

# CSET419 – Introduction to Generative AI

## Lab – 4

### Objective

The objective of this lab is to design and implement a **simple text generation model** that can learn patterns from a given text corpus and generate new, meaningful text sequences.

### Learning Outcomes

After completing this lab, students will be able to:

1. Understand the basics of text generation
2. Preprocess textual data for neural networks
3. Implement a sequence-based neural network model
4. Generate new text using a trained model
5. Analyze the quality of generated text

### What is Text Generation?

Text generation is a Natural Language Processing (NLP) task where a model learns patterns from text data and generates new text that resembles the training data.

### Common Approaches

- N-gram language models
- Recurrent Neural Networks (RNN)
- LSTM / GRU networks
- Transformer-based models (advanced)

### Experiment:

Component–I: RNN / LSTM Based Text Generation

### Tasks

1. Load and preprocess text data
2. Perform tokenization (character-level or word-level)
3. Create input-output sequences
4. Design RNN/LSTM architecture
5. Train the model
6. Generate new text using seed input

## Dataset

Take any text for example,

artificial intelligence is transforming modern society.

it is used in healthcare finance education and transportation.  
machine learning allows systems to improve automatically with  
experience.

data plays a critical role in training intelligent systems.  
large datasets help models learn complex patterns.  
deep learning uses multi layer neural networks.  
neural networks are inspired by biological neurons.  
each neuron processes input and produces an output.  
training a neural network requires optimization techniques.  
gradient descent minimizes the loss function.

natural language processing helps computers understand human  
language.

text generation is a key task in nlp.  
language models predict the next word or character.  
recurrent neural networks handle sequential data.  
lstm and gru models address long term dependency problems.  
however rnn based models are slow for long sequences.

transformer models changed the field of nlp.  
they rely on self attention mechanisms.  
attention allows the model to focus on relevant context.  
transformers process data in parallel.  
this makes training faster and more efficient.  
modern language models are based on transformers.

education is being improved using artificial intelligence.  
intelligent tutoring systems personalize learning.  
automated grading saves time for teachers.  
online education platforms use recommendation systems.  
technology enhances the quality of learning experiences.

ethical considerations are important in artificial intelligence.  
fairness transparency and accountability must be ensured.  
ai systems should be designed responsibly.  
data privacy and security are major concerns.  
researchers continue to improve ai safety.

text generation models can create stories poems and articles.  
they are used in chatbots virtual assistants and content creation.

generated text should be meaningful and coherent.  
evaluation of text generation is challenging.  
human judgement is often required.

continuous learning is essential in the field of ai.  
research and innovation drive technological progress.  
students should build strong foundations in mathematics.  
programming skills are important for ai engineers.  
practical experimentation enhances understanding.

### Expected Output

- Generated text samples

## Component–II: Transformer Based Text Generation

### Tasks

1. Use the same dataset as Component–I (**you may reduce the data size if memory exhaust problem occurs**)
2. Apply subword or word-level tokenization
3. Implement positional encoding
4. Design Transformer encoder blocks
5. Train the model
6. Generate text

### Expected Output

- Generated text samples