

Univ.-Prof. Dr.  
Jan Marco Leimeister

e leimeister@uni-kassel.de  
t (0561) 804-6068  
f (0561) 804-6067

Pfannkuchstraße 1  
34121 Kassel

22.02.2024

## Daten im Zeitalter von Large Language Models und generativer KI

### Hintergrund:

Das enorme Potenzial von generativer KI unstrukturierte Daten wie Texte zu verarbeiten, verändert die Art und Weise wie Unternehmen und KI-Entwickler in Zukunft Daten aufbereiten müssen. Während im Bereich des traditionellen maschinellen Lernens, Daten strukturiert und annotiert werden mussten, ist dies bei großen Sprachmodellen nicht mehr nötig. Zum Beispiel verspricht die Anbindung von PDF-Dokumenten oder Excel-Tabellen an generative KI über sogenannte Retrieval Augmented Generation, eine hohe Genauigkeit in der Beantwortung von domänenspezifischen Fragestellungen und Interaktionen. Und dennoch gilt: *Garbage in – Garbage out*. Die zugrundeliegenden unstrukturierten Daten müssen ebenfalls inhaltlich gepflegt werden. Außerdem können besser strukturierte PDF-Dokumente positiven Einfluss auf die Qualität der KI haben. Und letztlich müssen in Retrieval Augmented Generation Systemen einzelne Abschnitte (Chunks) extrahiert werden, die als Basis der Systeme dienen. Daher stellt sich die Frage, wie im Zeitalter von generativer KI, Daten in Unternehmen zielführend gepflegt und strukturiert werden können.

### Mögliche Themen für BA/MA Arbeiten

*Thema 1: Daten Management von LLMs und RAGS.* Eine erste Abschlussarbeit kann sich der Fragestellung widmen, wie Daten für Sprachmodelle und RAG-Systeme aufbereitet werden müssen. Hierfür können sowohl eine Literaturrecherche als auch eine qualitative Interviewstudie mit KI-Experten durchgeführt werden.

*Thema 2: Wissensarbeit und Datenarbeit.* Da generative KI-basierte Systeme weiterhin eine verlässliche Datengrundlage benötigen, sollen in Rahmen einer Abschlussarbeit die Prozesse im Bereich der Wissensarbeit und der Datenarbeit zur Pflege von künstlicher Intelligenz (bspw. Chatbots) untersucht werden. Welche Rollen und Verantwortlichkeiten ändern sich dadurch in Unternehmen? Welche Skills werden dafür benötigt?

*Thema 3: Tools zur Datenaufbereitung.* Dieses Thema beschäftigt sich mit der Analyse von vorhandenen Tools zur Aufbereitung von Daten für das Trainieren von KI-Systemen insbesondere RAG-Systemen. Auf dieser Basis soll ein Tool für das Management von Daten in generativen KI-Systemen konzipiert, entwickelt und evaluiert werden.

### Fragen und Bewerbungen an:

Philipp, Reinhard  
Raum 1170 , ITeG, Pfannkuchstraße 1, 34121 Kassel  
0561/804- 0 6021, philipp.reinhard@uni-kassel.de