Vertrauenswürdigkeit von KI-Lösungen (Implikationen im Data Science und Software-Engineering)

Veranstaltung der GI-Fachgruppe "Measurement & Data Science" (FG 2.1.10) im Rahmen der ESAPI-Community

15. November 2022 (Gastgeber: Fraunhofer IESE Kaiserslautern – direkt oder virtuell)

Motivation

Das Vertrauen in Anwendungen der künstlichen Intelligenz ist von multidimensionalen Aspekten abhängig. Die „Ethics Guidelines for Trustworthy AI“ der Europäischen Kommission definieren verschiedene Prinzipien und Handlungsempfehlungen, wie das Abwenden von Schaden, Fairness oder transparente Prozesse, als Grundlage für Vertrauenswürdigkeit. Eine ausschließliche Berücksichtigung der technischen Eigenschaften entwickelter Lösungen, die sich z.B. an der ISO 25000 orientiert, ist zwar sinnvoll, reicht aber zur Gewährleistung vertrauenswürdiger KI-Lösungen nicht aus. Die VDE-Anwendungsregel VDE-AR-E 2842-61 des DKE-Arbeitskreises 801.0.8 bricht Vertrauenswürdigkeit in einzelne Qualitätsaspekte herunter, wie Zuverlässigkeit, Verfügbarkeit, Wartbarkeit, Funktionale Sicherheit, Cybersicherheit, Privatsphäre, Benutzerfreundlichkeit, Ethik/Moral und Robustheit. Mit Hilfe von KI-Lösungen gewonnene Klassifizierungen, Prognosen oder auch Bild-, Audio- und Videoanalysen implizieren Bedürfnisse hinsichtlich der Erklär-, Interpretier- und Reproduzierbarkeit. Dabei geht es nicht zuletzt um die Vermeidung diskriminierender Ergebnisse eingesetzter KI-Algorithmen. Die Reproduzierbarkeit erzielter Analyseergebnisse wird durch das BSI als direkte Vorrausetzung für die Verbreitung vertrauenswürdiger KI-Ansätzen genannt[1]:

> „Furthermore, reproducibility is a requirement for establishing causality for the interpretation of model results and building of trust towards the overwhelming expansion of AI systems applications.“ (Quelle des Zitats: BSI 2022)

Unter Berücksichtigung der aufgezeigten Komplexität des Begriffs der Vertrauenswürdigkeit im KI-Diskurs bedarf es dennoch einfach zu handhabender Prinzipien und Methoden, die eine Auseinandersetzung mit sinnfälligen KI-Lösungen von vornherein nicht obsolet machen. Folgende Themenbereiche dienen der Anregung für potentielle Beiträge, selbstverständlich sind weitere Aspekte im denkbar.

Potentielle Themenbereiche:

- Vertrauenswürdigkeit cloudbasiert eingesetzter KI-Algorithmen.
- Umgang mit kompositorisch erstellten KI-Lösungen und deren Vertrauenswürdigkeit.
- Risikogetriebene Ansätze zur Vertrauenswürdigkeit von KI-Lösungen.
- Empirische Analysen zur nutzerzentrierten Bewertung der Vertrauenswürdigkeit.
- Möglichkeiten von Tests und Analysen.
- Bedarf rechtlicher Rahmenbedingungen.
- Implikationen im Softwareentwicklungsprozess.
- Umgang mit vertrauensbildenden Kriterien im Diskurs der eingesetzten Daten.
- Sinnfälligkeit potentieller Zertifizierungsansätze.

[1] Quelle: Deep Learning Reproducibility and Explainable AI (XAI) Results of BSI's project research, Federal Office for Information Security 2022 https://www.bsi.bund.de, letzter Zugriff 13. September 2022

Workshop-Beiträge

Gesucht werden Beiträge von Praktikern und Wissenschaftlern, die sich mit den aufgezeigten Themenschwerpunkten bereits auseinandergesetzt haben. Die Beiträge sollen mit Hilfe von Impulsvorträgen und Postern präsentiert werden, darüber hinaus ist eine Publikation in einem korrespondierenden Tagungsband vorgesehen. Bitte nutzen Sie für die Einreichung der Beiträge im docx- oder pdf-Format (5 bis 8 Seiten) ausschließlich das EasyChair-System! Die Formatierungsrichtlinien werden ebenfalls auf den genannten Webseiten zur Verfügung gestellt.

Termine

10.10.2022	Einreichung von Beiträgen (via Easychair)
24.10.2022	Annahme/Ablehnung (via Email)
01.11.2022	Abgabe der druckreifen Beiträge (unbedingt einzuhalten)
15.11.2022	Workshop (direkt oder virtuell)

Webseite zum Workshop

Weitere Informationen:
https://fg-data-science.gi.de und
https://blog.hwr-berlin.de/schmietendorf

Paper Submission:
https://easychair.org/conferences/?conf=esapi2022

Programmkomitee

S. Aier, Universität St. Gallen	F. Balzer, IBM Deutschland	M. Binzen, DB Systel GmbH
E. Dimitrov, T-Systems	R. Dumke, Uni Magdeburg	J. Marx Gómez, Uni Oldenburg
M. Bauer, CECMG	J. Heidrich, Fraunhofer IESE	A. Johannsen, TH Brandenburg
S. Kosterski, Toll Collect	M. Lother, Robert Bosch GmbH	P. Mandl, HS München
M. Mevius, HTWG Konstanz	S. Schmidt, Deutsche Bahn AG	A. Jedlitschka Fraunhofer IESE Kaiserslautern
A. Fiegler, Microsoft	A. Schmietendorf, HWR Berlin	F. Simon, Zurich Insurance Group
F. Victor, TH Köln	C. Wille, TH Bingen	G. Gurczik BMVI
T. Wiedemann, HTW Dresden	M. Wißotzki, HS Wismar	R. Zarnekow, TU Berlin

Kontakt zur Initiative

Andreas Schmietendorf
HWR Berlin - Berlin School of Economics and Law
E-Mail: andreas.schmietendorf@hwr-berlin.de

Jens Heidrich
Fraunhofer IESE Kaiserslautern:
E-Mail: Jens.Heidrich@iese.fraunhofer.de