

Bachelorarbeit

Künstliche Intelligenz verstehen: Gestaltung erklärbarer und nachvollziehbarer KI-Systeme

Durch die voranschreitende Entwicklung im Bereich maschinelles Lernen (insbesondere Deep Learning) ist es möglich, maschinelle Ausgaben weitgehend automatisiert ohne ‚human in the loop‘, sondern rein datenbasiert zu erzeugen. Die Ausgaben sind auch für Fachkräfte nicht interpretierbar – KI-Systeme als Black-Box. Erklärbare künstliche Intelligenz (XAI) umfasst verschiedene Prozesse und Methoden, die es Anwendenden ermöglichen soll, die von maschinellen Lernalgorithmen erzeugten Ergebnisse und Ausgaben zu verstehen. XAI-Frameworks stellen somit eine Möglichkeit dar, Verständnis und Nachvollziehbarkeit auf Seite der Anwendenden zu ermöglichen, mit dem Ziel, Vertrauen in die Systeme und ihre Nutzung zu erreichen.

Im Rahmen des BMBF-geförderten Forschungsprojekts KARL (*Künstliche Intelligenz für Arbeit und Lernen in der Region Karlsruhe*, [hier geht's zum Projekt](#)) werden die Auswirkungen verschiedener XAI-Methoden auf den arbeitenden Menschen erforscht. Vorbereitend hierfür ist es notwendig, einen Überblick über die Forschung in diesem Bereich zu erhalten.

Im Rahmen der Abschlussarbeit sollen mittels einer systematischen Literaturrecherche relevante wissenschaftliche Arbeiten im Bereich XAI identifiziert und systematisiert werden. Das Ziel besteht darin, eigenständig eine Bewertungssystematik zur umfassenden Beschreibung und Kategorisierung der verschiedenen Frameworks zu erstellen. Dabei soll ein besonderer Fokus auf dem Aspekt der menschen-zentrierten XAI liegen. Eine nähere thematische Eingrenzung sowie die Formulierung präziser Forschungsfragen als Rahmen für die Recherche erfolgt durch Rücksprache mit den Betreuenden.

Aufgaben:

- Entwicklung und Formulierung einer/mehrerer Forschungsfragen
- Umfassende und systematische Literaturrecherche im Bereich XAI/ human-centered XAI
- Systematisierung und Aufbereitung der Literatur
- Entwicklung einer Bewertungssystematik und anschließende Übertragung auf die recherchierten Arbeiten
- Literaturbasierte Beantwortung der Forschungsfrage und Ableitung relevanter wissenschaftlicher Erkenntnisse

Haben Sie Interesse an dieser Arbeit?

Dann nehmen Sie gerne Kontakt auf:

M. Sc. Lena Kölmel

Tel: 0721 – 608 -44717

Lena.koelmel@kit.edu