

MGFD Design

version 0.1 2025/08/11

前言：

這個引導式漏斗框架 (Multiturn Guided Funnel Framework) 透過從根本上改變對話式 AI 的設計理念，從而被動回答問題轉變為主動引導用戶達成特定商業目標，例如銷售和客戶服務中的有價值的成果。

以下是此框架如何轉變現有對話式 AI 以實現商業目標的幾個關鍵方面：

- 從「拉取」模型轉變為「推送」模型 (**Shift from "Pull" to "Push" Model**):

- 傳統的 Retrieval-Augmented Generation (RAG) 應用通常採用「拉取」模型，即等待用戶查詢並檢索相關答案，用戶完全控制對話方向。
- 而引導式漏斗框架則採用「推送」模型，代理會主動將對話導向預定目標，例如「引導潛在客戶」、「發起互動」和「培育潛在客戶」。這要求更複雜的架構來維持目標、制定策略並根據用戶反應進行調整。

- 複製傳統銷售漏斗流程 (**Replicating the Sales Funnel**):

- 「對話漏斗」將傳統的行銷和銷售漏斗階段（如認知、興趣、評估、參與、行動和保留）直接轉換到自動化對話領域。
- 對話代理不再是被動的資訊儲存庫，而是主動的策略指南，系統性地引導用戶經歷這些階段。例如，透過歡迎用戶建立認知，提出問題激發興趣，提供資訊進行評估，提供促銷進行參與，直接在聊天中促成行動，並在購買後提供支援以促進保留。

- **實現顯著商業優勢 (Yielding Significant Business Advantages):**
 - 實施此漏斗模型帶來包括提高潛在客戶轉換率、改善銷售效率和更高的客戶參與度等顯著優勢。
 - 透過自動化資格審查和培育流程，銷售團隊可以專注於高價值活動，從而加快交易完成並大幅節省成本。
- **核心架構元件支持目標導向 (Core Architectural Components for Goal-Oriented Design):**
 - **對話管理器 (Dialogue Manager, DM):** 作為系統的「認知核心」，它管理對話狀態並控制對話流程。DM 包含對話狀態追蹤 (**Dialogue State Tracking, DST**)，用於維護對話的結構化記憶，以及對話控制 (**Dialogue Control**)，用於基於當前狀態決定下一步行動，例如提問、推薦或升級給人類代理。
 - **「思考，然後行動」循環 ("Think, Then Act" Cycle):** 這是該框架的核心實現，將 DM 的功能分解為兩個由 LLM 驅動的步驟，以確保可靠性和可控性。
 - **「思考」階段 (Think Phase) (prompt-lvl1):** LLM 作為推理引擎，分析對話狀態並決定下一步的策略性行動，其輸出是結構化指令 (如 JSON 物件)，而非用戶可見的文本。
 - **「行動」階段 (Act Phase) (prompt-lvl2):** 另一個 LLM 調用將「思考」階段的結構化指令轉化為自然、引人入勝且面向用戶的回應。這種分離限制了 LLM 在每個步驟的任務，減少了其「幻想」或偏離商業邏輯的風險。
 - **槽位填充 (Slot Filling):** 這是推動對話漏斗前進的主要機械過程，用於收集完成任務所需的特定資訊。每個填滿的槽位都代表用

戶深入漏斗一步，將通用查詢轉化為具體的偏好。系統會主動識別並生成針對性問題來引導用戶填寫缺失的必需槽位。

- **高級考量與生產準備 (Advanced Considerations for Production Readiness):**

- 優雅的錯誤處理和對話修復：系統設計用於處理模糊輸入、離題、用戶挫敗或打字錯誤。透過重新表述問題、提供多選選項或在無法解決時升級給人類代理，確保了流暢的用戶體驗並防止負面客戶體驗。
- 使用 **RAG** 增強建議：透過將用戶偏好合成為語義查詢，從向量資料庫中檢索最相關的產品文檔，並將其注入到推薦提示中，LLM 可以生成更豐富、更具說服力且事實更準確的個性化推薦，同時減少幻覺的風險。
- 持久化狀態實現長期記憶和個性化 (**Persisting State for Long-Term Memory and Personalization**): 儘管 LLM 本質上是無狀態的，但透過將 `DialogueState` 對象保存到外部資料庫並在用戶返回時載入，代理可以實現長期記憶。這使得個性化問候、語境連續性、主動推薦和建立用戶信任與忠誠度成為可能，這是一個成功的對話漏斗的關鍵目標。

根據所提供的來源及我們之間的對話，多輪引導漏斗框架 (Multiturn Guided Funnel Framework) 的實作成功可能性非常高，因為它從根本上改變了對話式 AI 的設計理念，使其從被動響應轉變為主動引導用戶達成明確的商業目標。這種轉變旨在解決傳統對話式 AI 在商業應用中的局限性，並提供一套結構化、可預測且強健的解決方案。

以下是評估其成功可能性的關鍵因素：

核心架構的優勢與穩固性

1. 從「拉取」轉向「推送」模式：傳統 RAG 應用是被動等待用戶查詢的「拉取」模型，用戶完全掌控對話方向。而此框架採用「推送」模型，代理會主動將對話導向預定目標，例如「引導潛在客戶」、「發起互動」和「培育潛在客戶」。這種主動性是實現商業目標的關鍵。
2. 複製銷售漏斗流程：框架將傳統行銷和銷售漏斗的階段（認知、興趣、評估、參與、行動、保留）轉化為對話流程。對話代理不再是資訊儲存庫，而是主動的策略指南，系統性地引導用戶經歷這些階段。這直接對應了銷售和客戶服務的商業需求。
3. **「Think, Then Act」循環**：這是框架的核心，透過兩層的 LLM 調用實現了對話管理器 (DM) 的認知功能，顯著提升可靠性和可控性。
 - **「Think」階段 (prompt-lvl1)**：LLM 作為推理引擎，分析對話狀態並決定下一步的策略性行動，輸出為結構化指令（如 JSON 物件），而非用戶可見文本。這大幅減少了 LLM「幻想」或偏離商業邏輯的風險。
 - **「Act」階段 (prompt-lvl2)**：另一個 LLM 調用將「Think」階段的結構化指令轉化為自然、引人入勝且面向用戶的回應。這種分離限制了 LLM 在每個步驟的任務，提高了準確性和可靠性。
4. **Dialogue Manager (DM) 與 Dialogue State Tracking (DST)**：DM 作為系統的「認知核心」，負責管理對話狀態並控制對話流程。DST 則負責維護對話狀態的結構化記憶，包括用戶意圖、已收集的資訊和對話歷史，這對於生成連貫且相關的回應至關重要。這確保了對話的連貫性和目標導向性。
5. **Slot Filling (槽位填充)**：這是推動對話漏斗前進的「主要機械過程」。它用於從用戶那裡收集完成任務所需的特定資訊，每個填滿的槽位都代表用戶深入漏斗一步，將通用查詢轉化為具體的偏好。系統會主動識別並生成針對性問題來引導用戶填寫缺失的必需槽位。

6. 技術堆疊的適用性：框架建議使用 **Python**、**LangChain** 和 **LangGraph**。

- **LangChain** 提供 LLM 互動所需的基本構件，如 Chat Models、Prompt Templates 和 Memory。
- **LangGraph** 更是關鍵，它允許開發者透過圖形表示來構建有狀態、多代理的應用。其狀態圖、節點與邊緣、條件邊緣和循環功能，完美契合了引導漏斗框架中複雜、有狀態和循環的對話邏輯，能夠有效地實作「Think, Then Act」循環。

支援生產就緒的進階考量

1. 優雅的錯誤處理和對話修復：框架設計考慮了用戶輸入模糊、離題或表達挫敗等情況。透過重新表述問題、提供多選選項或在無法解決時升級給人類代理，確保了流暢的用戶體驗並防止負面客戶體驗。
2. 使用 **RAG** 增強建議：透過將用戶偏好合成為語義查詢，從向量資料庫中檢索最相關的產品文檔，並將其注入到推薦提示中，LLM 可以生成更豐富、更具說服力且事實更準確的個性化推薦，同時減少幻覺的風險。
3. 持久化狀態實現長期記憶和個人化：透過將 **DialogueState** 物件保存到外部資料庫並在用戶返回時載入，代理可以實現長期記憶。這使得個人化問候、語境連續性、主動推薦以及建立用戶信任與忠誠度成為可能，這是成功的對話漏斗的關鍵目標。

商業優勢

實施此漏斗模型可帶來顯著的商業優勢，包括提高潛在客戶轉換率、改善銷售效率和更高的客戶參與度。透過自動化資格審查和培育流程，銷售團隊可以專注於高價值活動，從而加快交易完成並大幅節省成本。這直接證明了其商業價值和成功潛力。

結論

綜合來看，多輪引導漏斗框架不僅提供了一個清晰且理論基礎堅實的對話式 AI 設計理念，更提供了一套完整的實作藍圖，涵蓋了從核心邏輯（Dialogue Manager、「Think, Then Act」循環、Slot Filling）到生產就緒考量（錯誤處理、RAG、長期記憶）的方方面面。

儘管任何複雜系統的實作都會面臨挑戰，例如需要仔細的對話設計、精確的 Slot Schema 定義、高品質的知識庫 以及熟練的開發團隊來利用 LangChain 和 LangGraph 的功能，但框架預先考慮並提供了解決這些問題的策略。因此，如果能夠投入足夠的資源和專業知識進行嚴謹的規劃與實作，多輪引導漏斗框架實作成功的可能性非常高，並且能夠有效且策略性地幫助企業達成其商業目標。

Chat-Turn概念: A turn of chat of user and LLM

[user query] → prompt-lvl1 → LLM → response-1 → (to user)

[user query] → prompt-lvl2 → LLM

1. 核心模組

a. Dialogue State Management

b. Slot Filling

c. Two Level Prompting: Think, then Act

i. Think

ii. Act

d. (optional not implement now) Long Term Memory

e. (optional not implement now)

Recommendation System Concepts Integration