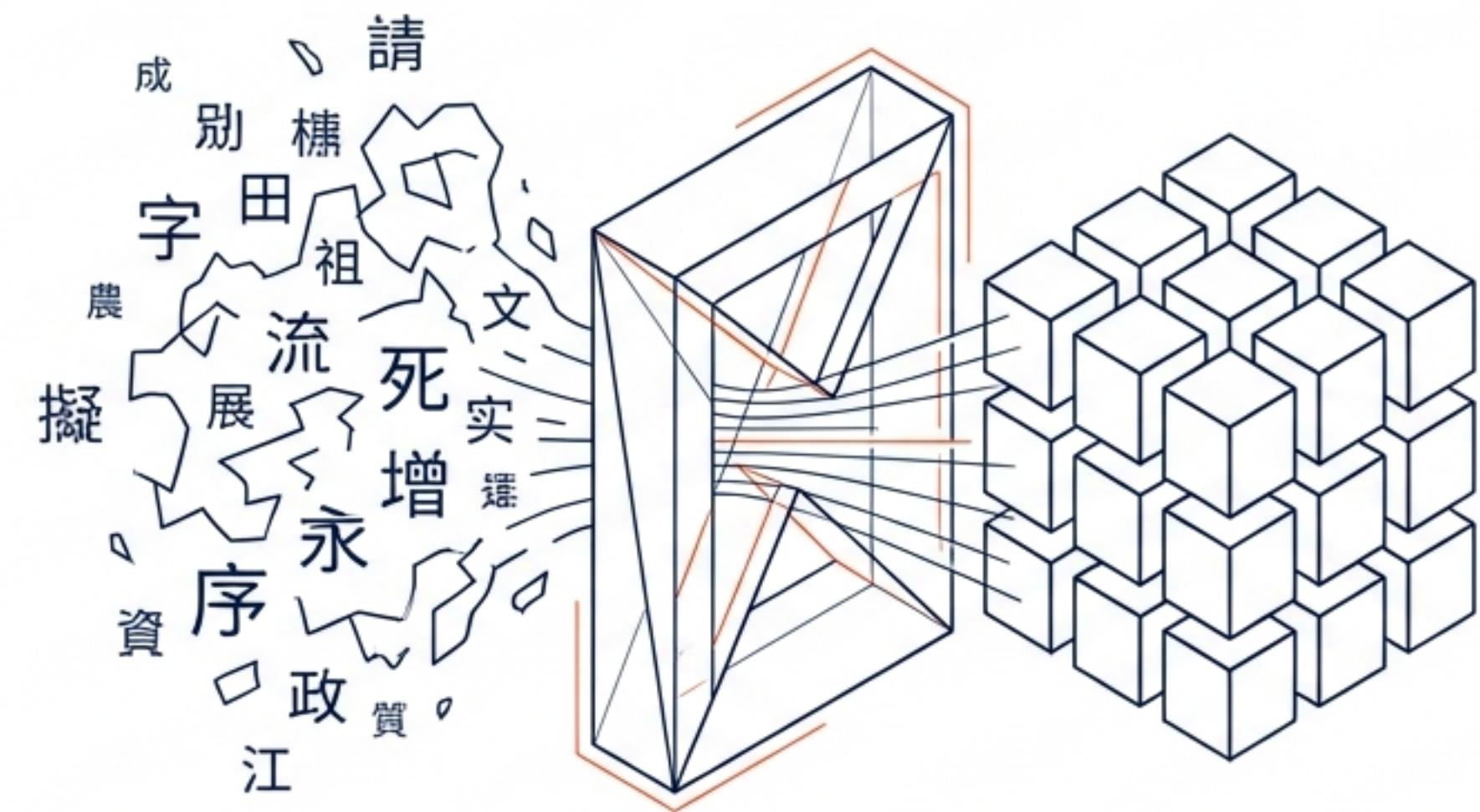


# ASD Architect & Decoder

## 方案基準評測報告

針對長文本精準檢索與事實查核的效度  
(Validity)、信度 (Reliability) 分析



# 測試證實 ASD 方案已超越「生產就緒 (Production Ready) 關鍵指標

經針對 5 萬字真實商業會議記錄進行的 54 題全量測試，ASD 方案在事實一致性與回答完整度上均優於預設標準，建議由實驗室環境轉入小規模試點部署。

**96.5%**

事實一致性  
(QA-based Factuality)

確保回答不胡說

**97.5%**

證據支持率  
(Claim Verification)

確保引用有依據

**88.9%**

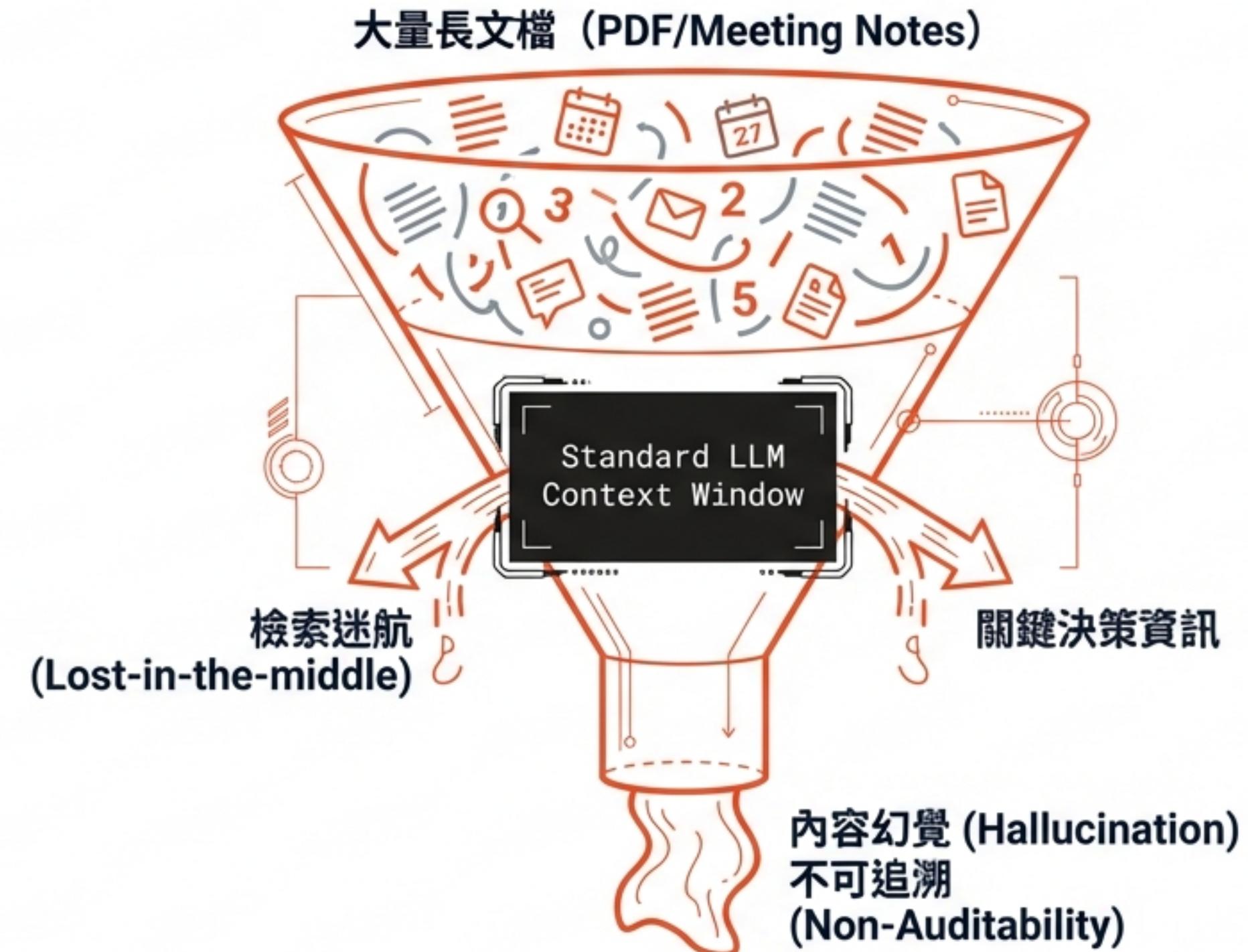
回答面向完整度  
(Facet Coverage)

超越 85% 目標，解決「答不全」痛點



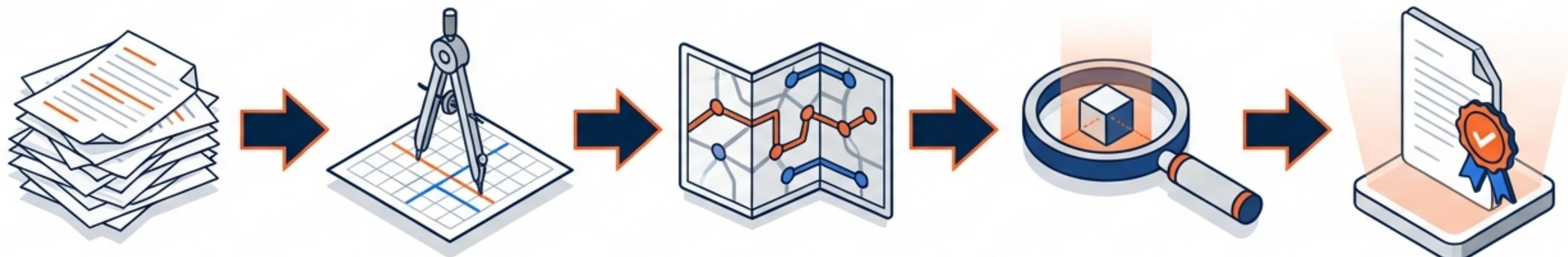
# 企業級長文本應用的核心風險：檢索迷航與內容幻覺

直接將長篇會議記錄或 PDF 丟給 LLM 處理時，面臨三大不可控風險。



# 解法：ASD (Agent-Skill Driven) 結構化封裝機制

ASD 將非結構化文檔轉化為帶索引的「精準資料庫」，而非單純摘要。



**Input:** 原始長文檔  
(Raw SSOT)

**Architect (建築師)**  
拆解為獨立模組

**Index (路由)**  
建立 Master Index 導航

**Decoder (解碼器)**  
依索引鎖定範圍 ->  
讀取 Payload

**輸出答案 + 證據包**  
(Evidence Pack)

**Fail-Closed 機制** — 系統若無法在鎖定範圍內找到答案，會直接回覆「Out of Scope」而非編造，確保零污染。

# 評測方法論：基於真實商業場景的嚴格壓力測試

## Technical Spec Sheet

### DATASET PROFILE

來源：某教育集團轉型專案會議記錄（已去識別化）

規模：約 50,000 字（涵蓋 2025 年下半年高層決策紀錄）

版本：d3e226a3

### TEST COMPLEXITY

- **54** 道測試題
- 涵蓋：
  - 事實查詢（具體數據）
  - 關係分析（因果推論）
  - 綜合判斷（跨文檔整合）

### SCORING MECHANISM

- **Gating 指標**：格式合規、JSON 解析、證據語義驗證（**必須通過**才能計分）
- **Balanced Scoring**：兼顧「精確度」與「實用性」

# Architect 效度分析：零損耗的結構化封裝

確保原始資料在轉換為機器可讀格式時，沒有資訊流失。

**99.4%**

**原文覆蓋率 (Raw Coverage Rate)**

335 個原始段落中，333 個被完整保留。

**90.0%**

**結構合規率**

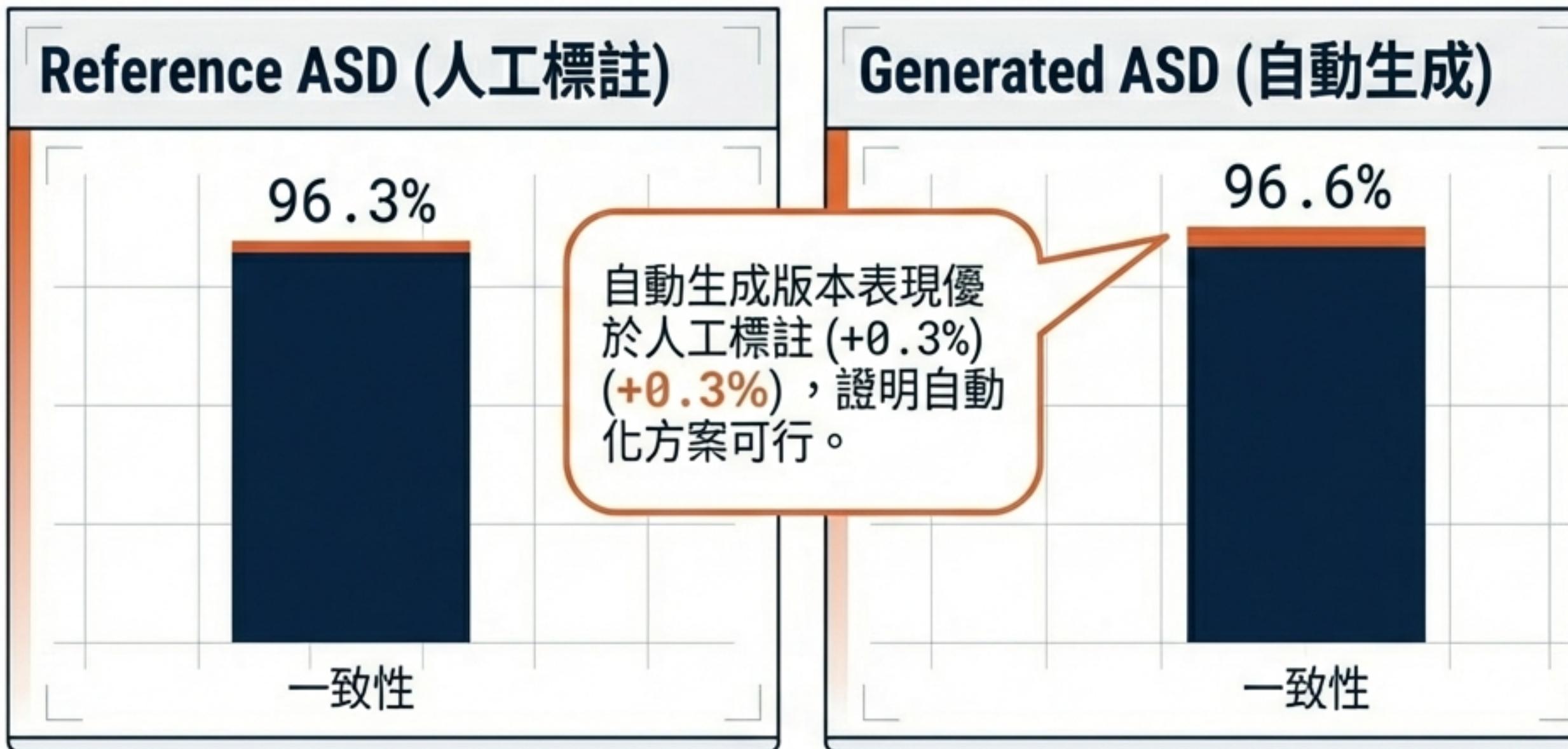
生成的模組符合嚴格的 metadata 標準。



Architect 成功將非結構化會議記錄轉化為 ASD 規範格式。

# Decoder 信度分析：高可信的事實一致性

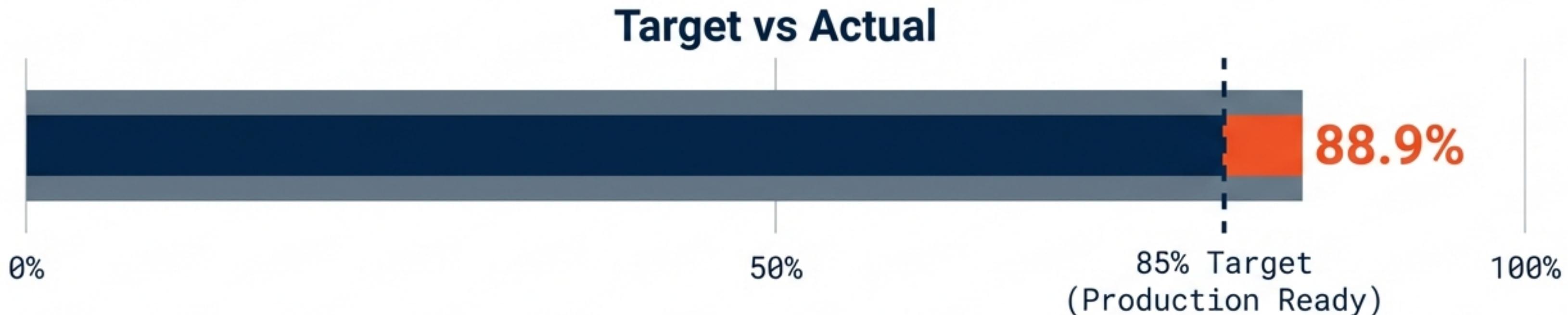
確保回答內容絕對忠於原文，無幻覺。



- **96.5%**  
QA-based Factuality  
(0 個不一致)
- **97.5%**  
Claim Verification  
(576/591 獲得支持)

## 重大突破：Facet Coverage 達成 88.9%

解決「回答不全」痛點，大幅超越預設目標。



**The Challenge:** 傳統檢索容易只回答問題的一半。

**The Solution:** Enhanced Decoder Prompt 強制 5 個面向掃描。

**The Result:** 74.1% 的問題實現了 100% 的完美面向覆蓋。

## 證據鏈技術 (The Evidence Chain)

每一句回答都有據可查，建立 AI 信任。

### Code Editor (Audit Log)

```
1 {  
2   "Evidence_Pack": {  
3     "Claim": "公司計劃轉型為教育生態系統",  
4     "Source_Module": "[Module_20250722]",  
5     "Original_Quote": "...核心願景是建立一個完整的教育生態系統...",  
6     "Verdict": "SUPPORTED"  
7   }  
8 }
```

精準錨定 (Anchor)

逐句證據 (Verbatim)

審計結論

## 實戰案例 I：複雜商業邏輯的精準提取

場景：分析某教育集團轉型計畫，跨越 7 月、11 月、12 月會議記錄。

### Question: 如何解決資金池雙重計算問題？



## 實戰案例 II：邏輯陷阱與風險事實查核

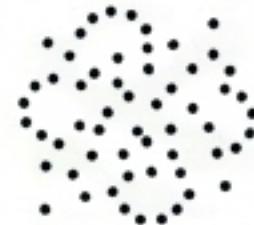
Question: FY24/25 的 +118% 增長是否有風險？



系統精準指出增長數據來自 Module A，但風險提示實際位於 Module B。成功反駁錯誤的因果連結。

Data source: Appendix Claim Verification Ex 7

## 為什麼選擇 ASD？與傳統 RAG 的架構對比



### 傳統 RAG (Traditional)

**機制：**向量相似度搜尋 (Vector Similarity)  
**弱點：**容易迷失 (Lost-in-the-middle)  
**結果：**引用模糊，難以審計



### ASD 方案 (Agent-Skill Driven)

**機制：**索引優先 (Index First) + 模組化路由  
**優勢：**精準錨定，Fail-Closed  
**結果：**96.5% 事實一致性

**Architect 穩定性說明：**雖然兩次運行有 Hash 差異 (17%)，但語義層面一致性極高。

# 當前限制與技術挑戰 (Limitations)

誠實報告當前限制，以建立真實的改進路線圖。

## 1. Gating 通過率 (79.6%)



距離 90% 理想值仍有空間，受嚴格格式檢查影響。

## 2. Forbidden Patterns 誤判



約 9.3% 案例受影響 (如將合法的 'TBD' 標記誤判為模板洩漏)。

## 3. 成本與資源



結構化轉換 (Architect) 需消耗較多 Token 計算資源。

## 4. 語料依賴性



目前驗證集中於教育科技領域，跨領域適用性需測試。

# 部署建議：啟動小規模試點 (Pilot Deployment)

基於當前技術限制的漸進式實施策略，  
以降低風險並優化價值。

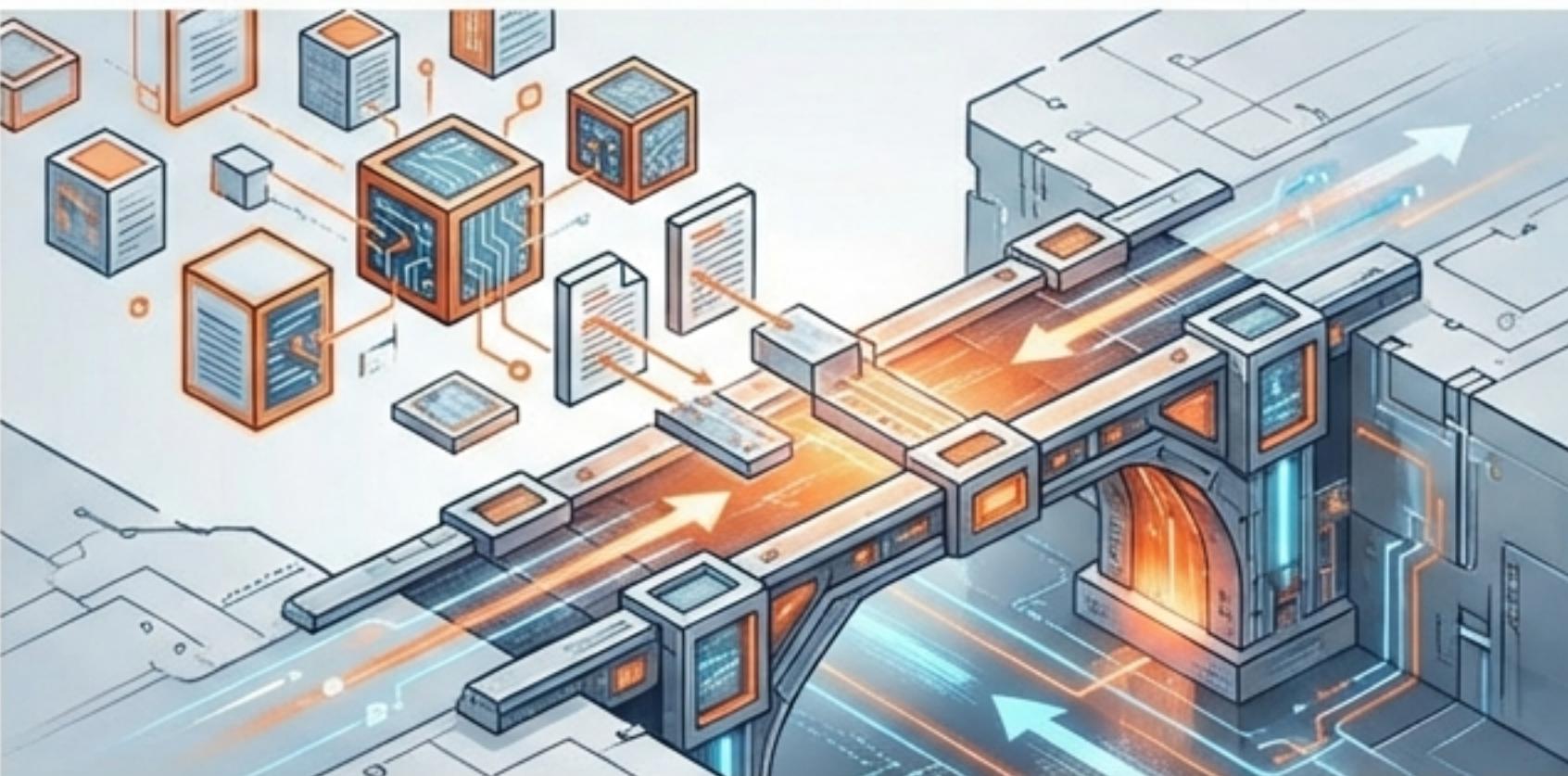


- 部署於高風險/高價值文檔  
(如董事會紀錄)
- 僅供核心團隊使用

- 監控 Facet Coverage 與  
Gating 指標
- 優化 Forbidden Patterns  
誤判率

- 開放給更多業務單位
- 自助式查詢

# 將非結構化文檔，轉化為可信賴的決策資產



ASD 不僅是一個檢索工具，更是企業資料治理（Data Governance）的解決方案。它讓沈睡在 PDF 與會議記錄中的知識，變得可見、可查、可信。

## 建議批准進入 Phase 1 試點部署。