# Urdu to Roman-Urdu Translation Model

## Introduction

This project focuses on developing a machine translation model that converts Urdu text, especially poetry, into Roman-Urdu while preserving semantic meaning and poetic expression. The project demonstrates the practical application of Neural Machine Translation (NMT) using Seq2Seq architectures with LSTMs, tokenization techniques, and extensive hyperparameter tuning.

## Methodology

We designed a Seq2Seq model with a BiLSTM encoder and multi-layer LSTM decoder. Byte Pair Encoding (BPE) was used for effective tokenization of both Urdu and Roman-Urdu corpora. The model was trained with variations in embedding dimensions, hidden sizes, and number of layers, and evaluated using BLEU, perplexity, and loss tracking. Experimentation allowed us to fine-tune hyperparameters and overcome common challenges like dimension mismatches, tokenization scope, and scheduler compatibility.

## Deployment

To make the model accessible, we deployed a demo application where users can test Urdu-to-Roman Urdu translation interactively.

■ Deployed Demo: https://naqsh-e-urdu.streamlit.app/

## Publications & References

1. Medium Article: https://medium.com/@abrar11ahmad99/developing-a-urdu-to-roman-urdu-translation-model-a-collaborative-effort-with-muhammad-taha-under-24fbdd02f724

2. LinkedIn Post: https://www.linkedin.com/posts/abrar-ahmad-a31a3b248_urdu-machinetranslation-nlp-activity-7376945936083800064-8Jvu?utm_source=share&utm;_medium=member_desktop&rcm;=ACoAAD1guwUBjlVO3e1HTlORNrLOWvSKtTuYDXU

## Conclusion

This collaborative effort successfully demonstrates how deep learning can be leveraged to bridge linguistic gaps in literature and culture. By translating Urdu into Roman-Urdu, we aim to enhance accessibility and preserve poetic essence for wider audiences.