1. According to the syllabus, what is the full form of the BERT model?
A. Bidirectional Encoding and Retrieval Transformer
B. Bidirectional Encoder Representations from Transformers
C. Binary Encoder Reconstruction Transformer
D. Bidirectional English Representations from Transformers

2. Which architecture is utilized by the GPT model?
A. Encoder-only Transformer
B. Encoder-Decoder Transformer
C. Decoder-only Transformer
D. Multi-stack Encoder

3. What is the primary training objective of the BERT model?
A. Next-token prediction
B. Reinforcement Learning from Human Feedback
C. Masked Language Modeling (MLM) and Next Sentence Prediction (NSP)
D. Causal self-attention modeling

4. Which type of attention mechanism is used in GPT?
B. Bidirectional self-attention
A. Causal (masked) self-attention
C. Cross-attention only
D. Sparse encoder attention

5. In terms of context direction, how does BERT process information?
A. Left-to-right only
B. Right-to-left only
C. Left + Right context
D. Unidirectional only

6. Which model is classified as "Autoregressive" according to the syllabus?
A. BERT
B. GPT
C. Transformer Encoder
D. NSP Model

7. What is the representational output type of the GPT model?
A. Representational
B. Generative
C. Discriminative
D. Classification-based

8. For which of the following use cases is BERT considered most suitable?
A. Creative writing
B. Chatbots
C. Named Entity Recognition (NER)
D. Code generation

9. What is the specific training objective for GPT?
A. Predict masked tokens
B. Predict sentence relations
C. Predict next token
D. Image-to-text alignment

10. GPT-4 is described as a Transformer-based architecture that builds upon which previous version?
A. GPT-1
B. GPT-2
C. GPT-3
D. BERT-Large

11. Which technique is used to fine-tune GPT-4 to ensure safer outputs and alignment?
A. Masked Language Modeling
B. Self-supervised learning
C. RLHF (Reinforcement Learning from Human Feedback)
D. Causal inference

12. What does the "Pronoun Resolution Attention Pattern" demonstrate in a Transformer?
A. Syntax awareness of verbs
B. Linking a pronoun to its antecedent (e.g., "she" to "student")
C. Long-range numerical dependencies
D. Next-sentence prediction accuracy

13. According to the syllabus, GPT-4 is "multimodal." What types of data was it trained on?
A. Text + Audio
B. Text + Image
C. Image + Video
D. Text only

14. Which feature is highlighted as an improvement in GPT-4 over GPT-3?
A. Reduced hallucinations
B. Switch to Encoder-only architecture
C. Use of bidirectional attention
D. Removal of feed-forward layers

15. The "Syntax Awareness" attention pattern specifically highlights which grammatical relationship?
A. Pronoun-antecedent resolution
B. Subject-verb agreement
C. Masked token prediction
D. Image-text alignment

16. What are the core components used by GPT-4 to efficiently process long sequences?
A. Bidirectional attention and MLM
B. Multi-head self-attention and deep feed-forward layers
C. Next sentence prediction and layer normalization

D. Causal attention and NSP

17. Which of the following is a best use case for GPT?
A. Search ranking
B. Document classification
C. Conversational AI
D. Sentiment analysis

18. What is the representational type of BERT?
A. Generative
B. Autoregressive
C. Representational
D. Multimodal

19. In the context of attention visualization, what does a pattern showing grammatical relationships betwee
A. The model ignores sentence structure
B. The Transformer learns sentence structure
C. The model is using next-token prediction
D. The model is performing RLHF

20. Transformers rely primarily on which mechanism?
A. Recurrent neural networks
B. Convolutional layers
C. Self-attention
D. Linear regression

21. Which model is best suited for "Sentiment analysis" and "Search ranking"?
A. GPT-4
B. GPT
C. BERT
D. Claude

22. GPT-4 uses which type of learning during its pretraining phase on massive datasets?
A. Supervised learning
B. Self-supervised learning
C. Reinforcement learning
D. Manual labeling

23. Which of the following is NOT a unique feature of GPT-4 mentioned in the syllabus?
A. Strong reasoning abilities
B. Multimodal understanding
C. Encoder-only architecture
D. Better factual accuracy

24. What architectural component is used in GPT-4 alongside multi-head self-attention and deep feed-forw
A. Bidirectional masking
B. Layer normalization

C. Recursive feedback loops
D. Sentiment classifiers

25. According to the syllabus, what is a best use case for GPT-4?
A. Named Entity Recognition
B. Document classification
C. Code assistants
D. Masked token prediction
1. What does the self-attention mechanism capture that traditional RNNs typically struggle with?
A. Short-term memory
B. Long-distance relationships
C. Linear data sequences
D. Tokenization speed

2. In the sentence "Although the project was difficult, the team completed it successfully," which mechanis
A. Linear regression
B. Recurrent processing
C. Self-attention
D. Data preprocessing

3. According to the syllabus, what is the primary function of fine-tuning?
A. To train a model from scratch
B. To adapt a pre-trained model to a specific domain
C. To replace the tokenization process
D. To increase the size of the model's parameters

4. Which of the following is an advantage of fine-tuning mentioned in the text?
A. It eliminates the need for evaluation
B. It makes AI systems more accurate and context-aware
C. It allows models to ignore local expressions
D. It reduces the need for large language models

5. Where does fine-tuning fit in the structured training pipeline?
A. Before data collection
B. During the tokenization phase
C. After pretraining
D. Before preprocessing

6. Which of the following is NOT a step in the structured training pipeline mentioned?
A. Data collection
B. Feature extraction
C. Tokenization
D. Evaluation

7. Why do general-purpose models often fail in content moderation for low-resource languages?
A. They are too large to process the data
B. They fail to capture cultural context, slang, and local expressions

C. They cannot be fine-tuned
D. They prioritize Nepali over other languages

8. For which specific low-resource language does the text suggest fine-tuning for content moderation?
A. Hindi
B. Sanskrit
C. Nepali
D. Tibetan

9. In the training pipeline, what does the model learn during the fine-tuning phase instead of learning langu
A. General intelligence
B. Task-specific patterns
C. Tokenization rules
D. Ethical guidelines

10. According to the syllabus, what is a transformer-based model like BERT particularly useful for?
A. Text classification
B. Reinforcement learning
C. Hardware acceleration
D. Unsupervised pretraining

11. What is the recommended use for GPT-style models in a fine-tuned context?
A. Data preprocessing
B. Controlled text generation
C. Human-verified labeling
D. Identifying abusive intent

12. What is the benefit of using curated datasets with human-verified labels?
A. Increased model size
B. Reduced bias and hallucination issues
C. Faster tokenization
D. Elimination of the pretraining phase

13. Which technique is mentioned as a way to further align a model with ethical guidelines?
A. Linear scaling
B. Instruction tuning
C. Manual preprocessing
D. Recurrent attention

14. What does the acronym RLHF stand for in the context of model alignment?
A. Random Linear Heuristic Function
B. Reinforcement Learning from Human Feedback
C. Recurrent Logic High Frequency
D. Refined Language Handling Framework

15. In the domain of Educational AI, what should a model be trained on to provide accurate explanations?
A. Social media slang

B. General-purpose encyclopedias
C. University-level notes and exam patterns
D. Unstructured web data

16. Fine-tuning enables AI systems to transition from "general intelligence" to what?
A. Artificial Super Intelligence
B. Practical, real-world intelligence
C. Basic token recognition
D. Traditional RNN logic

17. As models grow larger, fine-tuning remains essential for which of the following?
A. Eliminating the need for datasets
B. Personalization, safety, and domain relevance
C. Reducing the number of layers in a transformer
D. Moving back to RNN architectures

18. Fine-tuning a model for content moderation helps in identifying which of the following?
A. Pretraining errors
B. Abusive intent or content severity
C. Tokenization bottlenecks
D. RNN long-distance failures

19. Which process involves breaking down data before the training and evaluation phases?
A. RLHF
B. Tokenization
C. Instruction tuning
D. Hallucination reduction

20. How does a fine-tuned educational tutor model compare to a general chatbot?
A. It is slower but more creative
B. It provides more accurate and relevant explanations
C. It uses less domain-specific data
D. It avoids using university-level notes

21. What is the result of applying domain-specific datasets to fine-tuning?
A. Decreased performance in local contexts
B. Significantly improved performance
C. Increased bias and hallucinations
D. Transition to traditional RNNs

22. What is the core limitation of traditional RNNs mentioned in the text?
A. They cannot handle Nepali
B. They struggle with long-distance relationships
C. They require human-verified labels
D. They cannot be used for text classification

23. Which application is specifically mentioned as a domain for specialization?

A. Weather forecasting
B. Hate speech and content moderation
C. Stock market prediction
D. Image recognition

24. Fine-tuning allows a model to learn task-specific patterns, such as categorizing what?
A. Preprocessing steps
B. Content severity
C. Model parameters
D. RNN weights

25. According to the text, fine-tuning is considered one of the most powerful techniques in:
A. Traditional software engineering
B. Modern Generative AI systems
C. Basic data entry
D. Legacy RNN systems

***

Answer Key:
1. B
2. C
3. B
4. B
5. C
6. B
7. B
8. C
9. B
10. A
11. B
12. B
13. B
14. B
15. C
16. B
17. B
18. B
19. B
20. B
21. B
22. B
23. B
24. B
25. B