

# **ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope**

## **Summary on evaluation of LLM**

Large Language Models (LLMs) have been a popular research topic in recent years due to their ability to perform a variety of Natural Language Processing (NLP) tasks without adaptation on downstream data. ChatGPT, a 175-billion-parameter NLP model, has generated excitement in the NLP community because of its capacity to generate high-quality responses to human input. Several studies have been conducted to evaluate ChatGPT's performance on different tasks, such as answering medical exam questions and solving different NLP problems. Although ChatGPT performs well on several reasoning based tasks, it still faces challenges in solving some specific tasks like sequence tagging. Pre-training, adaptation tuning, utilization, and capacity evaluation are the major aspects of LLMs, which have been broadly researched, and the importance of model scaling in LLMs has been discovered. Despite ChatGPT's outstanding performance, researchers have conducted studies on the limitations and categorical archives of ChatGPT's failures in bias, reasoning, math, coding, and factual errors. These studies help in developing enhanced models that better capture language understanding, generation, and performance on different NLP tasks.