

# Gap Analysis

- Existing models show strong performance in English but struggle with Bengali.
- Limited availability of pre-trained Bengali NLP models.
- Lack of effective QA systems for Bengali-speaking communities.

cont..

# Methodology

**Data Collection:** Collected data from **Ajker Bissho**, **MediGK**, **Current Affairs**, etc.

**Preprocessing:**

- Tokenization
- Stop-word removal
- Data cleaning for improving dataset quality.

**Model Selection:** Fine-tune pre-trained models like **mt5-small**, **BanglaGPT**, **Dialo-GPT** or **GPT-2**.

**Training Process:** Train with the dataset using transfer learning techniques.

**Evaluation Metrics:** Use metrics like F1-score, BLEU score, and accuracy to measure model performance.

cont..