

Large Language Models (LLMs) like **ChatGPT** and **Grok** are built using advanced AI techniques, but how exactly do they work? How do LLMs **process and generate text**? What role does the **Transformer architecture** play? How do LLMs like **ChatGPT** and **Grok** learn from data and improve over time?

LLMs generate text by predicting one word at a time based on the words that came before, using a special deep learning structure called a **Transformer**. Transformers help the model understand the entire context of a sentence by paying attention to important words rather than reading sequentially, which makes the responses more relevant and coherent. These models learn from huge amounts of text data through a process called **pre-training**, where they figure out patterns in language, and then improve over time by **fine-tuning** on specific examples and human feedback, making their answers more accurate and helpful.