

Introduction to Large Language Models

1. What Are Large Language Models?

- AI models that generate, summarize, predict and understand text.
- Trained on massive datasets.
- Contain hundreds of billions of parameters.
- Produce grammatically correct and context-aware responses.

2. Evolution of LLMs

Three eras of NLP:

① Rule-based + Early ML

- Based on handcrafted rules.
- Weak in complex NLP tasks.

② Neural Network Era

- RNN, LSTM introduced memory + context.

③ Transformers + LLM Era

- Attention mechanism.
- Positional encoding.
- Parallel processing enabled scaling.

3. Foundation vs Fine-Tuned Models

Foundation Models

- Pretrained on massive corpora.
- Perform general tasks.

Fine-Tuned Models

- Customized for specific domains.
- Cheaper and faster.

4. Key Technical Components

Neural Networks

- Layered architecture with neurons.

Transformers

- Attention mechanism.
- Fast parallel token processing.

Attention Mechanism

- Assigns importance to each token.
- Improves context understanding.

5. Training Techniques

- Supervised learning (traditional)
- Unsupervised / Self-supervised learning (LLMs)
- No manual labels required

6. Popular LLMs

Examples:

- BERT, GPT-2, GPT-3
- RoBERTa, DeBERTa
- PaLM, MT-NLG, WuDao 2.0

7. Enterprise Use Cases

- Automation of text-heavy tasks
- Summarization & Chatbots
- Coding assistance
- Document analysis
- Customer support

8. Challenges of LLMs

- Bias and toxic output
- Ethical concerns
- Adversarial attacks
- Cost + Infrastructure requirements
- Lack of interpretability

9. Customizing LLMs

Two approaches:

- **Fine-tuning** on domain data
- **PEFT** (Adapters, LoRA, P-tuning, etc.)

Optimization: RLHF

10. Model Evaluation

Dimensions:

- Training data quality
- Model size
- Inference speed

Benchmarks:

- GLUE, SuperGLUE, CoQA