

Bachelorarbeit / Masterarbeit

Erklärbare KI: Verstehen Menschen maschinelle Entscheidungen besser, wenn sie mit KI-basierten Systemen frei interagieren können?

Durch die voranschreitenden Entwicklungen im Maschinellen Lernen ist es möglich, maschinelle Ausgaben weitgehend automatisiert und rein datenbasiert zu erzeugen. Die Ausgaben solcher Systeme sind allerdings – meist sogar für Fachkräfte – aufgrund der komplexen Funktionsweise nicht interpretierbar. KI-Systeme werden so zu einer „Black-Box“ und erschweren es dem Menschen, einen Einblick in ihre Funktionsweise zu bekommen und die Ausgaben inhaltlich zu verstehen.

Erklärbare Künstliche Intelligenz (XAI) umfasst verschiedene Methoden, die es Menschen ermöglichen sollen, die von maschinellen Lernalgorithmen erzeugten Entscheidungen, Empfehlungen und Prognosen zu verstehen. XAI-Ansätze stellen somit eine Möglichkeit dar, Verständnis und Nachvollziehbarkeit auf Seite der Anwendenden zu ermöglichen, mit dem Ziel, Vertrauen in die Systeme und dadurch eine langfristige 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](#)) wird in Zusammenarbeit mit dem Fraunhofer IOSB in einem Laborexperiment mit Versuchspersonen untersucht, wie sich die Interaktion mit Erklärungen auf Menschen auswirkt. Im Rahmen der Studie kommt eine am IOSB entwickelte Anwendung zum Einsatz, in der mit - auf einem künstlichen Datensatz basierenden - kontrafaktischen Erklärungen durch Gewichtungen und Ein-/Ausschluss von Faktoren interagiert werden kann.

Ziel der Studie ist es herauszufinden, ob interaktive Erklärungen zu einem besseren Verständnis führen und ob es Unterschiede bzgl. verschiedener subjektiver Maße im Vergleich zu nicht-interaktiven Erklärungen gibt.

Zur Bearbeitung des Themas ist kein tiefergehendes Wissen im Informatikbereich notwendig. Sie sollten jedoch Interesse an Fragestellungen bezüglich der Schnittstelle zwischen Mensch und KI sowie am empirisch-experimentellen Arbeiten mitbringen. Das Thema ist ab sofort zum nächsten möglichen Zeitpunkt zu vergeben.

Aufgaben:

- Literaturrecherche (v.a. experimenteller Studien, Journal Paper) im Bereich Personalisierte/Adaptierbare XAI, Kontrafaktische Erklärung, ggf. spezielle Bereiche in der Mensch-KI-Interaktion
- Integration einer eigenen entwickelten Forschungsfrage in den bestehenden Versuchsplan
- Versuchsleitung: Durchführung der Datenerhebung in einer Laborstudie mit Versuchspersonen
- Ausarbeitung des Auswertungsplans, Aufbereitung der Daten und Datenauswertung mittels z.B. SPSS
- 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