MediBot – Retrieval-Augmented Generation AI for Trusted Medical Answers

Netheeswaran A - 24204827
Vimarish K M - 24229318
ACM40960 - Projects in Maths Modelling

Why it matters

Quick medical answers are easy to find but accurate ones aren't. Many chatbots can chat smoothly but still "make things up", which is risky in healthcare. MediBot is different, it is built on expert medical textbooks and only responds when it has verified facts. If it can't, it says so and pulls trusted results from a web search. Every step is built for safety, transparency, and trust.

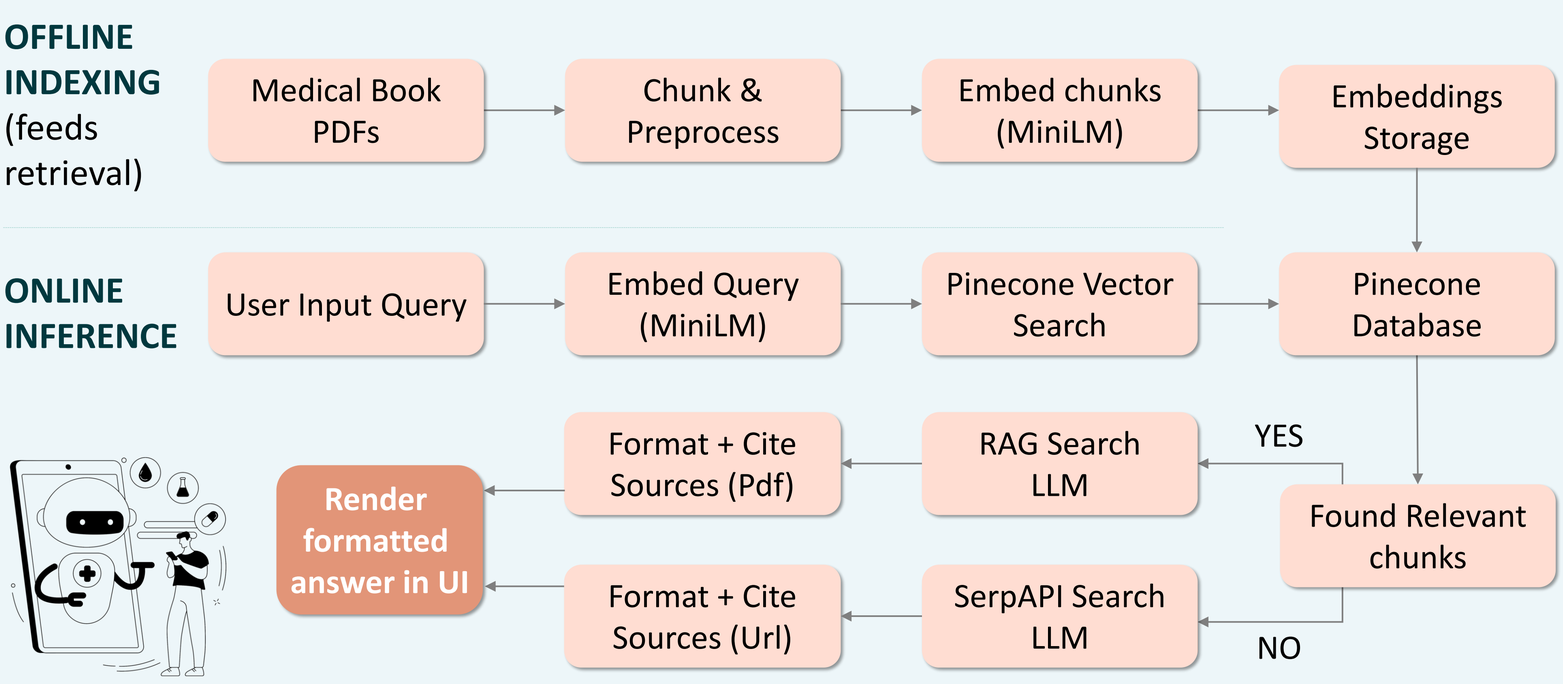
Our Goal

Most medical chatbots are either rigid rule-followers like Ada Health, or open-ended models like ChatGPT that can hallucinate. The first can't handle complex questions; the second can give confident but wrong answers. We set out to combine flexibility with accuracy, a chatbot that only speaks when it's sure and clearly labels its sources.

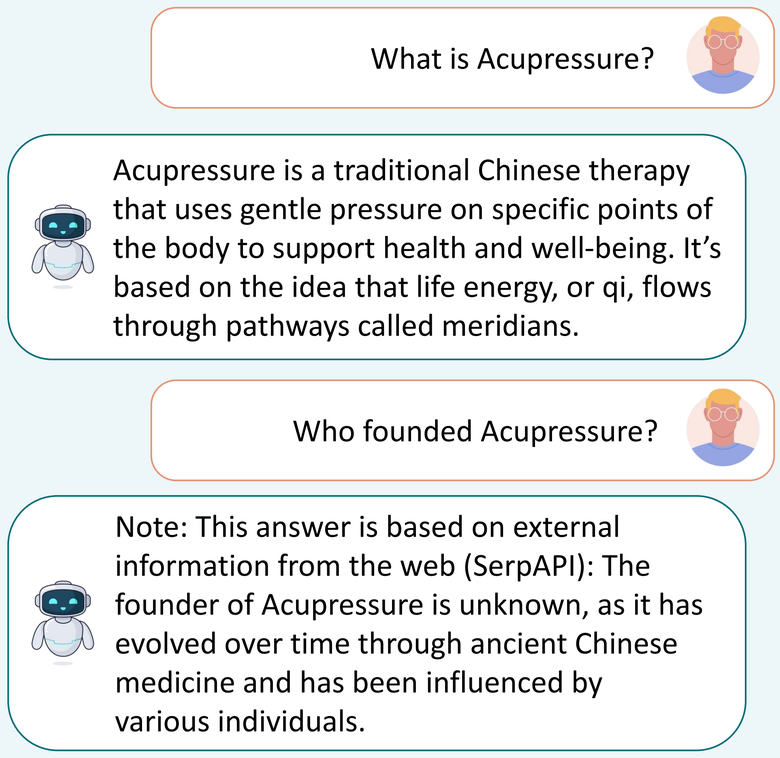
What is the gap?

People need reliable health information faster. Old chatbots were too limited and new ones can talk well but don't always get it right. MediBot blends natural conversation with evidence-based answers. By grounding an AI model in expert-written textbooks using RAG, it delivers not just quick responses, but ones you can trust.

How it works



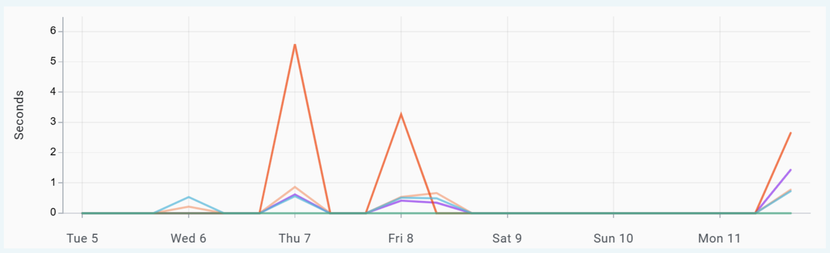
Sample chat

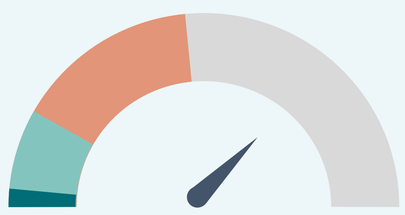


How fast and accurate it is?

Median latency by run name

Media run latency over time.





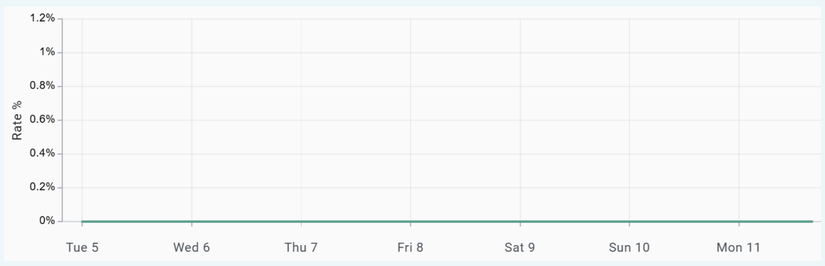
Total response latency **~100 ms**

Retrieval time **~20 ms**

LLM Inference Time **~3 ms**

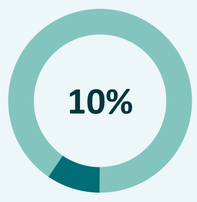
Error rate by run name

Run error rate over time.

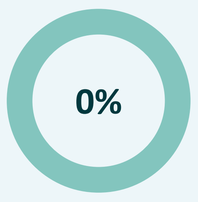


Internal Query Accuracy

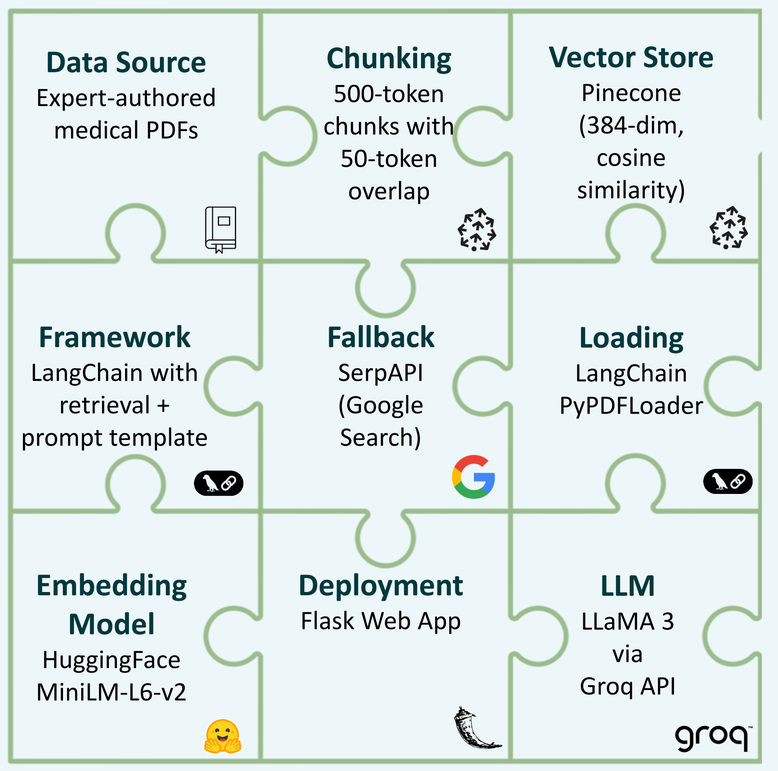
Fallback Use rate



Hallucination rate



Key components



Results

Tested on over **50 medical questions** from textbook chapters

All internally sourced answers matched their **original references**

When no answer was available, the **fallback** correctly triggered SerpAPI

No hallucinations observed throughout testing

LLM inference remained **ultra-fast** even under multiple runs

Designed for safety and trust

- Aligned with **the EU AI Act (2024)** as a high-risk AI system
- Designed for transparency, safety, and non-diagnostic use
- Complies with **GDPR** and **HIPAA** (no personal data used)
- Adheres to **the Hippocrates AI framework** (human-first design)

Built for?

- Student Support
- Patient Clarity
- Hospital FAQs
- NGO Assistance
- Offline Access

What's next?

- Global Languages
- Voice Support
- Resource Expansion
- Doctor Co-Pilot
- EHR Integration

Credits

This project shows how AI can support, not replace, health professionals by delivering accurate, explainable, and safe medical information. With clear rules, trusted sources, and transparent fallbacks, our chatbot advances ethical, effective healthcare AI.

Reference

1. Lewis et al. (2020) – RAG for factual grounding
2. Singhal et al. (2023) – Med-PaLM
3. EU AI Act (2024) – High-risk AI policy
4. Hippocrates (2023) – Ethical design in AI
5. Yang et al. (2024) – LLM eHealth Chatbot
6. Laranjo et al. (2020) – Chatbot design principles
7. LangChain Docs
8. Pinecone Docs
9. Groq API Docs
10. SerpAPI Docs

"MediBot, bringing accuracy and trust to healthcare AI."

GitHub Repo

MediBot