|  |  |
| --- | --- |
|  |  |
|  |
|  |

**High Impact Skills Development Program for**

**Gilgit-Baltistan Offered by NUST with the Collaboration of GBIT**

**Natural Processing Language based**

**Capstone Project**

**Report on Burushaski ChatBot Model**

**Saliha Nishad Farid**

[**https://github.com/Saliha-Nishat**](https://github.com/Saliha-Nishat)

[**salihanishad321@gmail.com**](mailto:salihanishad321@gmail.com)

[**https://github.com/Saliha-Nishat/Capstone-Project-brushobot**](https://github.com/Saliha-Nishat/Capstone-Project-brushobot)

**Section: 1**

**Course duration: 6 months**

**Trainers throught the course:**

**Eid Mohammad, Faiza Ali, Abdul Saboor & Wasif Mahmoud**

**Introduction**

The Burushaski ChatBot is an innovative language translation and AI-based application designed to facilitate communication by translating Burushaski phrases into English. Burushaski is a unique and lesser-known language, and this chatbot aims to bridge the communication gap while integrating AI-powered text responses for an enhanced user experience.

This application uses Python, Streamlit for deployment, and Google Generative AI for content generation. It incorporates a dataset of Burushaski phrases and their corresponding English translations.

**Objective**

The main goals of the Burushaski ChatBot are:

Translation: Provide accurate translations of Burushaski phrases into English using a pre-defined dataset.

AI-Powered Interaction: Use Google Generative AI to offer insightful and relevant responses based on user input and translations.

Suggestions: Recommend similar Burushaski phrases for incomplete or partially correct user inputs.

Ease of Use: Offer a simple, user-friendly interface for smooth interaction.

**Model Components**

1. Translation Model

**Functionality:**

The chatbot translates Burushaski text to English using a dataset (Bb.xlsx) loaded via Pandas.

It matches user input with phrases in the dataset to find corresponding English translations.

**Implementation:**

The translate\_to\_english function searches for exact or partial matches in the dataset.

If a match is found, it returns the English equivalent; otherwise, it notifies the user that no match is available.

2. Suggestion System

Functionality:

This module recommends similar Burushaski words or phrases when the user's input does not yield a direct match.

**Implementation:**

The suggest\_words function compares the first two characters of the user's input with the phrases in the dataset.

Matches are suggested to the user to refine their query.

3. AI-Powered Response Generation

Functionality:

Google Generative AI is used to generate a conversational response based on the translation.

**Implementation:**

The get\_text\_response function utilizes the gemini-1.5-flash model from Google Generative AI SDK.

It generates responses that enhance user engagement by providing additional context or explanations.

4. User Interface

Framework: Streamlit

**Design:**

Simple, intuitive interface with customized CSS for enhanced aesthetics.

Input field for user queries and dedicated sections for translation, suggestions, and AI responses.

Challenges Encountered

**1. Deployment Issues**

The model initially failed to respond on Streamlit Cloud due to missing API key configuration.

Resolution: API keys were added using Streamlit Cloud’s Secrets Manager.

**2. Dependency Management**

Missing dependencies (google-generativeai, pandas) during deployment.

Resolution: A requirements.txt file was created and included in the project.

**3. Dataset Accuracy**

Limited dataset size might result in untranslated phrases or suggestions.

Resolution: The dataset needs to be expanded with more Burushaski phrases and their English translations.

Strengths

Cultural Preservation: Supports a lesser-known language, making it accessible to a broader audience.

AI Integration: Leverages advanced AI models for improved user interaction.

Scalability: The modular architecture allows easy integration of additional features like voice-to-text translation or multi-language support.

Limitations

Dataset Dependency: Translations are restricted to the entries available in the dataset.

Network Dependency: AI response generation relies on a stable internet connection for API calls.

Latency Issues: The model may experience delays during heavy server loads or poor network conditions.

Future Enhancements

Dataset Expansion: Incorporate more phrases and idiomatic expressions to improve translation accuracy.

Voice Input and Output: Add speech recognition and text-to-speech capabilities for hands-free usage.

Offline Mode: Develop an offline version for translation using pre-trained machine learning models.

Error Handling: Enhance error messages to guide users better when issues arise.

**Conclusion**

The Burushaski ChatBot is a valuable tool for promoting cultural understanding and communication. By combining translation, AI interaction, and a user-friendly interface, it offers an engaging and practical solution for both native speakers and learners of Burushaski. While some challenges remain, the model has strong potential for growth and impact, particularly with further development and dataset enrichment.