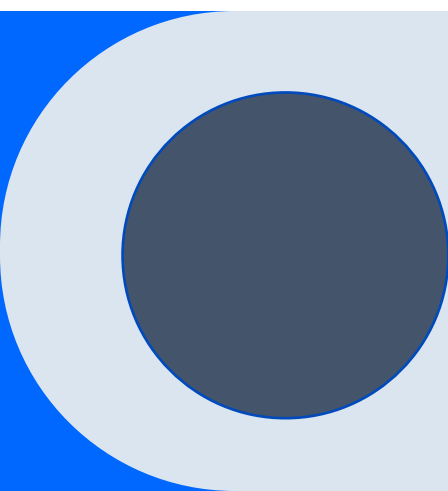# Newsreader Chatbot for the Telugu Community using NLP

A Project Overview

SE22UARI053 – G.Chaithri

SE22UARI020 – Anvitha Reddy

SE22UARI133 – P. Pravallika

SE22UARI202 – Vandana

SE22UARI101 – M. Sowmitha

Agenda

- Motivation
- Problem Statement
- Proposed pipeline
- Timeline
- Expected Outcome/Application

Motivation

- **Increasing Need for Regional Language News**: There is a rising demand for news content that is easily accessible in regional languages like Telugu.

- **Significance of Timely Information**: Real-time dissemination of news is essential for keeping the community informed about current events and developments.

- **Boosting Digital Engagement**: A newsreader chatbot can enhance digital interaction and foster a stronger connection within the Telugu-speaking community.

Problem Statement

A large section of the Telugu-speaking community continues to be neglected despite the growing digitalization of news material because there aren't many easily accessible and interesting news channels available in their mother tongue. The goal of this project is to bridge this gap by creating a thorough platform for Telugu news summaries.

Proposed Pipeline

- **Data Collection:** We will gather Telugu news articles from a variety of sources to ensure a broad and diverse range of content.

- **Preprocessing:** This step involves cleaning the text, breaking it down into manageable parts (tokenization), and performing necessary language processing to prepare it for analysis.

- **NLP Model Development:** We'll implement models designed to recognize user intents and extract important entities from the news articles, making the content more interactive.

- **Chatbot Framework:** We will build a conversational interface using tools like Rasa or Dialogflow, creating a user-friendly way for people to interact with the news.

- **Sentiment Analysis:** We will implement sentiment analysis techniques to gauge the emotional tone of the news articles, allowing the chatbot to respond to user queries with insights into public sentiment surrounding specific topics or events.

- **Testing:** We will conduct comprehensive testing of the chatbot and NLP model to evaluate their performance, ensuring accuracy in intent recognition, entity extraction, and sentiment analysis. This phase will involve user feedback and iterative improvements to enhance the overall user experience.

Timeline
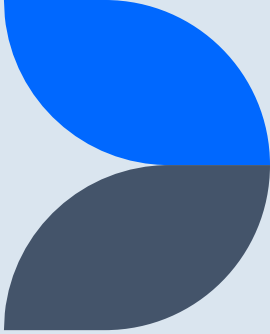
Week	Progress
1 - 2	Literature review and requirements gathering
3 - 4	Data collection and preprocessing
5 - 6	Model development and training
7 - 8	Building the chatbot interface
9 - 10	Testing and deployment

Expected Outcome/Application

- A user-friendly chatbot that delivers personalized news updates in Telugu
- Ability to handle user queries and provide contextual information
- Users will gain insights into public sentiment regarding specific topics or events, enhancing their understanding of the news landscape.
- Potential applications in education, local governance, and community engagement
- The framework developed can be scaled to include more languages or content types, broadening its applicability to different communities.
- The testing phase will provide data on the chatbot's performance, guiding further enhancements and ensuring reliability.

Q&A

Q. How will you train the NLP models, and what datasets will you use?

A. Data can be collected from online Telugu newspapers web scraping, blogs and social media platforms. This data will be preprocessed using methods such as Tokenization and Normalization. If using pre-trained transformer models (like BERT), we'll fine-tune them on the specific datasets to adapt them for Telugu language tasks. Once the models are trained and validated, test them on the unseen test dataset to assess their generalization capability.

Q. How will you address challenges related to understanding slang, idioms, or regional dialects in Telugu?

A. Collect a wide range of conversational data from different sources(and different regions), including social media, local news, and community forums to understand different context of language use. Context sensitive models would also be used in training. Model would also be regularly updated as languages are always evolving.

Thank you