

**H I T A M**

FIND YOUR PATH

HYDERABAD INSTITUTE OF TECHNOLOGY & MANAGEMENT

S O U V E N I R

A C A D E M I C P R O J E C T S

PROJECT EXPO - 2025

**Department of Emerging Technologies**

PROJECT GROUP ID : 15

**PROJECT TITLE : Unsupervised Language Model Adaptation For Low-Resource Languages**

**PROJECT MEMBERS : D. Manaswini**

**K. Nikshipta**

**S. Prasanna Lakshmi**

**V. Rishik Reddy**

**PROJECT GUIDE : Dr. M. Rajeshwar**

**SUMMARY OF THE PROJECT :** This project focuses on building a machine translation system for English ↔ Bengali using pretrained Transformer models. The dataset used consists of bilingual sentence pairs sourced from two files: ben.txt and english\_to\_bangla.csv. The data underwent standard preprocessing steps including lowercasing, removal of special characters, and tokenization using the BanglaT5 tokenizer. The dataset was then split into 80% training and 20% validation sets. Both models are built on the T5 (Text-to-Text Transfer Transformer) architecture, which employs an encoder-decoder structure. Each model includes 12 encoder layers and 12 decoder layers, with a hidden size of 768, 12 attention heads, and feedforward networks of size 3072. Dropout is applied at a rate of 0.1 to prevent overfitting. The implementation includes data preprocessing, model training for both directions (English → Bengali and Bengali → English), inference, and evaluation using multiple NLP metrics.

The performance of the machine translation models was assessed using three widely accepted evaluation metrics: BLEU (Bilingual Evaluation Understudy), METEOR (Metric for Evaluation of Translation with Explicit ORdering), and TER (Translation Edit Rate). These metrics provide a comprehensive view of the translation quality in terms of both precision and required human edits.

**BLEU Score**: 0.2867, **Corpus BLEU Score**: 0.3004

**METEOR Score**: 0.5910

**TER Score**: 33.3333

**PROJECT PHOTO :**



