Unterstützung von Software Entwicklung mit LLMs – „Chat with your own Code“ (Bachelorarbeit oder Masterthesis)

Hintergrund

Large Language Models (LLMs) finden zunehmend Verbreitung als Werkzeuge oder sogar zur eigenständigen Aufgabenbearbeitung in vielen professionellen Bereichen. So haben Large Language Models u.a. inzwischen auch in der Software Entwicklung Einzug gehalten. Neben professionellen Closed Source Lösungen und Tools wie GitHub Copilot und ChatGPT4 unter Verwendung von Plugins (wie z.B. AskTheCode) gibt es mittlerweile auch sehr leistungsfähige Open Source Lösungen. Zu erwähnen sind hier insbesondere die verschiedenen vortrainierten Coding Modelle auf der Basis von BERT oder LLaMa, wie z.B. CodeBERT, Code LLaMa, StarCoder o.ä.

Aufgabe

Im Rahmen dieser Arbeit soll untersucht werden, wie die Software Entwicklung mit LLMs unterstützt werden kann. Hierzu soll exemplarisch ein oder mehrere LLMs für Coding als Tool aufgesetzt werden. Anschließend muss die Wissensbasis in geeigneter Weise mit dem Code in einem eigenen privaten GitLab Repository (code.fbi.h-da.de) erweitert werden, um Bezug auf das spezielle Software Projekt nehmen zu können. Dies kann durch die Nutzung von Retrieval Augmented Generation mit Embeddings oder durch Fine Tuning des LLMs erfolgen. Das Ziel der Thesis ist es über ein Chat Modell mit dem Code zu kommunizieren und zu interagieren. Mögliche Usecases für die Software-Entwicklung sind dabei:

- Code Comprehension bzw. Explanation – Verständnis und Erklärung der Funktionalität eines vorhandenen Codes bzw. Software Projektes
- Coding Assistance bzw. Code Extension – Weiterentwicklung vorhandenen Codes aufgrund von kontextspezifischen Anforderungen (Prompts),
- Code Testing – Erstellung von Testcases für Code
- Code Documentation – Automatisches Generieren von Dokumentation für Nutzer zu bestehendem Code

Ergänzend soll ein geeignetes Benchmarking der Coding Unterstüzung durchgeführt werden.

Vorkenntnisse

Sie bringen folgende Kenntnisse mit oder haben Freude daran sich diese anzueignen und bestehende Lücken zu schließen:

- Arbeiten mit GitLab
- Programmierung in Python und C++
- Anwenderkenntnisse von Large Language Models

- Solide Grundkenntnisse zu maschinellem Lernen und Konzepten der künstlichen Intelligenz
- Gute Englischkenntnisse für die Literaturrecherche

Ansprechpartner

Prof. Dr. Andreas Müller
Raum 03.06
Gebäude D14
Schöfferstraße 8b
64295 Darmstadt
+49 6151 553-68448
andreas.mueller@h-da.de