

Grobgliederung zum Antrag Bachelorthesis

Student*in::

Maurice Krüger
Wuthenowstraße 10, 12169 Berlin+49 16096513291
krugerm Maurice97@gmail.comErstbetreuer*in: Prof. Dr.-Ing. Ingo Claßen
Zweitbetreuer*in:**Titel der Arbeit: Eignung von Large Language Models (LLMs) zur Generierung von Python Code zur Datenanalyse****Exposé:** Beschreiben Sie hier kurz Thema und Ziele der Thesis (circa ½ - 1 Seite). :

In den vergangenen Jahren haben Anwendungen wie ChatGPT immer mehr an Popularität sowohl im täglichen, als auch beruflichen Alltag gewonnen. Die dahinterliegenden Large Language Models (LLMs) machten über die Zeit enorme Fortschritte. Jedoch nicht nur in der menschlichen Sprache, sondern ebenso in sämtlichen Programmiersprachen, wie zum Beispiel Python. Mittlerweile ist es Normalzustand von Programmierern, LLMs zum Lösen von Problemen zur Hilfe zu nutzen. Durch den immer weiter voranschreitenden Fortschritt dieser öffnen sich neue Türen, genauso auch in der Datenanalyse, wo viele Aufgaben automatisiert werden können. Daher soll in dieser Bachelorarbeit untersucht werden, wie gut sich moderene LLMs zur Generierung von Python-Code zur Datenanalyse eignen. Das Ziel dabei ist es, systematisch zu untersuchen mit welcher Qualität LLMs diesen Code generieren können und wie dieser im Vergleich zu manuell geschriebenen Code abschneidet. Dabei wird ebenso untersucht, in was für Anwendungsfällen LLMs sinnvoll wären und wo dabei dessen Grenzen liegen. Zum Abschluss soll noch ein Zukunftsausblick mit den potentiellen Möglichkeiten von LLMs in der Datenanalyse erstellt werden.

Grobgliederung

1. Einleitung
 - 1.1 Problemstellung und Forschungsfragen
 - 1.2 Relevanz der Thematik
 - 1.3 Zielsetzung
 - 1.4 Aufbau der Arbeit
2. Grundlagen
 - 2.1 Einführung Large Language Models
 - 2.2 Einführung Python
 - 2.3 Einführung automatisierte Code-Generierung
3. LLMs in der Programmierung – aktueller Stand
 - 3.1 Überblick und Vergleich von verschiedenen LLMs
 - 3.2 Einsatzgebiete von LLMs in der Programmierung
 - 3.3 Vergangene Studien und Arbeiten zur Code-Generierung
4. Methodik
 - 4.1 Vorgehensweise der Untersuchung
 - 4.2 Testfälle der Datenanalyse
 - 4.3 Auswertungskriterien
 - 4.4 Verwendete Tools für die Auswertung
5. Auswertung der Python-Code-Generierung zur Datenanalyse durch LLMs
 - 5.1 Generierter Code zur Datenanalyse
 - 5.2 Manuell erstellter Code zur Datenanalyse
 - 5.3 Auswertung und Vergleich der Lösungen (Effizienz und Performanz des Codes)
 - 5.4 Fehler und Grenzen des generierten Codes
6. Fazit und Ausblick
 - 6.1 Zusammenfassung der Arbeit
 - 6.2 Beantwortung der Forschungsfragen
 - 6.3 Zukunftsperspektive
7. Anhang
 - 7.1 Literaturverzeichnis
 - 7.2 Quellcodeverzeichnis
 - 7.3 Tabellenverzeichnis
 - 7.4 Dokumentation der genutzten Tools