

Enhancing Patient Engagement in Hip Prosthesis Surgery: A Conversational Agent Using NLP-Based Question Answering and Self-Reporting Tools

Gogniat Saengrawi

Master-Thesis

External Partner: Prof. Dr. med. Reinhard Elke , Merian Iselin Clinic
Expert: Dr. Christoph Pimmer, Swiss Tropical and Public Health
Supervisor: Prof. Dr. Andreas Martin



Introduction

Recent advances in NLP and large language models (LLMs) have enabled conversation agents to deliver human-like conversations. However, in healthcare, challenges such as outdated data, lack of response control, and hallucinations limit their safe use.

This project develops a German-language conversational agent to support patients undergoing hip prosthesis surgery by answering questions and collecting health status data (EQ-5D). Using a lightweight Retrieval-Augmented Generation (RAG) approach, the system ensures more accurate, reliable, and context-specific responses while remaining resource-efficient.

Research Focus and Objectives

General RQ:

It is possible to develop a conversational agent that provides accurate and reliable answers in German to patient inquiries about hip replacement surgery while collecting and analysing patient-reported health status data to enhance postoperative care and patient engagement.

Main objective:

To develop an effective medical chatbot in German that answers hip surgery-related questions and facilitates patient-reported health status collection.

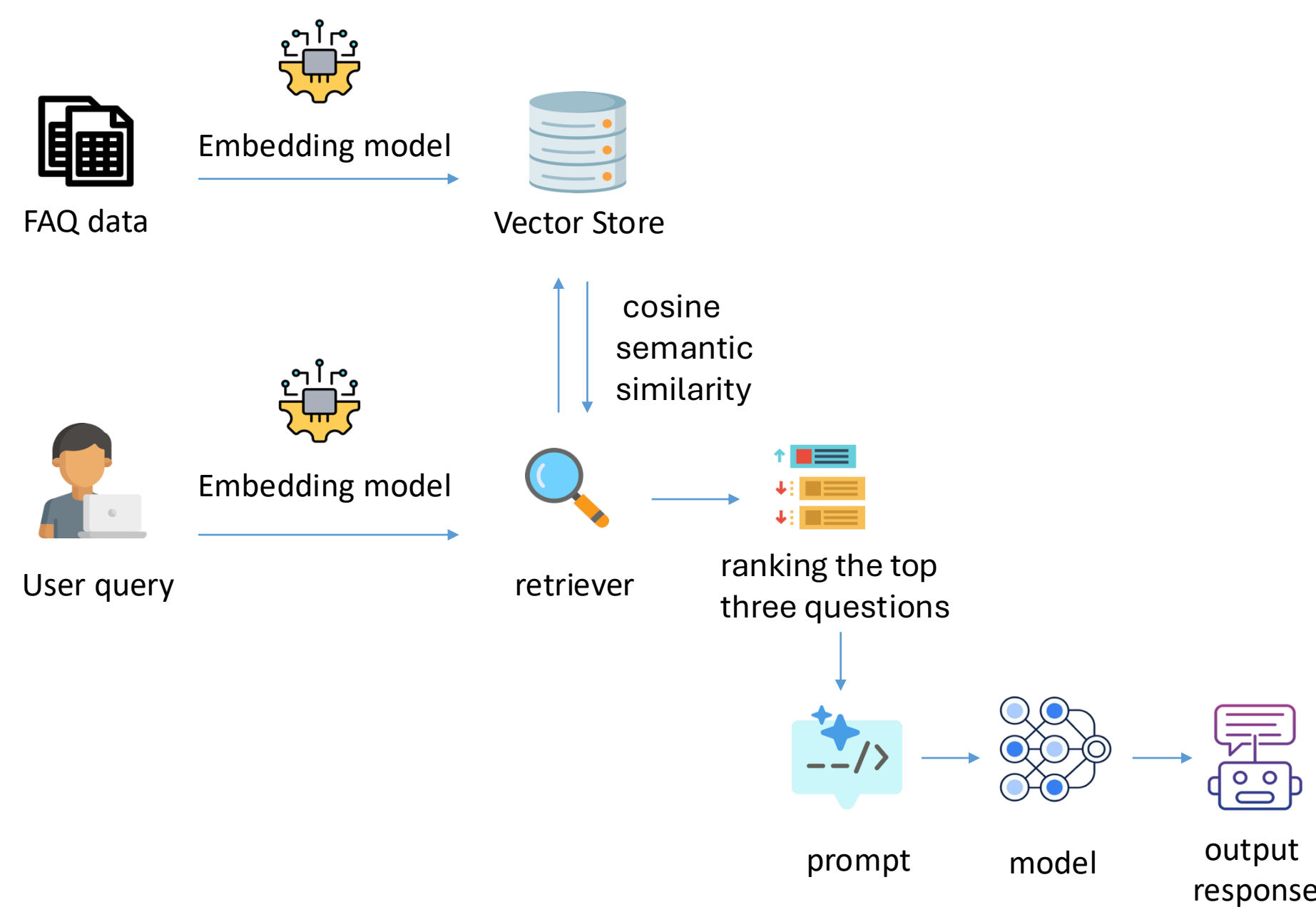
Data Collection & Preprocessing

- **Patient Information Sheets:** German-language documents related to hip prosthesis surgery.
- **FAQs:** Manually collected questions and answers from clinical sources.
- **PROMs:** EQ-5D-based self-reporting templates for integration.
- **Data Sources:**
 - Verified educational materials from Swiss hospitals and global institutions.
 - All content reviewed and adjusted by medical expert.
- **Preprocessing:** Data was cleaned, structured into Q&A pairs, and embedded using model-specific tokenizers.

Model Selection & Optimization

- Compared German-compatible embedding models on domain-specific data
- Selected models evaluated across different semantic similarity thresholds using metrics: Accuracy, Recall, F1, MRR, FPR
- BERTScore used to evaluate alignment of generated answers with reference answers
- Expert evaluation (Likert scale): >90% correctness.

Modular Architecture: Adaptive RAG



Willkommen beim OrthoMate ✨

Wie kann ich Ihnen helfen?

- 🔍 Ich habe eine Frage zur Hüftprothese
- 📊 Ich möchte meinen Gesundheitsstatus melden (SQ-5D)

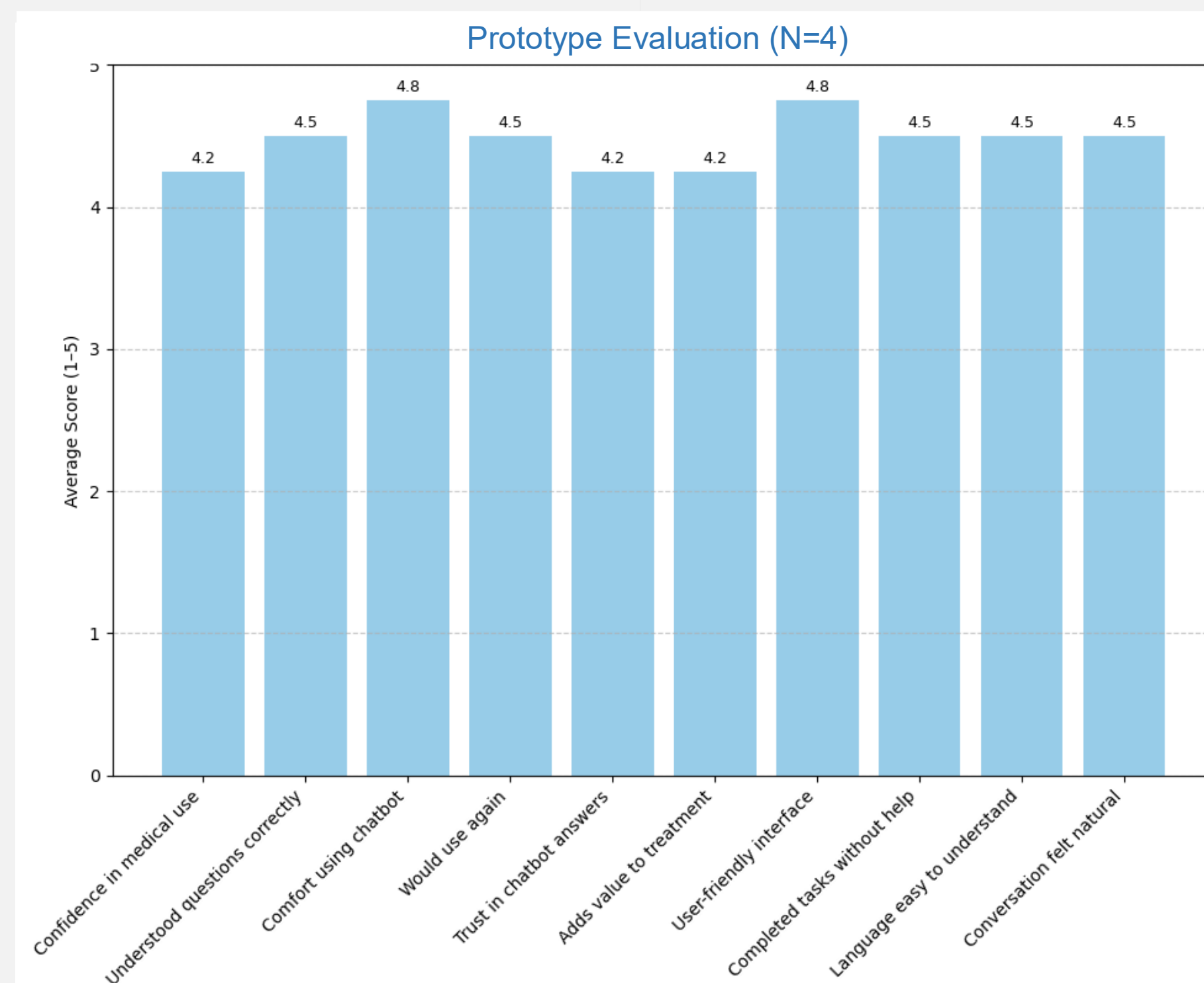
OrthoMate

Ein KI-gestützter Chatbot für Fragen zu Hüftprothesen

melden SQ-5D

- 🗨️ Stellen Sie eine Frage zur Hüftprothese...
- 🗨️ Wie lange dauert die Operation?
- 🗨️ Die Operation selbst dauert in der Regel etwa 90 Minuten. Danach verbringt man noch ein bis drei Stunden im Aufwachraum, um sich von der Anästhesie zu erholen. Bei besonders komplizierten Eingriffen kann sich die Gesamtdauer jedoch verlängern.
- 🗨️ Haben Sie noch eine weitere Frage?

Maximum Similarity Score: 0.9829 (FAQ Index: 103)
 Top 3 Matches (Before Threshold Filtering):
 Rank 1: Score=0.9829 | Question=Wie lange dauert die Operation?
 Rank 2: Score=0.9056 | Question=Wie viel Zeit nimmt die Operation in Anspruch?
 Rank 3: Score=0.8356 | Question=Wie lange dauert der chirurgische Eingriff?
 3 valid matches found. Sending to LLaMA.



System Architecture

A modular NLP pipeline was developed using an adaptive RAG framework. The system combines semantic search with generative responses to answer medical questions and collect PROMs data.

Key Components:

- Vector database created using *snowflake-arctic-embed-1-v2.0*
- Semantic similarity search with cosine similarity (threshold = 0.6)
- LLaMA-3.3-70B used for response generation based on retrieved answers
- Streamlit frontend with two features:
 - Hip prosthesis Q&A
 - EQ-5D-5L self-reporting module

Results

System output and user-facing experience

- Conversational agent successfully integrated into a user-friendly Streamlit app
- Top 3 semantically similar matches shown in terminal for transparency
- Fallback messages generated when no relevant match is found
- Self-reporting feature allows users to submit EQ-5D responses for analysis

Evaluation

The prototype was tested by four users: a domain expert, a volunteer patient, and two elderly individual with hip prosthesis-related questions. Each completed an evaluation covering:

- **Feasibility:**
 - ✓ Reliable system performance
 - ✓ No technical issues
 - ✓ Suitable for clinical settings
- **User Acceptance:**
 - ✓ High trust in responses
 - ✓ Comfortable using the chatbot
 - ✓ Willingness to reuse
- **Usability:**
 - ✓ User-friendly interface
 - ✓ Clear, natural language
 - ✓ Tasks completed without assistance

Discussion & Conclusion

A lightweight, RAG-inspired conversational agent for answering hip prosthesis surgery questions delivered accurate, user-friendly responses without hallucinations. Despite being built in a low-resource environment, the system performed well, highlighting the potential for efficient, resource-conscious development. However, further expert feedback and real-world testing are needed to guide future improvements.

Future Work

- Broader testing with real patients
- More domain expert feedback
- Larger, more coverage dataset
- Fine-tuned embeddings for medical terminologies
- Clinical impact assessment (Efficiency & Efficacy)