The Evolution of Large Language Models: A Comprehensive Survey

Exploring the Advancements, Applications, and Future of LLMs

Puneet Bajaj

CMPE 255 - Data Mining

Introduction

- Large Language Models (LLMs) are reshaping AI with advanced capabilities.

- Applications span chatbots, coding, and multilingual understanding.

- Key emergent abilities: in-context learning, reasoning, and instruction following.

Key LLM Families

- GPT (Generative Pre-trained Transformers)
  - GPT-3: Few-shot learning with 175B parameters.
  - GPT-4: Multi-modal capabilities (text + image).
- LLaMA (Meta AI)
  - Open-source innovation fostering research.
  - Models like Alpaca and Vicuna extend usability.
- PaLM (Pathways Language Model)
  - Med-PaLM for domain-specific applications.
  - Focus on efficiency and scaling.

How LLMs Work

- Development pipeline:

  - Data preprocessing: Cleaning and deduplication.

  - Tokenization: Techniques like BytePairEncoding.

  - Training: Masked language modeling, autoregressive modeling.

- Fine-tuning and reinforcement learning refine results.

Emergent Abilities

- Instruction Following: Adapting to diverse tasks.

- Multi-step Reasoning: Breaking down problems.

- Augmentation: Using tools like retrieval-augmented generation (RAG).

Challenges in LLMs

- High computational costs and resource intensity.
- Ethical concerns: Biases, misinformation, and security.
- Adaptability: Real-time dynamic interactions remain limited.

Future Directions

- Efficient and smaller models for wider accessibility.

- Multi-modal capabilities integrating text, images, and audio.

- Retrieval-Augmented Generation (RAG) for real-world interactions.

Conclusion

- LLMs are transforming AI with significant advancements.
- Challenges remain in efficiency, ethics, and multi-modality.
- Collaboration and responsibility are key for future growth.