(b) What are the major ethical issues in Large Language Models? Explain any four with real-world implications. (5)

(c) Explain the three types of Transformer-based language model architectures with suitable diagram. (6)



[This question paper contains 8 printed pages.]

**Your Roll No.....**

**Sr. No. of Question Paper : 7970** **L**

Unique Paper Code : 2343010029

Name of the Paper : Natural Language Processing

Name of the Course : **B.Sc. (H)/B.A (P)/B.Sc. (P) – DSE**

Semester : VIII

Duration : 3 Hours Maximum Marks : 90

**Instructions for Candidates**

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. **Section A** is compulsory.
3. Attempt any **FOUR** questions from **Section B.**
4. Parts of a question should be attempted together.

## Section A

1. (a) What are stopwords? Why are stopwords removed in text processing? (3)

(b) What is tokenization in natural language processing and why is it required? (3)

(c) State the Markoy assumption that underlies bigram language models and write the bigram probability formula it produces. Give one limitation of the bigram language model. (3)

(d) Explain the Bag-of-Words (BoW) model and its limitations. How does Word2Vec address these limitations? (3)

(e) Why are word embeddings important in NLP, and how do they differ from one-hot encoding? (3)

(f) State one application for each of the three types of Transformer-based models. (3)

(g) What is pooling in neural NLP models, and how do mean pooling and max pooling work? (3)



(c) Explain the XOR problem in neural networks. Why is it not linearly separable, and how does a multi-layer perceptron overcome this limitation? (5)

6. (a) Write short notes on **Any two** of the following:

I. Feedforward Neural Networks

II. Recurrent Neural Networks

III. Activation Functions (6)

(b) What is self-attention in Transformer models, and why is it important for language modeling? Also, explain how the Query, Key, and Value matrices are used to compute attention. (6)

(c) What is prompting in large language models? Explain prompt engineering with one example. (3)

7. (a) Explain how text is generated in language models. Describe any one sampling methods used during generation. (4)

(b) Compute the TF, IDF, and TF-IDF score for the term *"cat"* in document D1 given the corpus: D1 = *"the cat sat"*, D2 = *"the cat ate"*, D3 = "the dog sat".

(c) Compute the cosine similarity between the vectors A=[1,2,3] and B=[2,1,2]. Show the dot product and magnitude calculations, and interpret the result. (4)

5. (a) What is Part-of-Speech (POS) tagging? Perform for the sentence *"The quick brown fox jumps over the lazy dog."* Assign an appropriate tag to each word. (5)

(b) Apply NER tagging to the sentence "*Sundar Pichai joined Google in Mountain View on Monday.* İdentify all named entities and assign appropriate BIO tags along with their entity types. (5)



(h) Explain perplexity as a metric for language model evaluation. If a bigram model assigns probability 0.001 to a test sentence of 10 words, compute its perplexity. (3)

(i) Why is Minimum Edit Distance computed using dynamic programming instead of a brute-force approach, and what are the edit operations involved in calculating it? (3)

(j) Define Term Frequency-Inverse Document Frequency (TF-IDF). Explain its significance in text representation. (3)

**Section B**

2. (a) Given the regular expression r'\b[A-Z][a-z]+\b' and the text *"Alice went to Paris in January"*.

I. List all the words matched by the pattern.

II. Explain the purpose of each part of the pattern (\b, [A-Z], [a-z]+).

III. Modify the pattern so it also matches words written entirely in uppercase, and apply it to the same text. (5)

(b) Explain the differences between lemmatisation and stemming with one example. (5)

(c) Define Natural Language Processing (NLP). Explain any four real-world applications with suitable examples. (5)

3. (a) Perform two iterations of the Byte Pair Encoding (BPE) algorithm on the given corpus: 'l o w </w>' (×5), 'l o w e r </w>' (×2), 'n e w e s t </w>' (×6). Compute the frequency of all adjacent symbol pairs before each merge, identify the most frequent pair, update the corpus after each merge, and list the resulting vocabulary. (5)



(b) Compute the Minimum Edit Distance (MED) between "Saturday" and "sunday" using dynamic programming with unit costs for insertion, deletion, and substitution. (7)

I. Write the recurrence relation used.

II. Construct the Dynamic Programming/ word edit distance matrix.

III. Show one optimal sequence of edit operations.

(c) Explain the zero-probability problem in n-grams language models. Using Laplace (Add-1) smoothing, compute the probability P(cat|the) given Count(the cat)=0, Count(the)=15, and vocabulary size V=500. (3)

4. (a) What is Vector Semantics? Explain the intuition behind representing word meanings as vectors. What is the difference between static embeddings and contextual embeddings? (6)