

After the release of ChatGPT, the number of papers published every day about large language models (LLMs) increased more than 10-fold. The number of parameters in these LLMs grew from 340 millions in the first implementation to billions of parameters in models like GPT-4. A large language model (LLM) is a neural network with many layers of word embeddings and hidden layers which are used to represent the meaning of characters, words, and sentences. Although LLMs have a fixed vocabulary, they also respond to out-of-vocabulary inputs, which can lead to brittle models that cannot generalize to new languages.