Empirical Study to Understand Changes in Markdown Files in GitHub Projects After the Release of ChatGPT

Prachnachai Meakpaiboonwattana

June 5, 2024

Abstract

This paper presents an empirical study to understand changes in Markdown (.md) files in GitHub projects after the release of ChatGPT. We analyze snapshots of repositories mentioning terms like "LLM" and "ChatGPT" to identify any significant shifts in content and structure. The study aims to provide insights into how the introduction of large language models impacts the documentation practices within open-source projects.

1   Introduction

The introduction of @pranaCategorizingContentGitHub2018 ChatGPT and other large language models (LLMs) has had a profound impact on various domains, including software development and documentation practices. This study aims to explore the extent to which Markdown (.md) files in GitHub projects have changed since the release of ChatGPT. We focus on repositories that specifically mention terms related to LLMs and ChatGPT to understand any significant trends or shifts in documentation.

2   Related Work

This section reviews the existing literature on the impact of LLMs on software development and documentation. We discuss previous studies that have analyzed changes in code and documentation within GitHub repositories, and how these findings relate to our research.

3 Methodology

3.1 Data Collection

We collected a dataset containing snapshots of GitHub repositories that mention "LLM" and "ChatGPT". From this dataset, we filtered commits based on the release date of ChatGPT and the presence of ".md" files in commit messages.

3.2 Data Analysis

Our analysis focuses on identifying changes in the content and structure of Markdown files. We employ various quantitative and qualitative methods to examine the nature of these changes, including keyword analysis, sentiment analysis, and structural analysis of the documents.

4 Results

This section presents the findings of our analysis. We provide detailed results on the changes observed in Markdown files, including any notable trends or patterns. We also discuss the implications of these changes for the broader open-source community.

5 Discussion

In this section, we interpret our findings in the context of the existing literature. We discuss the potential reasons behind the observed changes and their implications for documentation practices in GitHub projects. We also consider the limitations of our study and suggest directions for future research.

6 Conclusion

We conclude by summarizing the key findings of our study and their significance. We highlight the contributions of our research to the understanding of how LLMs like Chat-GPT impact documentation practices in open-source projects.

Acknowledgements