Due date: Friday 27th September 2024, EOD – IST.

Important Instructions for submitting solutions

1. Submit the solution to all questions in the assignment should be submitted in a **single PDF file with not more than 500 words**.
2. Any plagiarism if detected will automatically attract **zero marks** for that assignment.
3. It is preferable if the **text of PDF file can be extracted** through a PDF extractor e.g. PyPDF. For example, pictures of handwritten text are not extractable, whereas PDF generated by MS Word, Latex, etc., are.
4. Exceptionally good solutions with extractable text may receive **special appreciation** from the teachers.

Assignment

1. What is Prompt Engineering?

2. What is prompt injection? What are the different types of prompt injection?

3. What are the key advantages of using Retrieval Augmented Generation (RAG)?

4. What are the essential components of the ReAct prompting framework?

5. What are the main advantages of Dense Retrieval over Sparse Retrieval in RAG systems?

6. An LLM is generating text, and the next word prediction has the following probability distribution:
   Word A: 0.15
   Word B: 0.10
   Word C: 0.45
   Word D: 0.20
   Word E: 0.05
   If you are using top-k sampling with k=3, which words are in the candidate set for the next word?

7. An LLM is generating text, and the next word prediction has the following probability distribution:
   Word A: 0.45
   Word B: 0.25
   Word C: 0.15
   Word D: 0.10
   Word E: 0.05

If we use top-p sampling with p=0.7, what is the smallest candidate set that meets the probability threshold?

8. Which factors could be influenced by the "temperature" setting when interacting with an LLM?

9. What are the main advantages of COT over ZERO-SHOT prompting?

10. What are the advantages of Auto-COT over COT?

11. What are the advantages of meta prompting?

12. A research team is exploring the use of LLMs for creative writing. How should they control temperature and top-k parameters?