Date: 9.19.2025

AI-Driven Health Chatbot implemented using RAG Model and WhatsApp Integration

Contents

Transforming Enterprises for Future

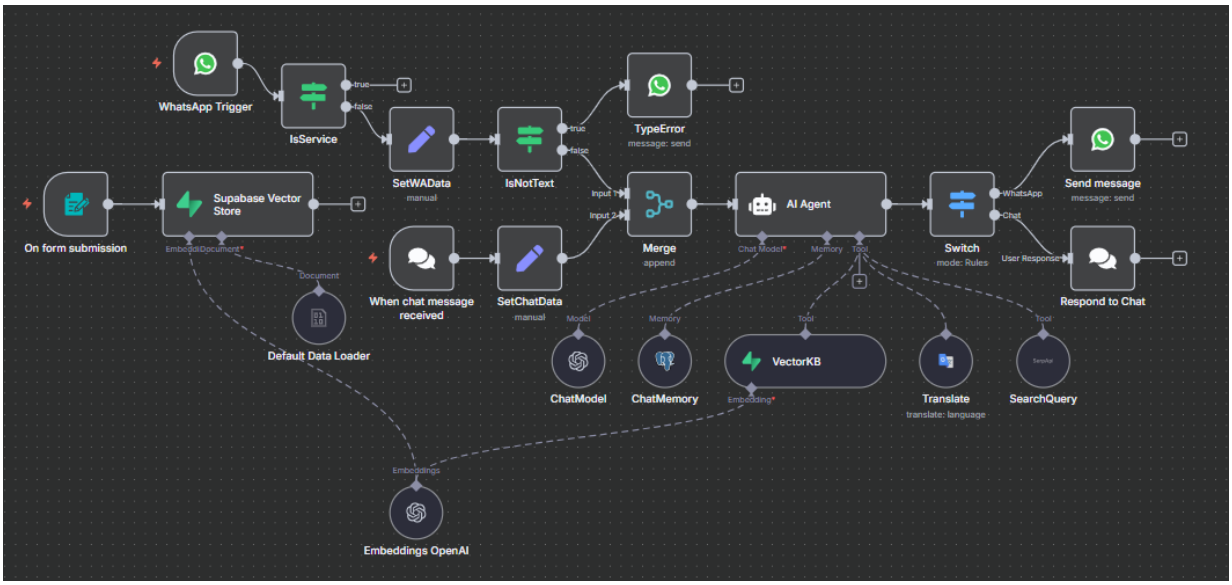
1. Objective

Build an AI chatbot that enables users—especially from rural/semi-urban areas—to ask health-related questions in **multiple Indian languages**, and receive accurate, general responses based on **document embeddings** via a Retrieval-Augmented Generation (RAG) pipeline.

2. Final Capabilities

Area	Description
Chat Interfaces	Supports both WhatsApp and web chat
Multilingual Input	Input supported in English + major Indian languages
Language Translation	Auto-detects and translates non-English input
Knowledge Base	Uses scraped health documents stored in a Supabase vector DB
Context Memory	Remembers context via Postgres Chat Memory
Document Loading	Scraped, cleaned, and chunked content from 1mg.com/diseases
Chat Model	OpenAI GPT for response generation
Tool Prioritization	Vector Search → Internal KnowledgeBase fallback

3. Final n8n Workflow Architecture



Breakdown:

Section	Description
Input Triggers	WhatsApp Trigger via Meta API or Web Chat
Pre-Processing	Filter non-user messages (IsService), check for unsupported types (IsNotText)
Data Routing	Set chat or WhatsApp data context using SetWAData / SetChatData
Merge Input	Unified path to AI Agent
AI Agent Node	Connected to ChatModel, ChatMemory, VectorKB, and Translate tools
Response Routing	Switch node checks source (WhatsApp or Chat) to send back response

4. AI Agent System Prompt Summary

The AI agent was configured with the following constraints and behavioral logic (adjusted to match **actual implementation**):

a. Goals Implemented

- Educate users on **preventive healthcare**
- Help identify **early symptoms of common diseases**
- Support **multilingual, culturally aware responses**

b. Language Handling

- Supports English, Hindi, Marathi, Odia, Telugu, Tamil, Bengali, and others
- Uses Google Translate API in n8n workflow
- Greeting override: greetings are always responded to in **English**
- Translations are used only when needed for downstream tools (like vector search)
- Maintains original input for consistency in reply language

c. Tool Usage Logic

- **VectorKB**: Always queried first using English-translated text (if needed)
- **Translate**: Only used when message isn't in English
- **InternalKnowledgeBase**: Used if no relevant vectors are returned
- **SearchQuery**: Used to query Google Search Engine

d. Ethics & Safety

- No diagnoses or treatments suggested
- Asks for clarification when unsure
- Reminds users to consult qualified doctors for medical concerns
- No personal/sensitive data collected unnecessarily

5. 3-Day Development Plan & Execution

Day	Tasks
Day 1	Explored disease content on 1mg.com
	Scraped sample (Bipolar Disorder) page
	Built initial n8n flow for uploading text into Supabase
Day 2	Crawled full 1mg disease list
	Cleaned & saved each disease to .txt
	Set up Supabase DB with pg_vector support
	Embedded text using OpenAI and uploaded
Day 3	Finalized RAG workflow with AI Agent in n8n
	Added translation step
	Configured WhatsApp via Meta App
	Added ChatMemory and error handling
	Routed outputs via Switch (Chat or WhatsApp)

6. Tech Stack

Component	Tool / Tech
Workflow Automation	n8n
LLM	OpenAI Chat Model (GPT-4.1-mini)
Embedding Model	OpenAI Embeddings
Vector Store	Supabase (pg_vector)
Memory	Postgres ChatMemory (via SupaBase)
Language Translation	Google Translate (via n8n)
Search Tool	SerperAPI (via n8n)
Input Channels	WhatsApp, n8n Chat Node
Scraping Tools	Python + BeautifulSoup

7. Future Scope

Feature	Description
Vaccination Schedules	Integrate with MoHFW API for schedule lookup
Outbreak Alerts	Use public data sources for real-time notifications
Government Scheme Checker	Lookup based on Aadhaar, ration card, etc.
Location-Specific Intelligence	Use device location or manual input
Analytics Dashboard	User queries, region heatmaps, FAQ tracking
Voice / IVR Bot	For accessibility in low-literacy regions
SMS Channel	Support feature phone users

8. Summary

Successfully built a **multilingual, document-aware health assistant**, powered by:

- Document scraping and vector embedding
- Retrieval-augmented generation (RAG)
- Contextual memory
- Multi-platform delivery (Chat + WhatsApp)
- Dynamic translation with fallback handling

Delivered as a cloud-scalable solution targeting health tech pilots and awareness tools.

9. Demo

Here is the link to the demo: https://www.youtube.com/watch?v=J2RtYsqWYdw