# KONKANI POEM NEXT WORD PREDICTION

## AN AI-POWERED LANGUAGE MODEL FOR KONKANI POETRY

SHARVARI NAIK

ZAINAB RAJABALI

# INTRODUCTION: WHY KONKANI POETRY AND AI?

- Konkani: A culturally rich Indo-Aryan language spoken in Goa, India.
- Limited digital tools for preserving Konkani literature.
- Goal: Use AI to generate and predict Konkani poetry.
- Fine-tuned GPT-2 on Konkani poems for creative text generation.

# PROJECT OBJECTIVES

- Preserve Konkani poetry through a curated dataset.
- Generate coherent Konkani poetic sequences.
- Predict the next word in a poetic prompt.
- Build a scalable framework for regional languages.
- Evaluate model performance (e.g., perplexity, loss).

# DATASET & PREPROCESSING

- Source: Konkani Sentiment Analysis Corpus (GitHub: anniedhempe)
- 54 Konkani poems in .docx format.
- Preprocessing: Extracted text, removed metadata, saved as .txt.
- Split: 70% train (38 poems), 15% validation (8 poems), 15% test (9 poems).

# MODEL & TRAINING

- Model: GPT-2 (base), fine-tuned for Konkani poetry.
- Training: 3 epochs, learning rate 5e-5, batch size 3.
- Tools: Hugging Face Transformers, PyTorch, Google Colab (GPU).
- Features: Poetry generation and next-word prediction.

# KONKANI POETRY GENERATION

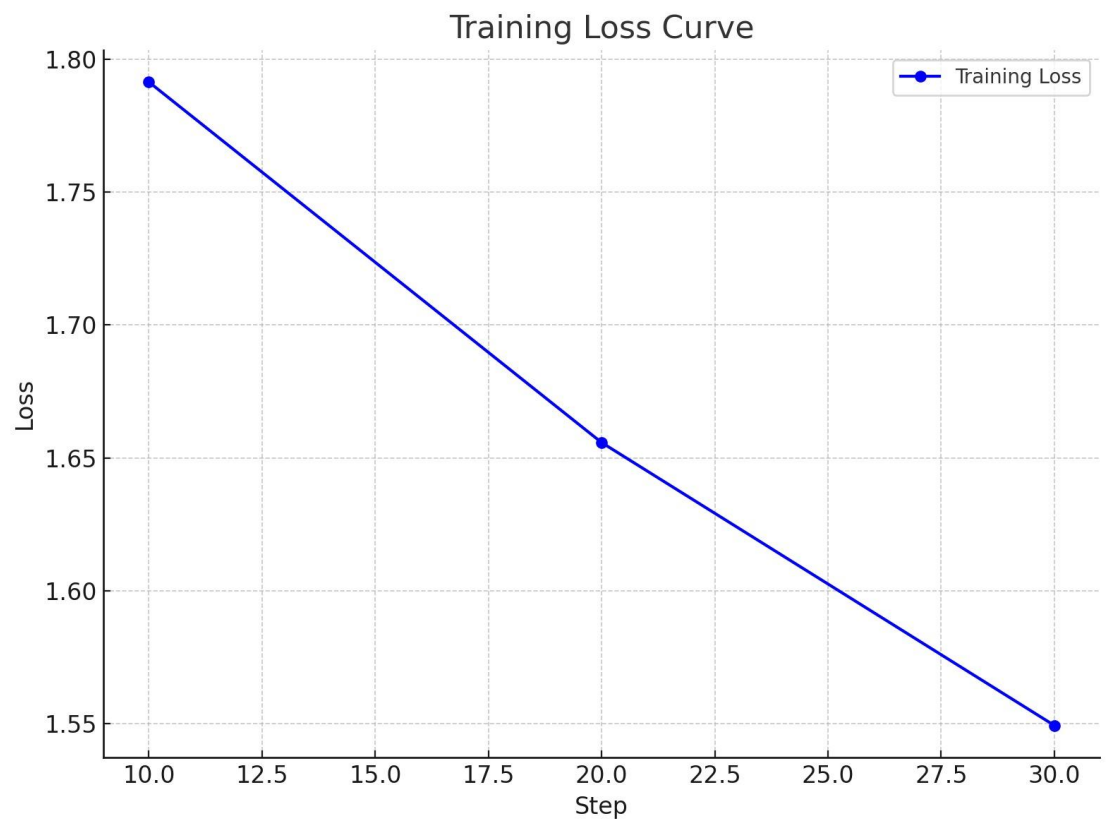
- Loading the fine-tuned GPT-2 model.
- Generating poetry from a prompt.
- Predicting the next word in a sequence.
- Exploring creative outputs with temperature variation.

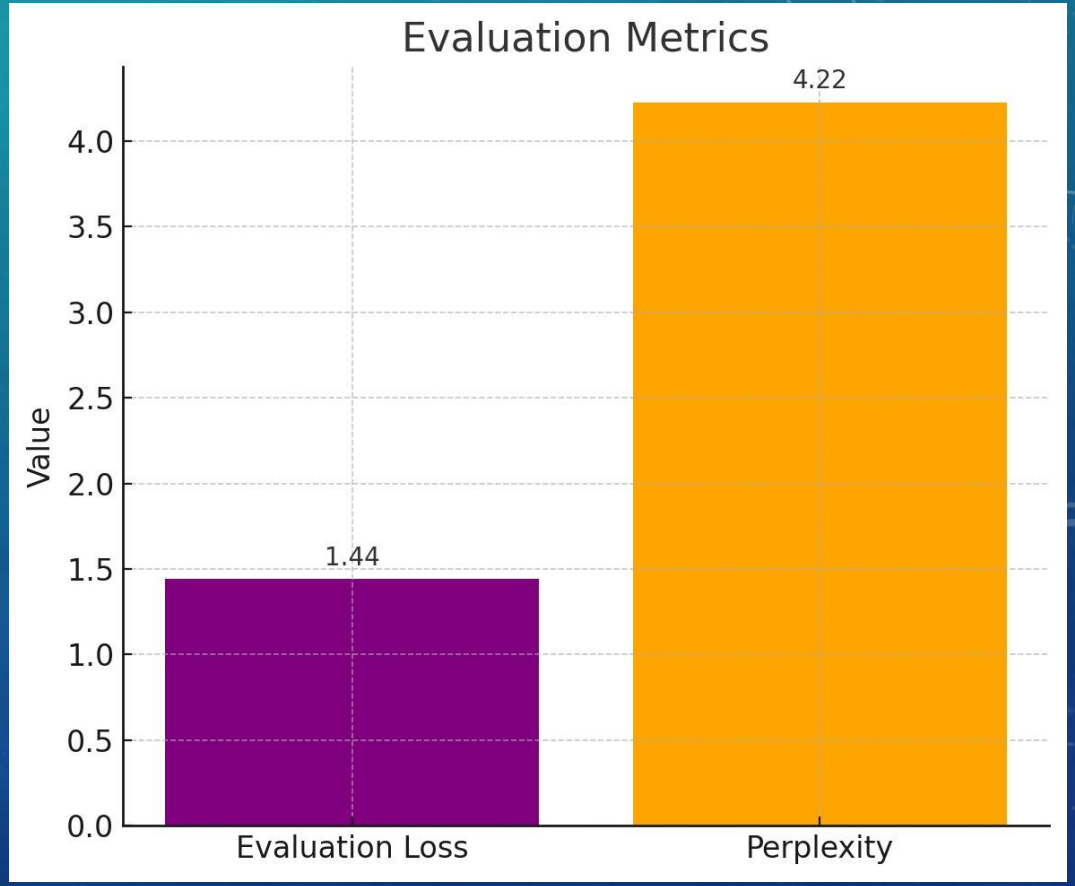
# RESULTS

- Training Loss: Decreased to 1.6310 over 3 epochs.
- Evaluation Loss: 1.4409.
- Perplexity: 4.2243 (good for small dataset).
- Example Output: "तुझें दिल म्हजें आकाश" → Poetic sequence generated.

# RESULTS

📊 Evaluation Loss: 1.4409
📉 Perplexity: 4.2243



# DISCUSSION

- Strengths: Captures Konkani poetic style, low perplexity.
- Limitations: Small dataset (54 poems), some incoherent outputs.
- Future Work: Expand dataset, add UI, support other languages.

# CONCLUSION

- Successfully fine-tuned GPT-2 for Konkani poetry.
- Demonstrates AI’s role in preserving regional languages.
- Foundation for scalable NLP tools for Konkani and beyond.

# REFERENCES


- Hugging Face Transformers Documentation: https://huggingface.co/docs/transformers
- PyTorch Documentation: https://pytorch.org/docs/stable/index.html
- Rao, S., et al., 2021. "Fine-Tuning Language Models for Regional Languages." Journal of NLP Research.
- GPT-2: Radford, A., et al., 2019. "Language Models are Unsupervised Multitask Learners."
- Konkani Sentiment Analysis Corpus: https://github.com/anniedhempe/Konkani-sentiment-analysis-corpus