



Department of Artificial Intelligence and Machine Learning

Date: 24.06.2025	Improvement Test	Max. Marks: 50+10
Semester: VI	UG	Duration: 1 $\frac{1}{2}$ Hrs + $\frac{1}{2}$ Hr
Course Title: Natural Language Processing and Transformers		Course Code: AI363IA

Part A

Sl No	Questions	M	BT	CO												
1.	Write the equation for the scalar dot product used in the attention mechanism	2	1	1												
2.	What is the purpose of adding a classification head to a Transformer model?	2	2	1												
3.	What does the feed-forward layer in a Transformer do?	2	2	2												
4.	What is the difference between extractive and abstractive summarization?	2	2	1												
5.	Given the following probabilities for 5 tokens from a language model output: If $k = 3$, which tokens will be considered for sampling? <table><tr><th>Token</th><th>Probability</th></tr><tr><td>A</td><td>0.40</td></tr><tr><td>B</td><td>0.25</td></tr><tr><td>C</td><td>0.15</td></tr><tr><td>D</td><td>0.12</td></tr><tr><td>E</td><td>0.08</td></tr></table>	Token	Probability	A	0.40	B	0.25	C	0.15	D	0.12	E	0.08	2	3	3
Token	Probability															
A	0.40															
B	0.25															
C	0.15															
D	0.12															
E	0.08															

Part B

Sl No	Questions	M	BT	CO
1.a	Compare and analyse top-k and nucleus sampling techniques in text generation tasks. How do they balance diversity and coherence in generated text?.	5	4	3
1.b	Explain the concept of self-attention in transformer architecture and explain how the attention score is calculated	5	2	1
2.a	What are some common methods used for text summarization and how do models like T5, BART and PEGASUS contribute to improving the state of art in summarization tasks?	5	3	3
2.b	Describe the positional embedding's in the transformer model and how they help capture the sequential information in the input	5	2	1
3.a	Explain Character and Sub word tokenization, and also mention the importance of numericalization in tokenization process	5	3	1
3.b	What is layer normalization in Transformers? Explain why and where it is applied in the Transformer architecture	5	3	1
4	Explain the role of transformer models as feature extractors in training a text classifier. How does this approach differ from fine-tuning the transformer model?	10	4	2
5	Explain greedy search decoding and beam search decoding in the context of text generation using transformer models	10	2	3