Module-1

1. What is a language model? List and explain any two applications of LMs.
2. Explain the architecture of a Feedforward Neural Language Model.
3. What is the difference between CBOW and Skip-gram in Word2Vec?
4. Describe the components of the attention mechanism in Transformers.
5. Define zero-shot and few-shot learning with an example.
6. Given a sentence “I love machine learning”, construct a bigram language model and compute the probability of the sentence assuming made-up probabilities.
7. How does a chatbot use an LSTM-based language model for predicting the next word?
8. Apply the concept of masked language modeling to fill in the blanks in a given sentence using BERT-like modeling logic.
9. Describe the architecture of an LSTM cell with appropriate equations and diagram.
10. Compare and contrast the Transformer and RNN-based models.
11. Explain the encoder-decoder structure in the Transformer architecture with a neat diagram.
12. What are the differences between GPT and BERT language models?
13. Show how a text generation task can be implemented using an LSTM model. Outline the input-output flow and training steps.
14. Construct a trigram language model from a short corpus and apply additive (Laplace) smoothing.
15. Explain how an instruction-tuned model like FLAN-T5 generates a response when prompted with a user query.
16. Explain the concept of Reinforcement Learning with Human Feedback (RLHF) and its role in aligning LLMs like ChatGPT.
17. Describe how pretraining and fine-tuning are used in LLMs. Give examples from BERT or GPT models.
18. Describe the components and functioning of Multi-head Attention in Transformers.
19. Explain how ChatGPT can be used to solve instruction-following tasks in education or customer support.
20. A company wants to create a multimodal AI assistant. Explain how a model like Gemini can be used, integrating vision and language inputs.
21. Consider a sentence-level translation task. How would a Transformer model handle this task from input to output?
22. Discuss the evolution of Large Language Models from GPT-1 to GPT-4. Highlight improvements at each stage.
23. Explain the full architecture of a Transformer with a labeled diagram, covering encoder, decoder, attention, and feedforward layers.
24. Compare LSTM, GRU, and RNN in terms of architecture, vanishing gradient handling, and use cases.
25. Suppose you are designing a question-answering system. Explain how BERT and GPT can be used differently in your application.
26. Design a chatbot using instruction-tuned models like FLAN or Claude. Describe the input pipeline, training strategy, and deployment method.
27. A news summarization system is being developed. Explain how a T5 or BART model can be trained and fine-tuned for this task.

**Module-2**

1. What is prompt engineering? Why is it important in influencing LLM behavior?
2. List and explain any three components of a good prompt.
3. What is soft prompting? How is it different from hard-coded prompts?
4. Define Chain-of-Thought (CoT) prompting with a simple example.
5. What are delimiters in prompt design? Mention their role in LLM responses.
6. Write a well-structured prompt to generate a summary of a news article using an LLM.
7. Create a zero-shot CoT prompt for solving a basic math word problem.
8. Demonstrate how a soft prompt can guide an LLM to produce a poem.
9. Compare direct answer prompting with step-by-step reasoning using CoT prompts.
10. What is prompt tuning? Explain the process involved with a diagram or example.
11. Explain multi-hop reasoning with an example use case. How does it differ from single-hop reasoning?
12. Write a few-shot CoT prompt for a logical reasoning task. Briefly explain how it works.
13. Show how retrieval-augmented generation (RAG) can be used in multi-hop question answering.
14. Given a factual task (e.g., climate data interpretation), write a role-based prompt to guide the model as a climate scientist.
15. Explain the different stages involved in multi-stage prompting. How does it differ from multi-turn dialogue?
16. What is goal-driven prompting? How does it help control LLM behavior? Give one case example.
17. Discuss the concept of self-reflection prompting and how it improves model reasoning.
18. A student wants to write a research paper. Design a multi-stage prompt flow to guide the student through topic selection, literature review, and outline creation.
19. Demonstrate how multi-hop reasoning can be applied in a medical diagnosis chatbot using a real-world dataset example like HotPotQA or BioASQ.
20. Create a case study-based prompt that uses delimiters, role prompting, and CoT to simulate a courtroom argument generation by a virtual lawyer.
21. What is Chain-of-Thought (CoT) prompting? Explain with examples, its inductive bias toward reasoning, and how it improves benchmark performance.
22. Compare and contrast the following prompt engineering strategies: Prompt tuning, soft prompting, and CoT prompting.
23. Describe Advanced Prompt Patterns (role prompting, self-reflection, delimiters, and goal-driven prompting) with suitable examples.
24. Design a multi-stage, goal-driven prompt strategy for an AI legal assistant tasked with answering user legal queries. Highlight how the prompts evolve across stages.
25. Write a complete walkthrough showing how self-ask + RAG techniques are combined for multi-hop reasoning in a real-world context like financial fraud detection.
26. Create a chain-of-thought prompt strategy to solve high-school level physics problems, and explain how few-shot examples boost performance.

Module-3

1. What is benchmarking in prompt engineering? Why is it important?
2. List and explain any three characteristics of a good prompt.
3. Compare prompt-based evaluation with fine-tuning evaluation.
4. Define self-consistency in prompt evaluation. When is it useful?
5. What is prompt calibration? Mention two techniques used for it.
6. Provide sample prompts for a classification and a QA task using benchmark datasets.
7. Design a prompt for a reasoning task and explain how self-consistency can be applied.
8. Explain how accuracy and F1-score would be used to evaluate prompts in a named entity recognition (NER) task.
9. Explain the importance of using standard benchmark datasets in prompt evaluation.
10. Discuss various categories of prompt-based tasks and how task goals influence prompt design.
11. Define and compare BLEU, ROUGE, and Perplexity as metrics for evaluating text generation tasks.
12. Create a prompt for a text generation task and describe how you would evaluate its fluency, coherence, and relevance using human evaluation.
13. Given a QA benchmark like SQuAD, write a few-shot prompt and describe how accuracy and log-likelihood metrics could be used to assess its performance.
14. Demonstrate how re-ranking outputs can improve prompt-based task accuracy for a multi-choice exam question.
15. Explain how self-consistency improves performance over greedy decoding in CoT prompts.
16. Discuss common challenges in designing benchmark prompts and how benchmarks guide design improvements.
17. Describe calibration techniques in prompt engineering and explain how they help manage uncertainty and hallucination.
18. Design prompts for a math reasoning benchmark (e.g., GSM8K) and explain how self-consistency evaluation helps improve answer quality.
19. Create two prompts (for classification and QA tasks) and explain which evaluation metrics you would use and why.
20. Apply temperature control and top-p sampling to a sensitive prompt generation scenario like medical advice or legal suggestion. Explain how calibration improves response safety.
21. Describe in detail the **consolidated prompt evaluation workflow**, from prompt creation to evaluation and refinement, including key metrics.
22. Explain the role and benefits of benchmark datasets in prompt engineering. Include examples across classification, QA, and reasoning.
23. Discuss the role of task framing in prompt engineering. How does it affect different types of goals such as generation, classification, or reasoning?
24. A company is testing different prompts for a customer support bot. Design a benchmarking and evaluation workflow using human and model-level metrics to compare performance.
25. Analyze how CoT-specific benchmarks improve LLM reasoning. Show a case study involving multi-hop factual QA, explaining how metrics like accuracy and log-likelihood are used.
26. You are asked to evaluate prompts for instruction-following in an educational app. Develop a multi-stage evaluation using self-consistency, human evaluation, and calibration techniques.

**Module-4**

1. What is AI ethics? Explain any two core principles from FATS.
2. Define narrow and general alignment in AI. Give one example each.
3. What is hallucination in large language models?
4. List any three sources of bias in LLMs.
5. What is model inversion? How does it threaten privacy?
6. Explain how fairness metrics can be applied to audit bias in a chatbot.
7. How can GDPR influence the design of AI systems using personal data?
8. Suggest two strategies to mitigate hallucination in generative models used for healthcare advice.
9. Explain the concept of AI alignment and its importance in AI development.
10. Discuss prompt injection attacks with an example.
11. Compare misinformation and disinformation, and their implications on democratic societies.
12. Suppose a generative AI is used in schools. Identify two ethical risks and propose mitigation strategies.
13. Write a short case study showing how transparency and auditability were implemented in a government AI project.
14. If a language model is trained on biased legal texts, how can this affect its fairness? Suggest ways to correct it.
15. Discuss the core principles of Responsible AI. How do model cards and datasheets support these principles?
16. What are the limitations of Reinforcement Learning with Human Feedback (RLHF)?
17. Describe the economic impact of LLMs on white-collar jobs. Mention both risks and opportunities.
18. You're deploying a generative AI in a healthcare setting. Explain how to ensure privacy and safety using legal frameworks like HIPAA or India’s DPDP Act.
19. Describe how human-in-the-loop (HITL) systems can be used to improve trust and reduce hallucination in AI-generated content.
20. Design a governance strategy for a company using LLMs in customer service, including roles of ethics boards and transparency tools.
21. Explain in detail the challenges of AI alignment and how RLHF addresses them. Include human feedback loops and limitations.
22. Compare and contrast the EU AI Act, UNESCO AI Ethics Guidelines, and OECD AI Principles.
23. What are the societal and cultural risks of deploying LLMs globally? Discuss ethical considerations in multilingual and multicultural settings.
24. Develop a case study on the misuse of generative AI in elections (e.g., deepfakes or disinformation). Analyze ethical, legal, and technical interventions.
25. You’re part of a policy team framing AI ethics in education. Create a guideline covering misinformation, student overreliance, and fairness.
26. Design a framework for responsible AI development for a startup building LLM-based tools for legal professionals. Address transparency, accountability, and bias mitigation.

**Module-5**

1. What is domain-specific prompting? How is it different from general prompting?
2. List four types of tasks supported by prompt engineering with one example each.
3. What is prompt drift? Why is it a concern in production systems?
4. Describe any two prompt adaptation techniques used for domain-specific tasks.
5. What is the role of tools like GitHub Copilot or Amazon CodeWhisperer in programming prompt engineering?
6. Give an example of a role-based prompt for a legal assistant analyzing a case brief.
7. Write a basic prompt for a summarization task in the medical domain using PubMed content.
8. How would you adapt a prompt to generate educational content for a 6th-grade student in mathematics?
9. Explain the structure and lifecycle of professional prompt engineering in domain-specific contexts.
10. Discuss the ethical challenges in applying prompt engineering in education.
11. Highlight three major challenges in medical domain prompting and suggest mitigation techniques.
12. Design a prompt to generate legal advice for a contract dispute using Harvey AI or Casetext.
13. You're developing a chatbot for journalism. Draft a news prompt and explain the challenges in controlling tone and fact consistency.
14. Compare the use of GPT-based tools like FinGPT and BloombergGPT for financial forecasting. Create a scenario-based application for one.
15. What are the main challenges faced while deploying prompt-based systems across domains? Discuss prompt drift, versioning, and memory limitations.
16. Describe how tone, structure, and personalization are adapted in prompts for EdTech platforms.
17. Explain the concept of multimodal prompts and their significance in future domain-specific applications.
18. Create a prompt chain for a healthcare chatbot diagnosing symptoms, generating a summary, and suggesting next steps.
19. You're tasked with designing prompts for a programming tutor. Use instruct and chain-of-thought formats to explain recursion.
20. Using Khanmigo, describe how prompts can be personalized for a high school student preparing for competitive exams.
21. Compare and contrast prompt engineering applications across three domains: Law, Medicine, and Computer Science. Focus on task types, tools, and challenges.
22. Discuss the future trends in domain-specific prompt engineering such as auto-prompting, prompt agents, and context-aware systems.
23. Explain the complete lifecycle of prompt engineering in a regulated domain (e.g., finance or healthcare), covering design, testing, deployment, and monitoring.
24. Design a cross-domain AI agent that can switch between legal, financial, and educational queries. Outline the prompt structuring logic, role adaptation, and deployment concerns.
25. Create a case study describing the use of FinGPT for interpreting tabular data and making investment suggestions. Include compliance, bias, and transparency challenges.
26. Propose a system for a journalism organization that uses AI to draft, verify, and publish news articles using Newsroom GPT and AI21 Studio. Describe prompt challenges and risk controls.