Understanding Transformer Models for Language Representation

Transformer-based models such as BERT, GPT, and LLaMA have revolutionized natural language processing by enabling large-scale pretraining on unlabeled text. These models learn contextualized word representations through attention mechanisms that capture relationships across entire sequences. This paper provides an overview of transformer architectures and highlights their applications in text classification, question answering, and summarization tasks. We also discuss challenges in model interpretability and fine-tuning for domain-specific tasks.