Trusted Data Sources for MediChat

1. openFDA API

Source: U.S. Food and Drug Administration (FDA)

What it provides: Data on drugs, side effects, recalls, etc.

Use case: Show medicine details, dosages, adverse effects, brand/generic info.

Website: https://open.fda.gov/apis/

2. NIH / MedlinePlus

Source: National Institutes of Health (NIH)

What it provides: Home remedies, disease info, symptoms, first aid

Website: https://medlineplus.gov/

3. World Health Organization (WHO)

Trusted for: Global health advice, precautions, disease outbreaks

Use case: Medical awareness, prevention tips, vaccine info

Website: https://www.who.int/

4. Drugs.com API

Note: Paid API or scrape with permission

Use case: Drug interactions, dosage guides, home remedies

Website: https://www.drugs.com/

5. Healthline / WebMD

Use: For inspiration and rewriting (no direct copy)

Trusted Data Sources for MediChat

Content: Home remedies, disease guides, wellness tips

Websites: https://www.healthline.com/ and https://www.webmd.com/

Best Practice: RAG Architecture

1. Fetch documents (drug/home remedy info)

2. Chunk them

3. Convert to vectors (LangChain + OpenAI + ChromaDB/Pinecone)

4. When user queries, fetch relevant chunks send to LLM

Suggested Datasets

DailyMed Drug Data - https://dailymed.nlm.nih.gov/

Home Remedies Dataset - Search on https://kaggle.com

WHO Health Topics - https://www.who.int/health-topics

DrugBank (Commercial License) - https://go.drugbank.com/

Legal & Ethical Use

Don't scrape protected content without permission

Always credit sources

Use data with open licenses (like CC BY 4.0)