Large Language Model (LLM)

A **Large Language Model (LLM)** is an artificial intelligence model trained on vast amounts of text data to understand and generate human language in a coherent and context-aware way.

What is an LLM?

An LLM uses deep learning, particularly the **transformer** architecture, to process and predict language. It excels at predicting the next word in a sentence, enabling it to perform tasks such as:

- Text generation
- Translation
- Question answering
- Summarization
- Code completion

Training

LLMs are trained on large-scale datasets (books, articles, web pages, code) using supervised and unsupervised methods. Training requires immense computational resources.

Parameter Scale

Models are measured by the number of parameters (internal adjustable weights). Examples:

- GPT-3 → 175 billion parameters
- GPT-4 → unknown exact size (larger and multimodal)
- LLaMA 2 → 7B to 65B parameters

Key Examples

Model	Developer	Year	Notes
GPT-3/4	OpenAI	2020–25	Powers ChatGPT
Claude	Anthropic	2023–25	Emphasis on alignment and safety
Gemini	Google DeepMind	2023–25	Integrated with Google products
LLaMA	Meta	2023–25	Open-source, academic use
Mistral	Mistral.ai	2023–25	Efficient, performant smaller models

Strengths

- Versatile language tasks
- Learns from context

- Multilingual support
- Can be fine-tuned for specific domains

Limitations

- May produce incorrect or biased outputs
- No true understanding (statistical patterns only)
- High computational and energy cost
- Requires oversight in critical applications

Applications

- Chatbots and virtual assistants
- Automated writing and summarization
- Legal and medical drafting support
- Programming assistance
- Research and data extraction