UNIKASSEL
VERSITÄT

Fachgebiet
Wirtschaftsinformatik
FB 07 Wirtschaftswissenschaften

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

Explainable AI (XAI) – Erklärbare und transparente generative KI

Hintergrund:
Neben den zahlreichen Vorteilen und Potentialen von generativer KI, weisen insbesondere große Sprachmodelle auch Limitationen und Risiken auf. Bestes Beispiel sind Halluzinationen. Dabei generiert die KI-Antworten, die sehr überzeugend klingen, allerdings keinesfalls auf Fakten basieren. Um diese und weitere Herausforderungen im Umgang mit generativer KI zu überwinden, sollen im Rahmen dieses Themenschwerpunkts Mechanismen von Explainable AI (XAI) verprobt werden, um die Modelle und deren Antworten erklärbarer und transparenter zu gestalten. Damit werden im Sinne einer hybriden Intelligenz Nutzende bei der Entscheidungsfindung unterstützt.

Mögliche Themen für BA/MA Arbeiten
Thema 1: Design, Entwicklung & Evaluation
Im Rahmen einer Abschlussarbeit können bspw. dynamische und interaktive Erklärungen für generative KI-Systeme konzipiert und entwickelt werden.

Thema 2: Experimente zu genXAI. Ziel dieses Bereichs ist es, die Auswirkungen von Erklärungen auf die Mensch-KI Interaktion zu untersuchen. Dabei sollen online Experimente durchgeführt werden, in denen ausgewählte Einflussfaktoren untersucht werden. Hier eine Auswahl an potenziellen Fragestellungen:

- Wie beeinflusst es die Nutzererfahrung und die Entscheidungsfindung, wenn Erklärungen als eigenständige Features oder Merkmale präsentiert werden im Gegensatz zu ihrer Integration in die KI-Antworten?
- Wie können Nutzende über Prompts im Rahmen des sogenannten „Transfactual Reasoning“ die Funktionsweise und die Erklärbarkeit der Antworten spielerisch verstehen?
- Welchen Einfluss hat die Qualität der Datenquelle (Source und Argument Quality) in Bezug auf das Vertrauen der Nutzenden?
- Welchen Einfluss hat die Abfolge der Darstellung von Erklärungen auf die Nutzung? Gibt es „Priming Effekte“?
- Vergleich von statischen und interaktiven Erklärungen …

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