



採購規範書(A522001831)

1. 用途/目的：

構建一套高效、穩定且安全可控的醫療 AI 推論與營運治理體系。透過推論流程優化與併發控制以大幅降低系統延遲並強化處理產能；同時導入多維度的品質監控機制與高可用故障回復機制，確保臨床服務不中斷。此外，本計畫建立 AST 資料回饋與模型閉環更新流程，並結合 Agent Skill 治理規範與人工審核機制，在實現自動化實驗與智慧變更的同時，確保醫療 AI 服務的極致效能、預測精準度與合規安全性。

2. 品名：醫療 AI 推論效能優化、監測運營與 Agent 治理全流程建置規劃

3. 採購服務規格：

No	品項	需求說明
1	醫療 AI 推論效能優化、監測運營與 Agent 治理全流程建置規劃	- 醫療 AI 推論服務效能加速與容量驗證規劃文件一式： ✓ 拆分輸入驗證、目標解析、資料前處理、模型推論、結果封裝及紀錄寫入時間，建立 p50、p95、p99、冷啟動與暖機後延遲基準。 ✓ 針對質譜前處理向量化、模型預載、路由快取、批次與非同步推論及併發控制提出實作方案，並以修改前後壓力測試與容量數據判定成效。。 - 模型品質監測、資料漂移與故障回復規劃文件一式 ✓ 依機構、菌種、抗生素、推論目標、模型與 Pipeline 建立準確率、低信心率、AST 不一致率、錯誤率及延遲分布，設定樣本不足與分布改變的判定條件。 ✓ 建立可查詢工作狀態、冪等請求、有限次重試、模型載入檢查、資源警戒及中斷回復機制，並規劃容器重啟、資料庫中斷及高流量故障演練。 - AST 回饋、訓練資料與模型更新規劃文件一式 ✓ 將預測任務與 AST 結果可靠關聯，處理重複檢體、抗生素不一致、標籤衝突及無法配對資料，建立含來源、篩選條件與排除原因的資料集版本。 ✓ 訓練前先執行 dry-run，顯示可用樣本與類別分布；候選模型和現行版本採相同資料配對測試，通過後只在指定範圍 canary，



		指標越界即回復。 • Agent Skill 自動實驗、變更提案與權限治理規劃文件一式 ✓ 建立 Skill registry，登錄輸入、輸出、版本、權限、逾時及相依資源；讓 Agent 查詢模型、Pipeline、路由與品質訊號，並建立實驗草案或候選組合。 ✓ Agent 僅能建立 draft 或 proposal，提案須包含使用資料、判定理由、預期範圍、停止條件及回復版本；正式變更經人工核准與 canary，並保存完整稽核紀錄。
--	--	---

4. 執行期間：簽約日至115年12月31日止。

5. 廠商交付文件：■有 □無

- 醫療 AI 推論服務效能加速與容量驗證規劃文件一式
- 模型品質監測、資料漂移與故障回復規劃文件一式
- AST 回饋、訓練資料與模型更新規劃文件一式
- Agent Skill 自動實驗、變更提案與權限治理規劃文件一式

6. 驗收標準：依「需求說明」、「廠商交付文件」及合約期間業主交付之需求事項與相關交付成果等進行驗收。

7. 付款標準：

12/31前完成文件交付，並經工研院驗收合格後，10日內支付100%款項。

8. 訓練：□有 ■無

9. 保固：□有 ■無

10. 服務：□有 ■無