FG Software Engineering
Leibniz Universität Hannover
Welfengarten 1
3. Stock, Flur G

Master-Arbeit

- zu vergeben -

Kritische Anwendung eines Referenzmodells für Studien über LLMs und Menschen

Hintergrund

ChatGPT und andere Large Language Models (LLM) überraschen immer wieder mit Leistungen, die man bisher nur Menschen zugetraut hätte. Zudem werden wissenschaftliche Publikationen geschrieben, die evaluieren sollen, ob für die betreffende Aufgabe LLMs so gut wie Menschen oder sogar noch besser arbeiten. Aber solche Vergleiche sind erstaunlich schwierig; immer wieder wird nicht klar gesagt, wie vergleichbare Voraussetzungen geschaffen wurden, wie die Aufgabe per prompt oder Text vermittelt wurde, und mit welchen Kriterien der Vergleich durchgeführt wurde. Daher wurde am FG Software Engineering ein Referenzmodell erstellt, mit dem man dies etwas systematischer und fairer tun kann. Das Referenzmodell enthält einige Aspekte, die in Publikationen beschrieben werden müssen, zum Beispiel: Mit welchen Menschen wurde verglichen? Welche statistische Methode wurde eingesetzt?

Aufgabe

Im Rahmen dieser Masterarbeit soll das Referenzmodell nun auf eine größere Zahl einschlägiger Publikationen angewendet werden. Dabei interessieren einerseits die dabei identifizierten Arten, wie die Vergleiche durchgeführt wurden und beschrieben sind. Andererseits interessiert uns, ob die Anwendung des Referenzmodells noch verbessert werden muss.

Arbeitspakete

Wichtige Arbeitsschritte in der Masterarbeit sind:

1. Systematische Literatursuche (SLR mit Snowballing) nach Publikationen, die LLMs mit Menschen vergleichen.
2. Anwendung des Referenzmodells auf diese Publikationen. Das erfordert eine intensive Auseinandersetzung mit dem Referenzmodell und seinen Aspekten.
3. Auswertung und visuelle Darstellung der Ergebnisse: Was kann man mit dem Referenzmodell über die gefundenen Publikationen sagen? Zum Beispiel: Welche Vergleichstechniken sind üblich, welche selten?
4. Hat die Anwendung des Referenzmodells Unzulänglichkeiten oder Schwierigkeiten zutage gefördert? Vorschläge erarbeiten, was man dagegen tun kann.

Die Arbeit soll, wie üblich, in einer 65-99 Seiten umfassenden Ausarbeitung dokumentiert werden.

Organisatorisches

Betreuer:	Marc Herrmann, M. Sc. marc.herrmann@inf.uni-hannover.de
Prüfer:	Prof. Dr. Kurt Schneider
Beginn:	ab März
Randbedingungen:	Bereitschaft zum lesen von vielen akademischen Publikationen