Abstract

Zorka Albert

June 2025

Large Language Models (LLMs) have demonstrated powerful capabilities across a wide range of tasks, resulting in widespread usage. This thesis aims to critically explore the generalization capabilities of LLMs and to determine whether their performance resembles genuine understanding or sophisticated mimicry. After outlining fundamental principles and theoretical background, the work examines the examples of generalization in the key areas such as language, code, and logic. The work also presents original experiments on the leading LLMs to test their capabilities. The thesis concludes that while LLMs exhibit impressive pattern-matching skills, they largely operate without deeper understanding. The thesis also highlights the limitations of generalization, measures of generalization, and opportunities for improvement while reflecting on the insufficiencies of the current evaluation methods. It poses important questions about future AI development and highlights possible future advances in the field of LLMs.