EXPERIMENT 8

Seq2Seq Model with Attention Mechanism

OBJECTIVE

To build and evaluate an LSTM-based Seq2Seq model enhanced with attention mechanisms (Bahdanau and Luong) for improved translation quality.

MODEL ARCHITECTURE (WITH ATTENTION)

Encoder:

- Embedding Layer → LSTM Layer → All hidden states returned

Attention Mechanism:

- **Bahdanau (Additive)**: Alignment scores based on learned additive scoring
- **Luong (Multiplicative)**: Alignment scores using dot product between encoder and decoder states

Decoder:

- Embedding Layer
- Context Vector (from attention) concatenated with embedding
- LSTM → Dense → Softmax

Initialization & Regularization:

- Xavier/He initialization
- Tanh for LSTM/attention layers, Softmax for output
- No dropout or regularization
- Padding tokens masked during attention and loss computation

---

RESULTS

Model	Final Loss	BLEU Score
Bahdanau Attention	7.4280	0.1196

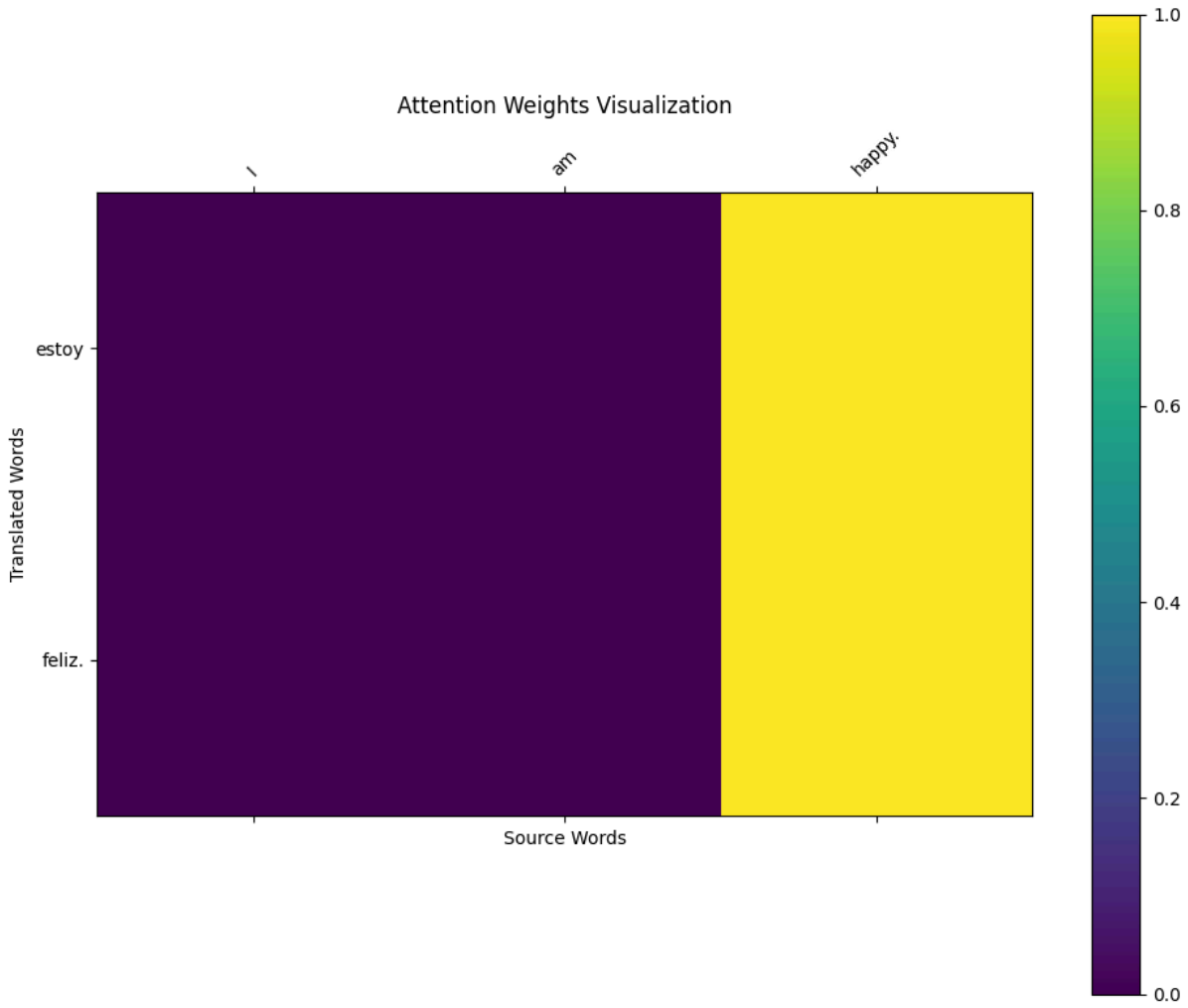
Luong Attention	6.8122	0.1309

SAMPLE TRANSLATIONS

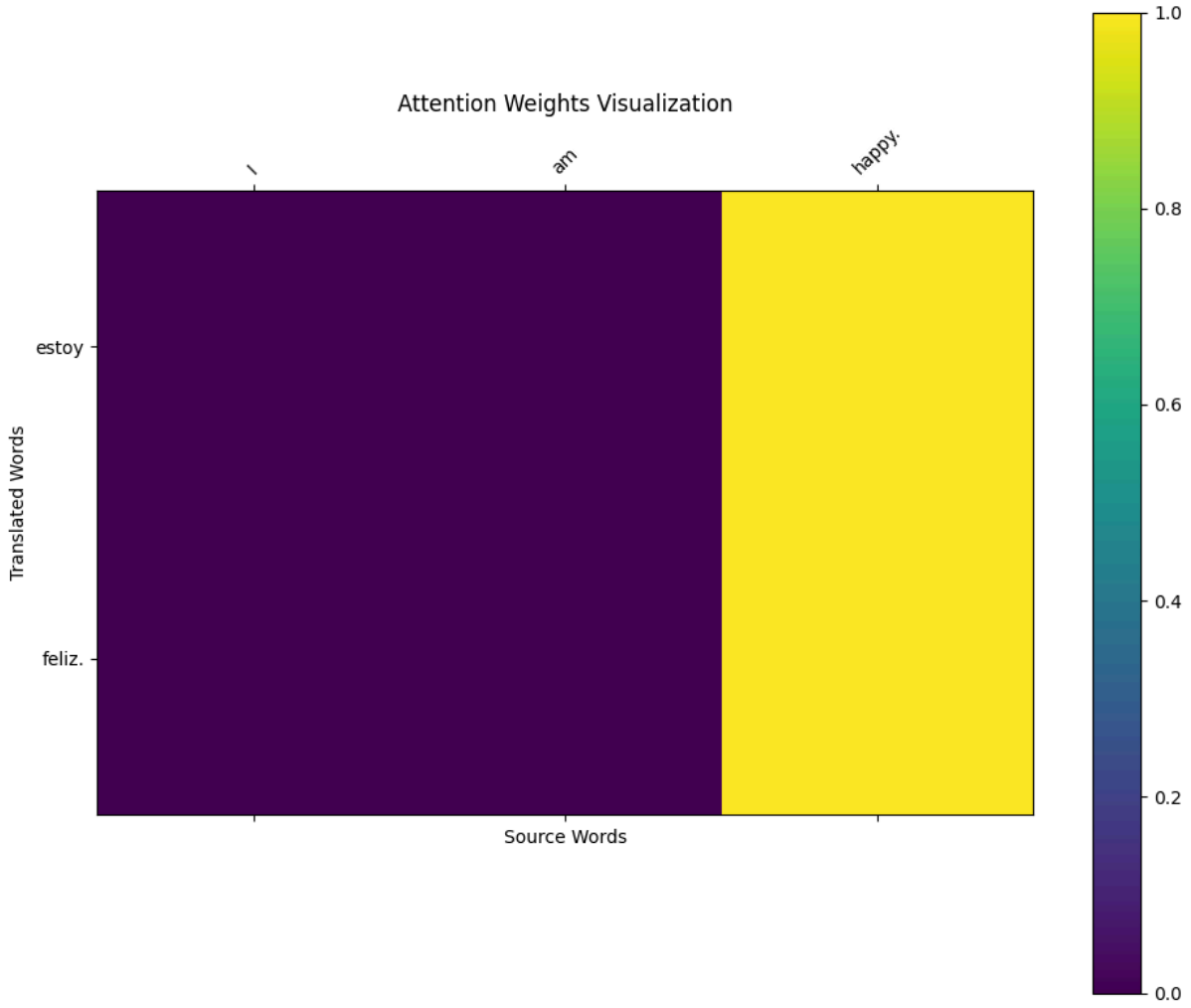
Input	Model	Output
I am happy	Bahdanau	estoy feliz.
	Luong	estoy feliz.

VISUALIZATION

- **Bahdanau:** Heatmaps show smooth, interpretable alignment weights over input tokens



- **Luong:** Sharp alignments ideal for short, direct phrases



CONCLUSION

- Attention boosts fluency and context handling
- Bahdanau provides richer sentence-level context
- Luong performs slightly better on short sequences
- BLEU scores and qualitative outputs improve with attention

FUTURE WORK

- Use Bidirectional LSTM encoders
- Add dropout regularization
- Use pretrained embeddings (GloVe/FastText)
- Fine-tune on larger bilingual corpora
- Explore Transformer and BERT-based encoder-decoder models