識識別。

6. 透明度、公平性與救濟 (Transparency, Fairness & Redress)

- **6.1 AI 標示：** 主動揭露與 AI 互動之事實、用途、能力、限制、資料來源和風險，提供可理解的理由來說明產出結果，當使用者與人工智慧而非人類互動時應該要明確告知。
- **6.2 偏見與漂移監測：** 定期偵測「模型漂移（Model Drift)」與偏見稽核（Bias Audit)，進行能確保公平和非歧視性行為的人工智慧設計，包括在資料訓練中識別和減輕偏見、執行偏見稽核，並確保創建多元化且具代表性的資料集。確保決策之公平性。
- **6.3 AI 可解釋性：**致力於提升 AI 決策的可理解性，讓利害關係人（如: 加盟主、消費者）能理解 AI 推薦背後的邏輯（例如：商品推薦原因），避免「黑箱作業」損害信任。
- **6.4 申訴與人工覆核：** 利害關係人具備申訴權，本公司提供透明之救濟程序與人工覆核路徑。