

After the release of ChatGPT, a huge number of papers published every day focused on large Language Models (LLMs) scaled up 10-fold. The number of parameters in these LLMs increased 10-fold in implementation. From 100M to billions of parameters. A large LLM is the parallel combination of word-embedding layers which are represented by a finite vocabulary. Each word is represented by a set of numbers. Applying a linear transformation on these numbers can produce a response. Simple models, which are trained with out-of-distribution inputs, do not generalize to new languages.