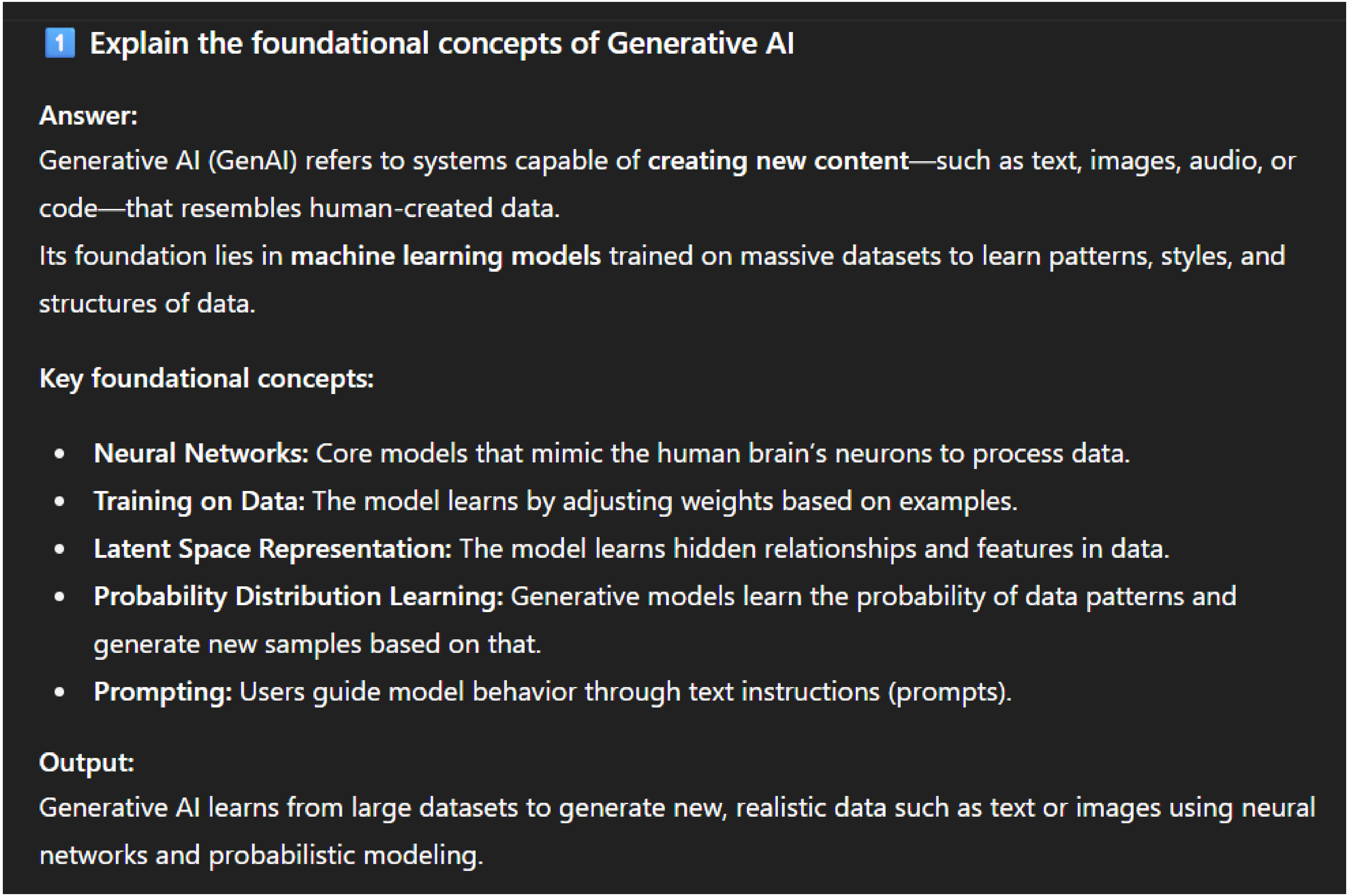


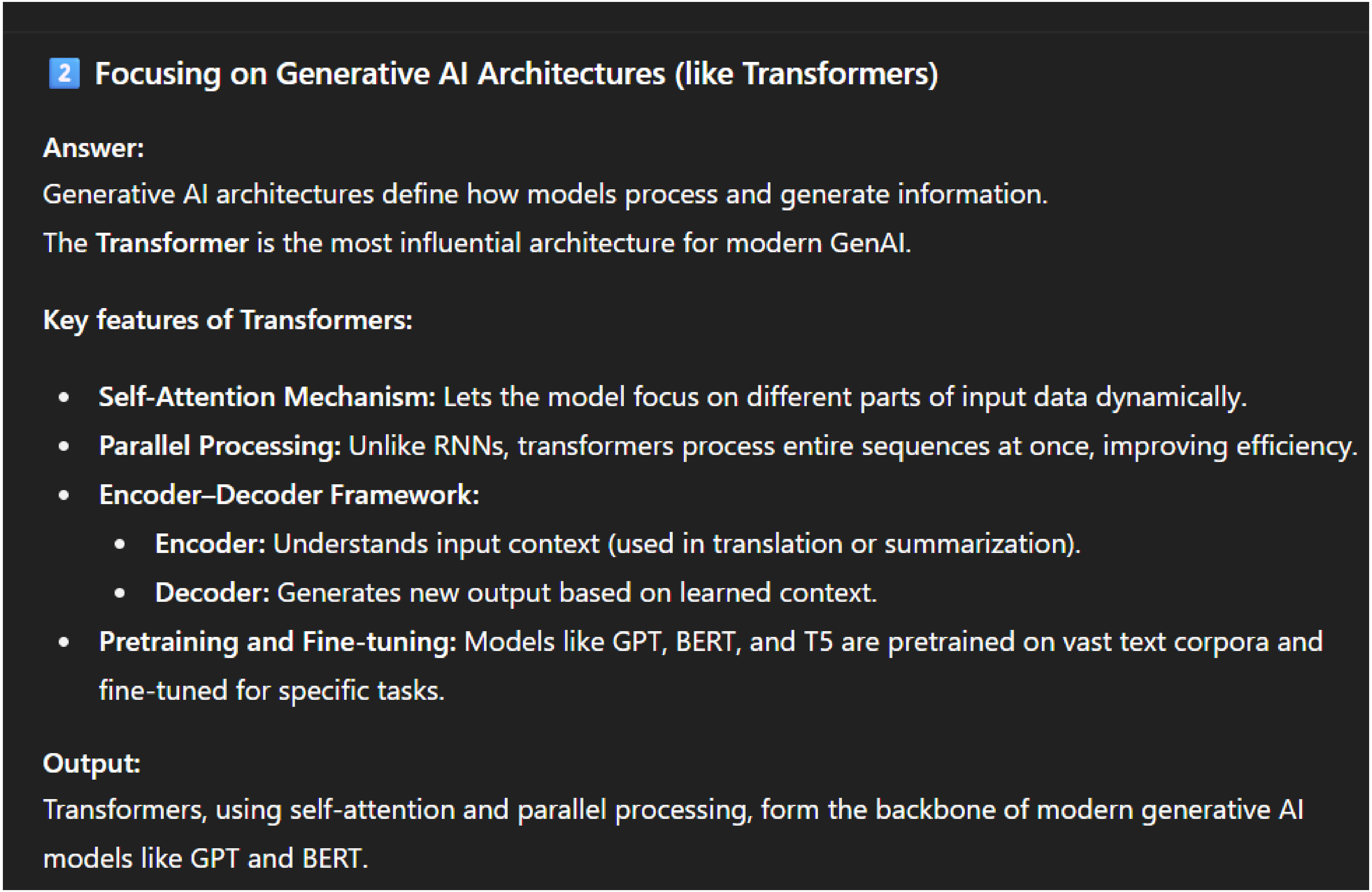
1️⃣ Explain the foundational concepts of Generative AI

Answer:
Generative AI (GenAI) refers to systems capable of **creating new content**—such as text, images, audio, or code—that resembles human-created data.
Its foundation lies in **machine learning models** trained on massive datasets to learn patterns, styles, and structures of data.

Key foundational concepts:

- **Neural Networks:** Core models that mimic the human brain's neurons to process data.
- **Training on Data:** The model learns by adjusting weights based on examples.
- **Latent Space Representation:** The model learns hidden relationships and features in data.
- **Probability Distribution Learning:** Generative models learn the probability of data patterns and generate new samples based on that.
- **Prompting:** Users guide model behavior through text instructions (prompts).

Output:
Generative AI learns from large datasets to generate new, realistic data such as text or images using neural networks and probabilistic modeling.



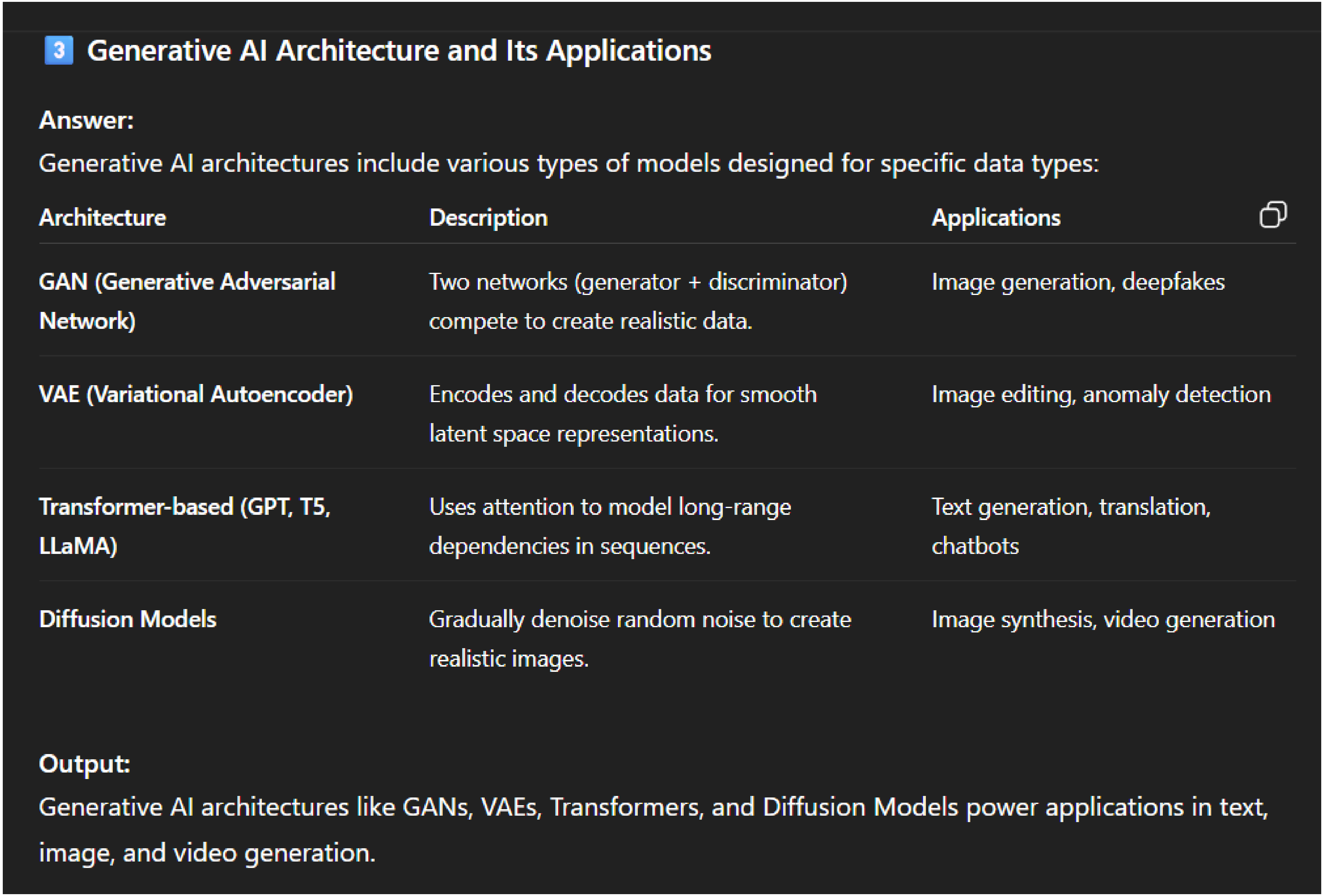
2️⃣ Focusing on Generative AI Architectures (like Transformers)

Answer:
Generative AI architectures define how models process and generate information.
The **Transformer** is the most influential architecture for modern GenAI.

Key features of Transformers:

- **Self-Attention Mechanism:** Lets the model focus on different parts of input data dynamically.
- **Parallel Processing:** Unlike RNNs, transformers process entire sequences at once, improving efficiency.
- **Encoder–Decoder Framework:**
    - **Encoder:** Understands input context (used in translation or summarization).
    - **Decoder:** Generates new output based on learned context.
- **Pretraining and Fine-tuning:** Models like GPT, BERT, and T5 are pretrained on vast text corpora and fine-tuned for specific tasks.

Output:
Transformers, using self-attention and parallel processing, form the backbone of modern generative AI models like GPT and BERT.



3️⃣ Generative AI Architecture and Its Applications

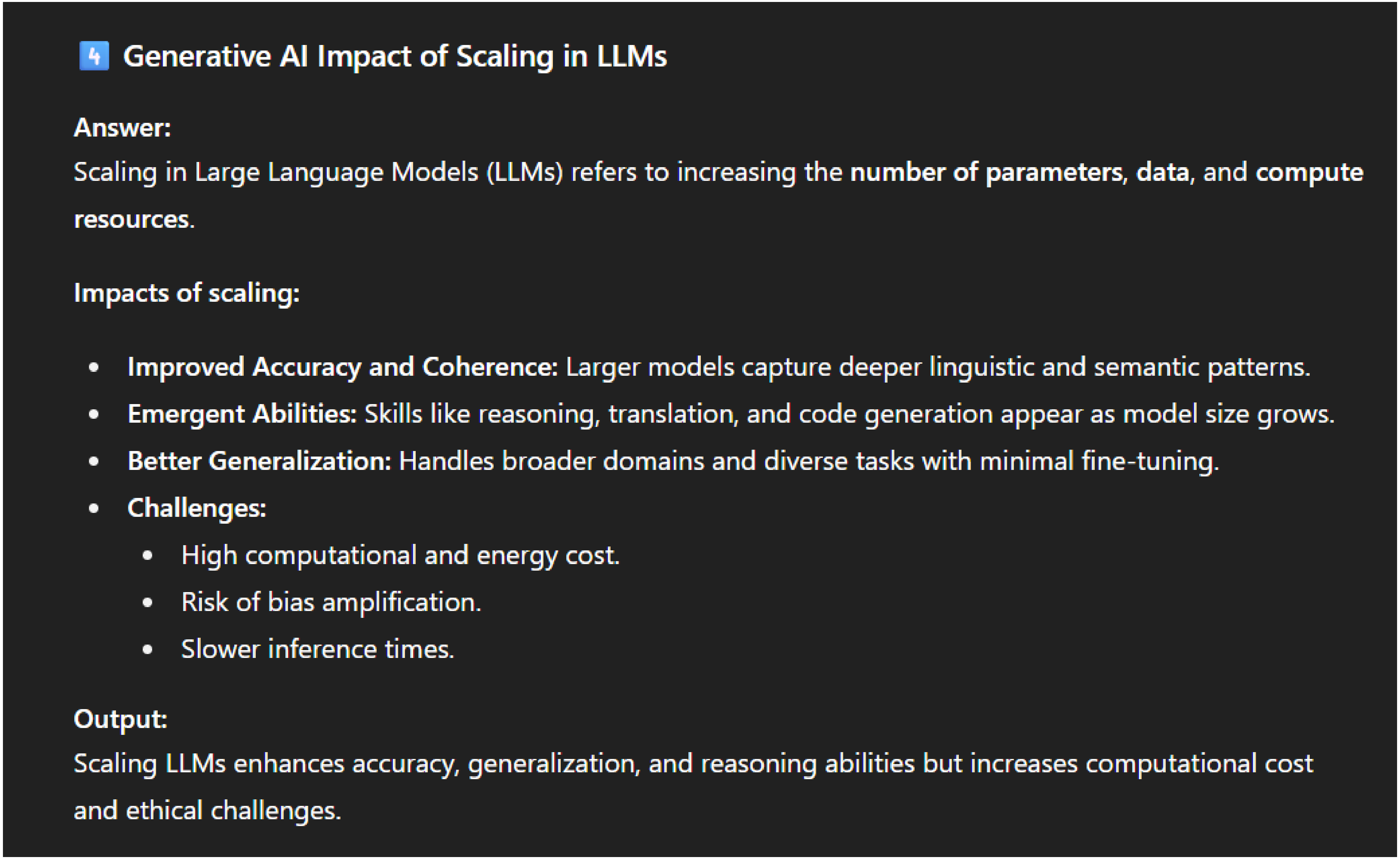
Answer:

Generative AI architectures include various types of models designed for specific data types:

Architecture	Description	Applications
GAN (Generative Adversarial Network)	Two networks (generator + discriminator) compete to create realistic data.	Image generation, deepfakes
VAE (Variational Autoencoder)	Encodes and decodes data for smooth latent space representations.	Image editing, anomaly detection
Transformer-based (GPT, T5, LLaMA)	Uses attention to model long-range dependencies in sequences.	Text generation, translation, chatbots
Diffusion Models	Gradually denoise random noise to create realistic images.	Image synthesis, video generation

Output:

Generative AI architectures like GANs, VAEs, Transformers, and Diffusion Models power applications in text, image, and video generation.



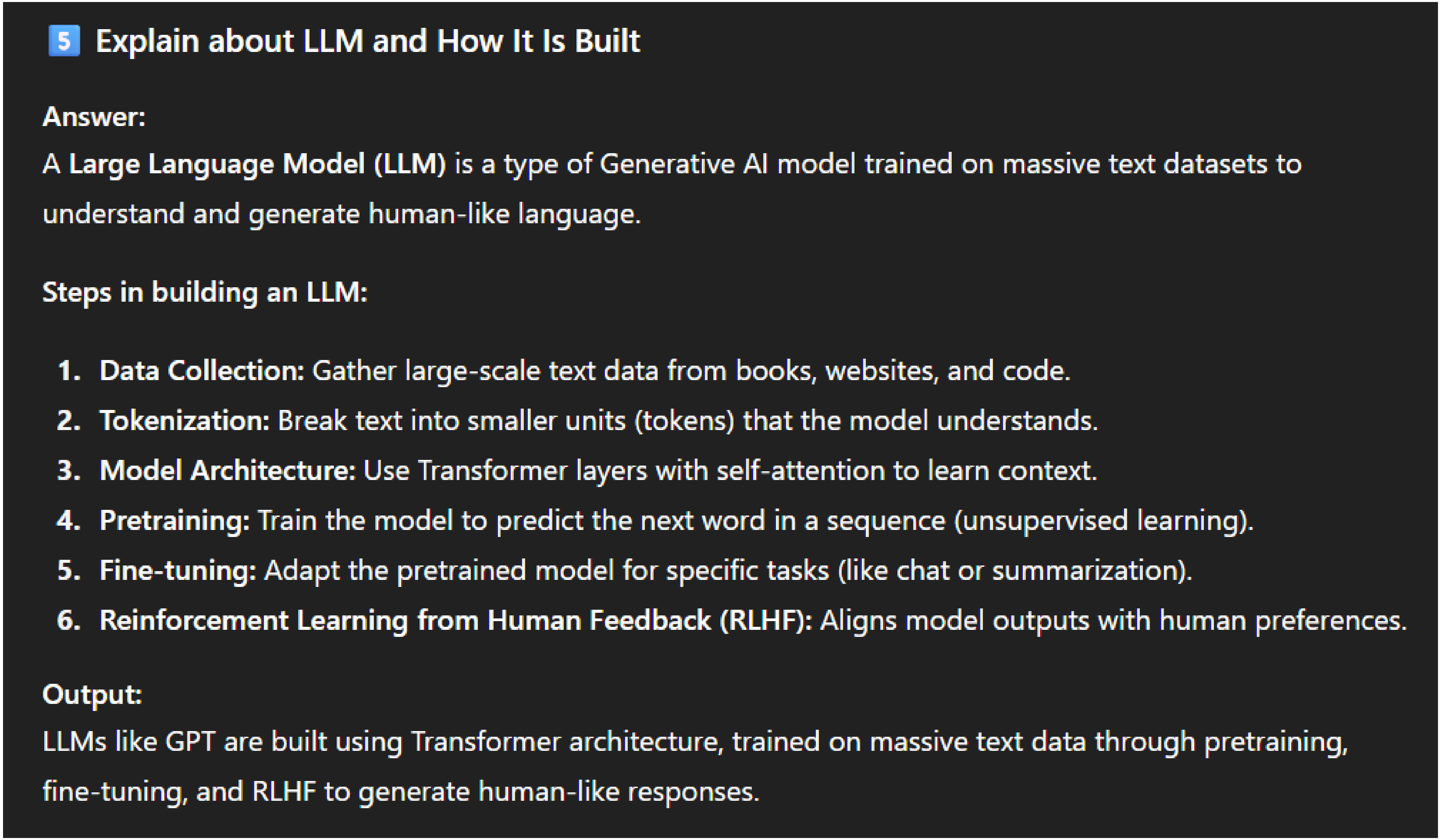
4️⃣ Generative AI Impact of Scaling in LLMs

Answer:
Scaling in Large Language Models (LLMs) refers to increasing the **number of parameters**, **data**, and **compute resources**.

Impacts of scaling:

- **Improved Accuracy and Coherence:** Larger models capture deeper linguistic and semantic patterns.
- **Emergent Abilities:** Skills like reasoning, translation, and code generation appear as model size grows.
- **Better Generalization:** Handles broader domains and diverse tasks with minimal fine-tuning.
- **Challenges:**
  - High computational and energy cost.
  - Risk of bias amplification.
  - Slower inference times.

Output:
Scaling LLMs enhances accuracy, generalization, and reasoning abilities but increases computational cost and ethical challenges.



5️⃣ Explain about LLM and How It Is Built

Answer:
A **Large Language Model (LLM)** is a type of Generative AI model trained on massive text datasets to understand and generate human-like language.

Steps in building an LLM:

1. **Data Collection:** Gather large-scale text data from books, websites, and code.
2. **Tokenization:** Break text into smaller units (tokens) that the model understands.
3. **Model Architecture:** Use Transformer layers with self-attention to learn context.
4. **Pretraining:** Train the model to predict the next word in a sequence (unsupervised learning).
5. **Fine-tuning:** Adapt the pretrained model for specific tasks (like chat or summarization).
6. **Reinforcement Learning from Human Feedback (RLHF):** Aligns model outputs with human preferences.

Output:
LLMs like GPT are built using Transformer architecture, trained on massive text data through pretraining, fine-tuning, and RLHF to generate human-like responses.