

Verwendung und Auswertung von generativer KI zur Generierung von Erklärungen für existierende Software-Systeme

Hintergrund

Moderne Softwaresysteme und ihre Algorithmen werden zunehmend komplexer und undurchsichtiger. Nutzende, die mit diesen Systemen interagieren, können oft nicht nachvollziehen, wie diese funktionieren und wie sie zu ihren Ergebnissen kommen. Im schlimmsten Fall verlieren die Nutzenden das Vertrauen in die Software und hören auf, diese zu benutzen. Erklärungen können dazu beitragen, Missverständnisse zu klären und das Vertrauen der Nutzenden in die Software zu stärken. Aus diesem Grund ist die Erklärbarkeit zu einem wichtigen Forschungsthema geworden. Die Implementierung von Erklärungen ist jedoch mit hohen Entwicklungskosten verbunden. Eine Idee, um diese Kosten zu verringern ist das Nutzen einer generativen künstlichen Intelligenz wie *ChatGPT*, um automatisch Erklärungen zu erzeugen. Durch online verfügbare Dokumente wie Handbücher, Forendiskussionen oder Datenschutzbestimmungen ist es denkbar, dass eine generative KI dazu in der Lage ist einige Fragen automatisiert zu beantworten. Dabei soll empirisch ausgewertet werden, ob Antworten der generativen KI zufriedenstellend für den Nutzer sind.

Aufgabe

Im Rahmen dieser Masterarbeit soll untersucht werden, für welche Arten von Erklärungsbedarf generative KI in der Lage ist, zufriedenstellende Antworten zu generieren. Dazu ist es zunächst notwendig, sich mit den Bereichen Erklärbarkeit und generative KI vertraut zu machen. Der nächste Schritt ist die Erarbeitung von Anfragen, die an die generative KI gestellt werden können. Die Basis für diesen Schritt ist eine bereits durchgeführte Nutzerstudie, in der Nutzende unterschiedliche Arten von Erklärungsbedarf zu verschiedenen Software-Systemen geäußert haben. Außerdem soll ein Webtool entwickelt werden, mithilfe dessen die generierten Antworten bewertet werden können (siehe Beispielbild). In dieses Webtool soll *ChatGPT* eingebunden werden, um Ad-hoc-Antworten auf die gestellten Fragen zu generieren. Anschließend soll eine Nutzerstudie konzipiert und durchgeführt werden (~50 Teilnehmende). Die Ergebnisse sollen schließlich mit Bezug auf verschiedene Erklärungs- und Softwarearten ausgewertet werden.



Arbeitspakete

1. Einarbeitung in den Bereich Erklärbarkeit & generative KI
2. Erarbeitung von Anfragen (Queries) für *ChatGPT*
3. Implementierung eines Tools unter Einbindung von *ChatGPT* zur Bewertung von generierten Antworten
4. Konzeptualisierung und Durchführung einer Nutzerstudie (~50 Teilnehmer) zur Bewertung der generierten Erklärungen
5. Auswertung der Ergebnisse und kritische Diskussion, was mit generative AI möglich ist / was nicht.
6. Verschriftlichung der Ergebnisse

Organisatorisches

Betreuer/in: Hannah Deters, M. Sc. hannah.deters@inf.uni-hannover.de

Prüfer: Prof. Dr. Kurt Schneider

Beginn: ab sofort möglich

Randbedingungen: motiviertes, eigenverantwortliches Arbeiten; selbstständige Teilnehmerakquise