LLM Assessment: Capabilities, Features, and Tools

Image Generation

1. **Multiple Choice:** Which of the following best describes the difference between FLUX and Gemini 2.5's image generation capabilities?

    - a) FLUX creates photorealistic images while Gemini 2.5 only creates cartoons
    - b) Gemini 2.5 offers more interactive refinement capabilities while FLUX is more one-shot
    - c) Gemini 2.5 can generate video while FLUX cannot
    - d) FLUX is only available in specific regions while Gemini 2.5 is global

Correct Answer: b) Gemini 2.5 offers more interactive refinement capabilities while FLUX is more one-shot

2. **True/False:** Both FLUX and Gemini 2.5 use natural language prompts as input for image generation.

    - a) True
    - b) False

Correct Answer: a) True

3. **Multiple Choice:** What is a key limitation in current image generation models like FLUX?

    - a) They cannot generate images of people
    - b) They often produce artifacts in complex scenes
    - c) They can only work with square aspect ratios
    - d) They require programming knowledge to use effectively

Correct Answer: b) They often produce artifacts in complex scenes

Artifacts in LLMs

4. **Multiple Choice:** What is the primary purpose of artifacts in LLM interactions?

    - a) To store user data between sessions
    - b) To create and manage self-contained content that can be referenced and updated
    - c) To improve rendering speed of text responses
    - d) To track conversation history for billing purposes

Correct Answer: b) To create and manage self-contained content that can be referenced and updated

5. **True/False:** Artifacts can be modified by both the LLM and the user throughout a conversation.
    - a) True
    - b) False

Correct Answer: a) True

6. **Multiple Choice:** Which of these is NOT a common type of artifact in modern LLM systems?
    - a) Code artifacts
    - b) Document artifacts
    - c) SVG graphics
    - d) Video artifacts

Correct Answer: d) Video artifacts

7. **Multiple Choice:** How do LLMs typically use artifacts in creative writing tasks?
    - a) They can edit and improve existing artifacts based on user feedback
    - b) They automatically generate artifacts without user input
    - c) They can only create artifacts of fewer than 100 words
    - d) They require external APIs to generate artifacts

Correct Answer: a) They can edit and improve existing artifacts based on user feedback

System vs. User Prompts

8. **True/False:** System prompts are visible to the end user in most commercial LLM applications.
    - a) True
    - b) False

Correct Answer: b) False

9. **Multiple Choice:** What is the primary difference between system prompts and user prompts?
    - a) System prompts are written in code while user prompts use natural language
    - b) System prompts help define the LLM's behavior and constraints while user prompts are specific requests
    - c) System prompts are optional while user prompts are required
    - d) System prompts only work for certain types of questions

Correct Answer: b) System prompts help define the LLM's behavior and constraints while user prompts are specific requests

10. **Multiple Choice:** Which of the following would most likely be part of a system prompt rather than a user prompt?

    - a) "Write a poem about autumn"
    - b) "Summarize this article"
    - c) "Always cite sources and avoid making up facts"
    - d) "How do I fix this Python error?"

Correct Answer: c) "Always cite sources and avoid making up facts"

Vibe Coding

11. **True/False:** Vibe coding refers to assigning numerical values to represent emotional tones in text.

    - a) True
    - b) False

Correct Answer: b) False

12. **Multiple Choice:** Which best describes how vibe coding works in LLM interactions?

    - a) It uses special symbols to trigger different emotional responses
    - b) It involves writing code that makes interfaces more visually appealing
    - c) It uses natural language indicators to guide the style and tone of LLM responses
    - d) It requires specialized programming languages only available to AI researchers

Correct Answer: c) It uses natural language indicators to guide the style and tone of LLM responses

Open Source LLMs

13. **Multiple Choice:** What does it mean that DeepSeek is "open source"?

    - a) Anyone can use it for free without restrictions
    - b) Its code and weights are publicly available for others to inspect, modify, and use
    - c) It can only process open data formats
    - d) It was developed collaboratively by multiple companies

Correct Answer: b) Its code and weights are publicly available for others to inspect, modify, and use

14. **True/False:** Open source LLMs like DeepSeek always perform worse than closed source models.

- a) True
- b) False

Correct Answer: b) False

Multimodal LLMs

15. **Multiple Choice:** Which of these capabilities is NOT typically associated with multimodal LLMs?

    - a) Understanding and describing images
    - b) Processing audio input
    - c) Physical manipulation of objects
    - d) Analyzing charts and graphs

Correct Answer: c) Physical manipulation of objects

16. **True/False:** Multimodal LLMs can understand the relationship between text and images in a single context.

    - a) True
    - b) False

Correct Answer: a) True

17. **Multiple Choice:** What is a key advantage of multimodal LLMs over text-only models?

    - a) They always provide more accurate responses
    - b) They can process and reason across different types of media
    - c) They require less computing power
    - d) They are always open source

Correct Answer: b) They can process and reason across different types of media

LLM Tools

18. **Multiple Choice:** In the context of LLMs, what are "tools" primarily used for?

    - a) Debugging model errors
    - b) Extending the model's capabilities beyond its training
    - c) Reducing inference costs
    - d) Simplifying prompt engineering

Correct Answer: b) Extending the model's capabilities beyond its training

19. **True/False:** LLM tools require specialized programming knowledge to implement.

    - a) True
    - b) False

Correct Answer: b) False

20. **Multiple Choice:** Which of the following is an example of a common tool integration with LLMs?

    - a) Weather forecasting
    - b) Password management
    - c) Physical robotics control
    - d) Cryptocurrency mining

Correct Answer: a) Weather forecasting

Vector Embeddings

21. **True/False:** Vector embeddings represent words as points in multidimensional space where semantic similarity correlates with spatial proximity.

    - a) True
    - b) False

Correct Answer: a) True

22. **Multiple Choice:** How are vector embeddings primarily used with LLMs?

    - a) To compress model size for faster inference
    - b) To enable better understanding of relationships between concepts
    - c) To translate between different languages
    - d) To reduce the risk of hallucinations

Correct Answer: b) To enable better understanding of relationships between concepts

Semantic vs. Keyword Search

23. **Multiple Choice:** What is the key difference between semantic search and keyword search?

    - a) Semantic search requires more computing resources
    - b) Keyword search finds exact word matches while semantic search understands meaning
    - c) Semantic search only works with certain languages
    - d) Keyword search is always more accurate for technical queries

Correct Answer: b) Keyword search finds exact word matches while semantic search understands meaning

24. **True/False:** Semantic search is inherently more accurate than keyword search for all types of queries.
    - a) True
    - b) False

Correct Answer: b) False

Chat Context Limitations

25. **Multiple Choice:** Why do chat conversations with LLMs typically have context length limitations?
    - a) To reduce the cost of providing the service
    - b) Because longer conversations lead to more hallucinations
    - c) Due to memory constraints and computational complexity
    - d) To encourage users to start new conversations more frequently

Correct Answer: c) Due to memory constraints and computational complexity

26. **True/False:** When an LLM reaches its context limit, it automatically deletes the oldest parts of the conversation.
    - a) True
    - b) False

Correct Answer: b) False

Additional Questions

27. **Multiple Choice:** Which statement about LLM-generated images is most accurate?
    - a) They are always completely original creations
    - b) They may contain artifacts that reveal their AI-generated nature
    - c) They can perfectly replicate copyrighted images
    - d) They can only be created in standard aspect ratios

Correct Answer: b) They may contain artifacts that reveal their AI-generated nature

28. **Multiple Choice:** What happens when providing feedback on LLM responses?
    - a) The model immediately learns from your feedback
    - b) Your feedback is collected to potentially improve future versions

- c) The model adjusts its parameters in real-time
- d) Feedback is only used for billing purposes

Correct Answer: b) Your feedback is collected to potentially improve future versions

29. **True/False:** Modern LLMs can remember information from past conversations after the chat session has ended.

    - a) True
    - b) False

Correct Answer: b) False

30. **Multiple Choice:** Which of the following best describes how an LLM processes a prompt?

    - a) It searches its training data for exact matches
    - b) It generates a response based on statistical patterns from its training
    - c) It connects to the internet to find the most recent information
    - d) It asks other AI systems for help

Correct Answer: b) It generates a response based on statistical patterns from its training

I'll add some questions specifically about reasoning models and how they work:

Additional Questions on Reasoning Models

31. **Multiple Choice:** What is the primary characteristic that distinguishes a "reasoning model" from other types of LLMs?

    - a) Reasoning models can only process numerical data
    - b) Reasoning models have an extended thinking mode for more complex problems
    - c) Reasoning models can only be accessed through specialized APIs
    - d) Reasoning models are always larger in parameter count than standard LLMs

Correct Answer: b) Reasoning models have an extended thinking mode for more complex problems

32. **True/False:** Reasoning models like Claude 3.7 Sonnet always use their extended thinking capabilities for every query they receive.

    - a) True
    - b) False

Correct Answer: b) False

33. **Multiple Choice:** How does the "extended thinking mode" in reasoning models typically work?

- a) It accesses external knowledge bases not available to standard models
- b) It allocates additional time and computational resources to process complex problems
- c) It only activates when specifically requested by developers
- d) It uses specialized hardware acceleration unavailable to regular models

Correct Answer: b) It allocates additional time and computational resources to process complex problems

34. **Multiple Choice:** Which types of tasks generally benefit most from reasoning models compared to standard LLMs?

- a) Simple fact retrieval and basic information queries
- b) Creative writing and storytelling
- c) Complex problem-solving involving multiple steps or logical deduction
- d) Text summarization of short documents

Correct Answer: c) Complex problem-solving involving multiple steps or logical deduction

35. **True/False:** The reasoning capabilities in models like Claude 3.7 Sonnet require users to input their queries using special formatting or syntax.

- a) True
- b) False

Correct Answer: b) False

36. **Multiple Choice:** What is a key technique often employed by reasoning models to improve their problem-solving abilities?

- a) They only use pre-programmed decision trees
- b) They employ chain-of-thought processes to break down complex problems
- c) They can only reason about topics present in their training data
- d) They rely exclusively on retrieving solutions from memory

Correct Answer: b) They employ chain-of-thought processes to break down complex problems

37. **Multiple Choice:** When using a reasoning model like Claude 3.7 Sonnet in extended thinking mode, what should users typically expect?

- a) Higher costs but no improvement in answer quality
- b) Slightly longer response times but potentially more thorough and accurate answers
- c) Immediate responses that are identical to standard mode

- d) The ability to solve problems that are fundamentally unsolvable by AI

Correct Answer: b) Slightly longer response times but potentially more thorough and accurate answers

38. **True/False:** Reasoning models can solve mathematical problems beyond the capabilities of standard LLMs without using external tools.

- a) True
- b) False

Correct Answer: b) False:

DataFrames

10. Consider the following code:

    What will be printed from this code?

```
import pandas as pd
data = {
    'student': ['Emma', 'James', 'Sofia'],
    'score': [85, 92, 88],
    'subject': ['Math', 'Science', 'Math']
}
df = pd.DataFrame(data)
print("Original DataFrame:")
print(df)
df = df.drop('score', axis=1)
df.iloc[2, 1] = 'Art'
print("\nModified DataFrame:")
print(df)
```

11. Given the following DataFrame:

```
import pandas as pd
data = {
    'student': ['Emma', 'James', 'Sofia'],
    'score': [85, 92, 88]
}
df = pd.DataFrame(data)
```

    Write code to add a new column called ‘status’ with values [‘Fail’, ‘Pass’, ‘Pass’] and print the resulting DataFrame.