

Einsatz von Large Language Models zur Analyse vorhandener Software-Artefakte

- Bachelorthesis
 Masterthesis

Aufgabenstellung

Large Language Models in Form von Tools wie ChatGPT sind aktuell in aller Munde. Es ist zu erwarten, dass ihr Einsatz zukünftig auch bei der Softwareentwicklung eine wichtige Rolle spielen und den Entwickleralltag deutlich verändern wird. Öffentlich verfügbare Tools wie GitHub CoPilot sind derzeit allerdings noch limitiert, da sie nur einen begrenzten Kontext aus dem in Entwicklung befindlichen Gesamtsystem nutzen können, z.B. die aktuell geöffnete Quelltext-Datei. Daher können sie für ihre Vorschläge zwar auf umfangreiches öffentlich verfügbares Wissen zu Frameworks und Design-Pattern zugreifen, aber nicht auf die Architektur-Spezifika des aktuellen Systems. Das Übertragen von eigenem Quellcode an den Tool-Anbieter ist im Hinblick auf Vertraulichkeit ggf. bereits bei kleinen Ausschnitten und erst recht für den Gesamtumfang problematisch.

Um diesen Begrenzungen zu begegnen baut die itestra GmbH derzeit eine private LLM-Infrastruktur auf. Hiermit sollen mögliche konkrete nutzbringende Einsatzszenarien erforscht werden. Das „Langzeitgedächtnis“ des LLM kann mit dem Quellcode eines existierenden Systems und/oder weiteren Artefakten wie Architektur-Dokumentation befüllt werden. In Folge sollen dann z.B. Fragen der Entwickler zu Architektur und Funktionalität des vorhandenen Codes beantwortet sowie Codevorschläge passend zum vorhandenen Design generiert werden können. Hierdurch können vsl. repetitive Entwicklungsarbeiten deutlich beschleunigt und die Einarbeitung neuer Entwickler und Migration oder Erweiterung von Legacy-Systemen vereinfacht werden. Auch Unterstützung im Bereich Code Review oder Qualitätsprüfung sind denkbar.

Zielsetzung der Arbeit

Ziel der Arbeit ist es, Bausteine für die oben beschriebene Toolkette zu konzipieren und umzusetzen und den Nutzen anhand realer Informationssysteme und Fragestellungen aus dem Entwickleralltag zu evaluieren. Details und genaue Abgrenzung der Aufgabenstellung sind je nach Startzeitpunkt der Arbeit noch festzulegen. Zwei mögliche Aufgabengebiete sind:

- Konzeption und prototypische Umsetzung einer Pipeline für das Einlesen von Sourcecode, aber auch weiteren Artefakten wie Buildkonfiguration und Dokumentation zum Aufbau einer Vektor-Datenbank als „Langzeitgedächtnis“
- Konzeption und prototypische Umsetzung eines "custom-GPTs" mit Fokus auf Mandantenfähigkeit: die Datenspeicherung soll für verschiedenen Kundenunternehmen getrennt erfolgen und die Antworten und Codevorschläge sollen sich am Design des jeweiligen Systems orientieren
- Konzeption und prototypische Umsetzung eines IDE Plugins

Unternehmensprofil

Diese Arbeit wird in Kooperation mit der itestra GmbH (www.itestra.de) angeboten. Die itestra GmbH ist ein innovativer, unabhängiger und international tätiger Software-Dienstleister im Bereich unternehmenskritischer Prozesse, Systeme und Anwendungen. Zum Portfolio gehören die Analyse von Geschäftsprozessen, der Entwurf und die Realisierung leistungsfähiger Lösungen in Projekten sowie das strategische Management bereits vorhandener Softwaresysteme.

Betreuung (itestra GmbH)

Pascal Bawidamann, Jonathan Streit (bawidamann/streit@itestra.de)