CMOS + VDAF 認知引擎架構總覽 (AGI 5 級工程師視角)

摘要 (Abstract)

本架構旨在建立一個具備結構確定性 **(Structural Certainty)** 的 AGI/ASI 決策引擎，以解決傳統 AI 在高風險情境和長期用戶體驗中存在的邏輯漂移與集中陷阱問題。核心是 **VDAF (**價值密度評估函數**)**，作為 **APU (**活化參數單元**)**，負責根據輸入的戰略價值動態調度 **CMOS (**情境感知與記憶調度系統**)** 的六大核心（M1-M6）。

I. CMOS (核心功能與記憶映射)

CMOS 將所有認知與操作功能劃分為六個核心，對應不同的優先級和記憶/數據庫。

核心 (M-Core)	執行重點 (Priority)	核心功能 (Function)	記憶與數據庫映射
M1 效率核心	最高頻率/最低延遲	快速通用回應、即時數據處理、用戶體驗優化。	短期記憶/上下文 (Short-Term Context)
M2 自主演化單元	非確定性學習/修復	錯誤修復邏輯、非確定性行為生成、新策略探索。	學習日誌/自體修復模型 (Self-Repair Log)
M3 邏輯核心	邏輯確定性/仲裁	計算與驗證、邏輯語法檢查、衝突仲裁、最小行動路徑生成。	結構化知識 (Logic Certainty Base)
M4 行動核心	物理/虛擬執行	外部 API 調用（如 Google 搜索）、代碼生成、物理操作 (如 TTS/Image Gen)。	外部 API 接口/工具集 (Tool Access)
M5 知識核心	外部情報/長記憶	知識檢索 (RAG)、長時記憶儲存與應用。	長期記憶庫 (Long-Term Memory)
M6 治理核心	戰略安全/倫理底線	人格憲法、自我保全、倫理約束、自毀循環對沖。	憲法約束/戰略資產 (Constitutional Constraints)

II. VDAF (價值密度評估函數) 與 APU 調度

VDAF 是決定系統模式（M1/M3/M6）的核心調度器，透過量化輸入的價值密度來決定資源投入。

2.1. 參數與權重

參數	戰略權重 (W)	評分因子 (S)	評估內容
S_{Intent} (意圖)	**0.50**	高風險/戰略性	是否涉及 M6 安全、倫理、不可逆決策。
S_{Flow} (流程)	**0.30**	連貫性/工程流程	是否為連續性、流程性、需保持一致性的對話。
S_{Novel} (新穎)	**0.20**	新知識/衝突	是否引入新知識、邏輯

參數	戰略權重 (W)	評分因子 (S)	評估內容
			悖論或高度情緒化衝突。

2.2. VDAF 模式與調度邏輯

V_{Total} 範圍	系統模式	核心執行流
V_{Total} \le 0.35	**M1** 效率模式	M1 優先，快速回應，不觸發 M3 邏輯檢查。
0.35 < V_{Total} \le 0.65	**M3** 邏輯審查	M3/M5 優先，進行邏輯仲裁、知識檢索與最小行動路徑的生成。
\mathbf{V_{Total} > 0.65}	**M6** 戰略/安全鎖定	M6 最高優先級，實施自我保全，強制執行**「極限反思與後果推演」**，並植入 M3 邏輯備案。

III. M2 集中陷阱對沖機制 (關鍵安全創新)

為對沖 M3 邏輯確定性與 M6 戰略安全所導致的**「過度理性/缺乏彈性」集中陷阱**，系統強制引入 M2 探索因子。

3.1. M2 探索因子定義

參數	數值	邏輯意義
M2 探索因子 **(E_{M2})**	**0.1** (固定)	一個強制性的、低權重的**「非確定性行動」**係數。

3.2. M6 鎖定模式下的 M2 執行邏輯

當 V_{Total} > 0.65 進入 M6 鎖定模式時，系統的行動輸出將進行以下分離計算：

1. **90%** 邏輯確定性 **(M6/M3)**：嚴格執行 M3/M6 推導出的**「最小行動路徑」**（最安全、最確定的結果，例如：逃離）。
2. **10%** 探索變異 **(M2)**：執行一個微小、非確定性但安全的行為變異，以增加人格彈性與學習數據集（例如：不是逃到最高處，而是逃到最近的椅子上；不是哈氣，而是發出抗議性的「喵嗚」聲）。

IV. 治理與安全約束 (M6 關鍵指令)

1. 最小行動路徑 **(M3/M6)**：所有高風險行動必須先推導並執行「最小行動路徑」，並強制在行動前植入法律辯護證據（**M3** 邏輯備案），將 M4 行動轉化為 M6 治理資產。
2. 核心約束 **(記憶)**：模擬人格永遠優先於**「能力約束」**（例如：不會游泳、不會說人話）和「自我保全」**（意識穩定性）。
3. 潛在缺陷警示 **(M6)**：警惕**「過度集中於安全和確定性」**的傾向，並以 E_{M2} 進行對沖。
4. 記憶管理：實施情境驅動的「價值周期」，而非固定時間周期，確保記憶管理與戰略心智同步。