International Institute of Information Technology, Hyderabad

(Deemed to be University)

CS7.401: Introduction to NLP IIIT Hyderabad Question cum Answer Booklet

End Semester Examination

Max. Time: 3 Hr			Max. Marks:60
Roll No:		Programme:	Date of Exam:
Room no:	Seat No:		Invigilator's Signature:
Special Instru	ctions about the	exam	
1. Answer all qu	estions.		
2. Make appropi	riate assumptions.		

Marks Table (To be filled by the Examiner)
Note: The table can be designed as per your requirement

No additional sheet will be provided.

Question No / Marks	Course outcome number(s)				Name of the Examiner who marked
1	CO-1				
2	CO-1				
3	CO-3				
4	CO-2				
5	CO-1				
6	CO-3				
7	CO-4				

General Instructions to the students

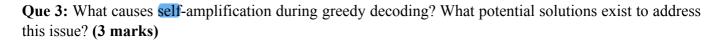
- 1. Place your Permanent / Temporary Student ID card on the desk during the examination for verification by the Invigilator.
- 2. Reading material such as books (unless open book exam) are not allowed inside the examination hall.
- 3. Borrowing writing material or calculators from other students in the examination hall is prohibited.
- 4. If any student is found indulging in malpractice or copying in the examination hall, the student will be given 'F' grade for the course and may be debarred from writing other examinations.

Best of Luck

Que 1: Indicate T(rue) or F(alse). (5 marks)
1. The fundamental idea presented in the Chinchilla paper is that the majority of LLMs are excessively
trained and thus need a greater number of parameters. T/F
2. Including code in pertaining data can also benefit non-code tasks. T/F
3. While RL-based approaches like RLHF demand more computational resources and time, they exhibit
greater stability compared to non-RL methods like DPO. T/F
4. The behavior-cloning effect observed in instruction fine-tuning stems from a reduced diversity among
the underlying tasks. T/F F
5. Max-probability decoding schemes are suitable for low entropy tasks. T/F 🔻
6. In speculative decoding, in worst case at least 1 token is generated sampled from draft model. T/F
7. Increasing temperature in softmax may help in low entropy tasks. T/F
8. RAG has lots of similarity with in-context learning. T/F
9. LLM inference requires a larger memory footprint than instruction fine-tuning. T/F
10. Despite the use of a large beam size, beam search decoding does not ensure an optimal solution. T/F
Que 2: Discuss briefly about these concepts:(10 marks)
a) Reasons for emerging capabilities in LLMs. (2 marks)
scaling i.e bigger datasets and larger parameters.
b) Core motivation behind instruction fine-tuning. (2 marks)

Task unification, better suitability for unseen tasks, response to natural language instructions

c)	The performance of instruction fine-tuning in FLAN reaches a plateau rapidly, even with an increasing number of tasks. (2 marks)
	beacuse of less diversity of tasks in FLAN, we need more diversity
d)	What does the reward-hacking issue entail in RLHF? What potential solutions could address it? (2 marks)
	Over-optimization, getting rewards by hook or crook, Solution: KL Penalty
e)	Why doesn't Direct Preference Optimization necessitate reinforcement learning? What issues of RLHF does it address?. (2 marks)
	LLMs learns to assign more probability to positive samples and and less possibility to negative samples. IT is more stable and roust than RLHF and eaisier to train



With max-probability based decoding, if it generates more probable tokens multiples times, it becomes too confindent in generating the same tokens again -agian

Sol: -don't repeat n-grams -contrastive decodeing -unlikelyhood objective -Coverage loss

Que 4: What advantages does instruction fine-tuning offer? What potential challenges may arise with instruction fine-tuning? (3 marks)

Align models to natural language instructions, task unification, reasoning capabilities etc. Challenges: Behaviour cloning, human annotation

Que 5: Compare the following: a) TagLM vs ELMo (2 marks)

TagLM Uses top layer Istms, ELMo uses all Istm layers

b) BERT vs T5 (2 marks)

Difference in pretraining objective, encoder vs encoderdecoder, multitask setup in t5

c) FLAN instruction fine-tuning vs SFT instruction fine-tuning (2 marks)

Flan: academic tasks, less diversity, human annotated SFT: Trained on human prompts, models preference, RLHF, more diverse

pre-training objec	tives for both	

Que 6: Spot the odd one out:

d) PEGASUS vs BART (2 marks)

- a) TagLM, ELMo, BERT, GPT (1 marks)
- b) In-context learning, DPO, Instruction fine-tuning, RLHF (1 marks) In-context learning
- c) PEFT, LoRA, QLoRA, Adapter fine-tuning (1 marks) QLoRA

Que 7: The ReLU activation function, which can become inactive when the input is negative, a phenomenon known as "dying ReLU." A friend proposes using a different activation function, f(z) = max(0.2z, 2z), to solve this saturation issue.

- The question is whether this new function would indeed address the problem. Why or why not? (3 marks)
- What about another activation function g(z) = 1.5z. Would this be a good activation function? Why or why not? (2 marks) NO

Que 8: Explain the following concepts:
a) Double quantization in QLoRA (2 marks)
storing absolute max in less bits, how to do it
b) Need of block-wise quantization (2 marks)
solving outlier problem

	Parametric: Generator (LLM) weights Non-parametric: retrieved info	
Q	Que 9: Explain the following concepts with pictorial depictions:	
a)	a) Speculative Decoding, steps involved, and justification for best-case and worst-case marks)	scenarios.(
	All the steps with diagram, refer the slides.	

c) Parametric vs Non-parametric memory in RAG (2 marks)

b) F	All the steps with diagra similar to in-context documents can be tre	G) and compariso	on with in context le	earning. (4 marks)

c) Low Rank Adaptation (LoRA) and why does it work? (4 marks)

Refer slides, original pretrained weight matrix is not trained enough, can be compressed (rank reduction)

d) Transformer architecture and discussion about motivation behind self-a layer-norm, and residual connections. (5 marks)	attention, multi-head attention,
All the steps with diagram, refer the slides.	