This project provides information on rabies through a chatbot that can understand and respond to questions asked in the Sepedi language.

Exploring open access AI tools to improve the efficacy of a Sepedi rabies awareness chatbot

BACKGROUND

Rabies is a serious public health concern in South Africa, particularly in rural and Sepedi-speaking communities where limited awareness, delayed treatment, and language barriers contribute to late diagnosis and fatal outcomes. This study presents an AI-powered chatbot designed to improve rabies awareness in these communities, an AI chatbot application that provides information in the Sepedi language has been created. The chatbot retrieves information using transformer models and utilizes large language models to provide responses to users' questions using validated medical data. The primary goal of this chatbot is to create an improved understanding of the disease and the importance of seeking immediate medical attention for those who may have been exposed to rabies.

METHODS

1. Dataset Preparation

- A rabies question-and-answer dataset was developed using 60 curated pairs based on NICD rabies educational materials. The data was first created in English, then translated into Sepedi using Gemini, and refined by a Sepedi language expert to ensure linguistic accuracy and cultural relevance. The final dataset was further validated by medical experts to confirm correctness of health-related information.

2. Text Representation and Retrieval System

- A transformer-based Sentence Transformer model was used to generate dense vector embeddings for both user queries and dataset responses. These embeddings were stored in a similarity search index (FAISS) to enable semantic retrieval of the most relevant answers. This Retrieval-Augmented approach improves system performance for unseen or rephrased Sepedi queries.

3. System Design (RAG Pipeline) + Translation Model

The final chatbot system follows a Retrieval-Augmented Generation pipeline:

1. User inputs a Sepedi question
2. Query is embedded using SentenceTransformer
3. FAISS retrieves the most similar Q&A pairs
4. An LLM is then used as a fallback to translate the query and help retrieve the most relevant answer.
5. Model generates a final response based on grounded information.

4.. Visualisation and Deployment

- The final system was deployed as an interactive chatbot using a lightweight framework such as Streamlit or Gradio. The interface allows users to input Sepedi questions and receive real-time responses generated through a combination of retrieval-augmented generation and the fine-tuned LLM. The deployment is designed to be accessible and suitable for low-resource environments to support rabies awareness in Sepedi-speaking communities.

Gemini

OpenAI

Hugging Face

RESULTS

- The chatbot successfully retrieved relevant rabies information in Sepedi.
- Handles variations in phrasing
- Retrieval-based approach provided more accurate responses than classification models.

DISCUSSION

- The chatbot was able to respond effectively to rabies-related queries in Sepedi, demonstrating the potential of AI for low-resource languages.
- The use of a retrieval-based model allowed the system to provide relevant answers based on existing verified data.
- However, the system sometimes returned incorrect or unrelated answers when the user query differed significantly from the stored dataset.
- To address this, GPT was integrated to improve, rephrase, and enhance responses, making the chatbot more flexible and user-friendly.
- Alias mapping also helped handle different ways of asking the same question, improving matching accuracy.
- Despite these improvements, performance is still limited by the size and quality of the Sepedi dataset.

VISUALISATIONS

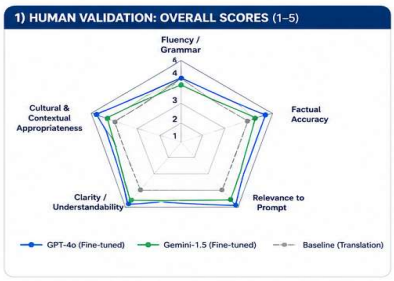

Figure 1 : Radar plot of human validation scores

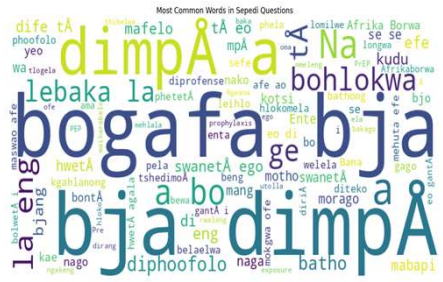

Figure 2 : WordClouds of LLM responses

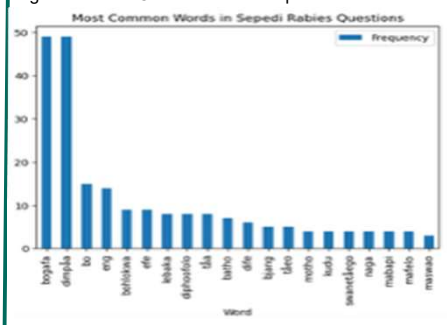

Figure 3 : Most Common words in Sepedi Rabies questions

Muphulusi Dzivhani , Nonhlanhla Mabila

Take a picture to download the full paper