

I. Luồng workflow AI + Metadata

1. Webhook nhận API

- Nhận dữ liệu đầu vào qua webhook.
- **Kiểm tra điều kiện (if/else):**
 - Loại bỏ spam.
 - Xác minh dữ liệu hợp lệ, đúng nguồn.

2. Xác định Metadata

- Gửi prompt để AI xác định **metadata chính (category)** cho dữ liệu.
- AI xử lý và trả về **metadata phù hợp**.
- Extract metadata → đưa vào **switch case** → phân luồng xử lý.

3. Phân luồng xử lý

Luồng 1: AI trò chuyện cơ bản

- Đưa prompt chứa **ngữ cảnh**.
- AI phản hồi.
- Trả về **API JSON**.
- Lưu kết quả:

- Ghi vào **database**, hoặc
 - Gửi qua **service text-to-speech (TTS)** → tạo file audio.
-

Luồng 2: AI RAG Chatbot

- Đưa prompt chứa **ngữ cảnh + dữ liệu ANN** (tìm kiếm gần đúng) dựa trên metadata + câu hỏi.
 - AI phản hồi.
 - Trả về **API JSON**.
 - Lưu kết quả:
 - Ghi vào **database**, hoặc
 - Gửi qua **service TTS** → tạo file audio.
-

II. Logic dữ liệu & metadata

- **Dữ liệu gốc phải được clean** (lọc trùng, loại bỏ rác).
 - Gắn **metadata chuẩn** → Đưa lên **Google Sheet** → Hệ thống tự động gắn metadata → Bắt đầu workflow.
-

III. Lưu ý quan trọng

- **Dữ liệu càng nhiều, logic càng phức tạp.**
 - Dữ liệu ít: Có thể không cần metadata chi tiết.
 - Dữ liệu lớn: Phải có **metadata & sub-metadata**, có giới hạn về **chi phí & thời gian**.
- **Ví dụ:**
 - 1 triệu rows → Từ khoá trùng lặp nhiều.
 - Metadata dễ **chồng lấn nghĩa**, gây khó cho AI:
 - **Tính năng (Feature) vs Công dụng (Benefit/Use):** Ít dữ liệu thì dễ phân biệt, nhiều thì AI (và cả con người) dễ nhầm.

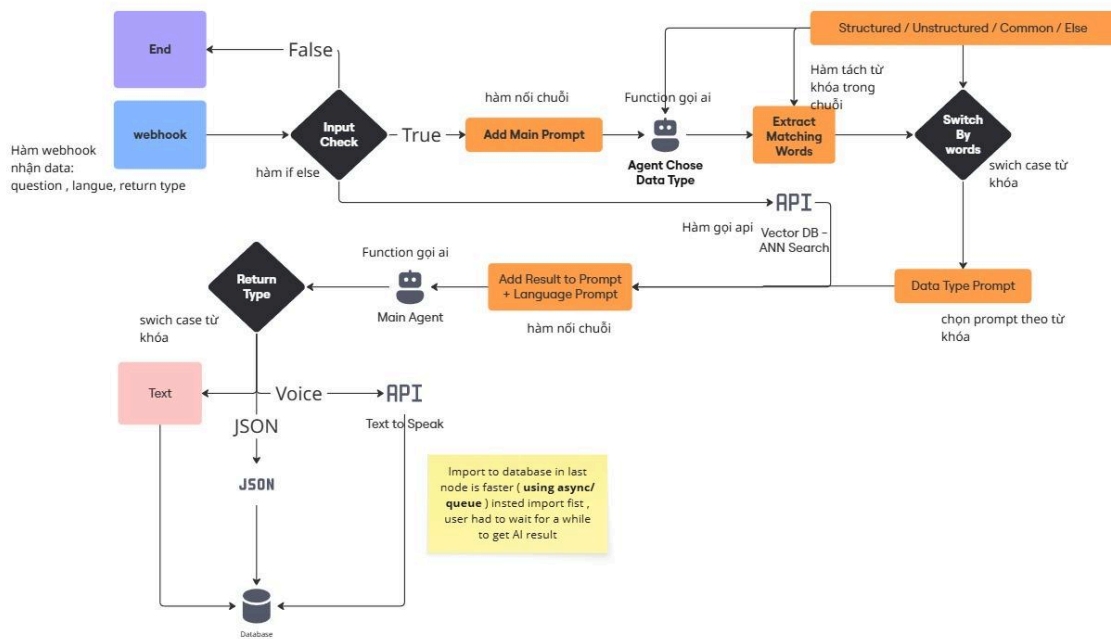
- Gắn metadata sơ sài:
 - VD: 1000 rows “Security” nhưng không rõ: Security cho **data/permission/firewall/gateway/request spam** → AI dễ hiểu nhầm.
-

IV. Các rủi ro nếu logic không chặt

- ANN tìm kiếm:
 - Metadata không rõ → ANN không bao quát đủ → Miss thông tin.
 - ANN scope nhỏ (ví dụ tối đa 100 rows) → Dữ liệu không đủ → AI trả lời thiếu.
 - ANN scope lớn → Tốn phí AI → AI bị loãng dữ liệu → Tăng rủi ro **bị/sai sót**.
 - Quá nhiều dữ liệu → ANN trả về khối lượng lớn → AI tiêu tốn chi phí xử lý → Dữ liệu bị nhiễu → Câu trả lời sai.
-

V. Kết luận

- **Clean data trước** → **Gắn metadata kỹ** → **Có quy tắc chuẩn hóa** → **Tách rõ các nhánh metadata để nhầm lẫn**.
- Thiết kế ANN & prompt phù hợp **quy mô dữ liệu**, tránh lãng phí chi phí & rủi ro **AI bị**.



Hình ảnh workflow được đơn giản hóa