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

Workshop der GI-Fachgruppe "Measurement & Data Science" (FG 2.1.10) und der Central Europe Computer Measurement Group (ceCMG) im Rahmen der ESAPI-Community

15. November 2022 (Gastgeber: Fraunhofer IESE - Fraunhofer-Platz 1, 67663 Kaiserslautern)

Die Vorträge am Vormittag können via MS Teams (siehe QR-Code) auch virtuell verfolgt werden!

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.

Anmeldung

Anmeldungen zum Workshop bitte über Herrn Sandro Hartenstein (Sandro.Hartenstein@hwr-berlin.de) bzw. unter www.cecmg.de realisieren.

Webseite zum Workshop

Weitere Informationen und QR-Code für virtuelle Teilnahme:

https://fg-data-science.gi.de,

https://cecmg.de/events

https://blog.hwr-berlin.de/schmietendorf

[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

Agenda

09:30 – 10:00 Eröffnung und Motivation zum Thema

Prof. Dr. Andreas Schmietendorf – Initiator des Workshops

Dr. Jens Heidrich – Gastgeber am Fraunhofer IESE

Dr. Andreas Jedlitschka – Sprecher der GI-Fachgruppe 2.1.10

10:00 – 12:00 Eingeladene Gastvorträge

Dr. Gaby Gurczik, Referentin für Grundsätze KI und Datenökonomie beim BMDV
KI-Innovationen als Standortchance für Deutschland und Europa

Dr. Rasmus Adler, Leiter des Programms Autonome Systeme am Fraunhofer IESE
Das Spaltmaß für KI-Systeme - Wie sieht es aus und was sind akzeptable Grenzwerte?

Prof. Dr. Katharina Zweig, Leiterin Algorithm Accountability Lab TU Kaiserslautern
Kann man mit Surrogatansätzen KI-Entscheidungen erklären?

12:00 Uhr bis 13:00 Uhr Mittagspause

13:00 – 14:30 Moderiertes World Cafe – potentielle Themenvorschläge

- Risiken wahrscheinlichkeitsbehafteter KI-Ergebnisse – Sandro Hartenstein
- Test und Validation von KI-Ergebnissen - Dr. Rasmus Adler
- Möglichkeiten für erklärbare KI-Lösungen - Dr. Andreas Jedlitschka

14:30 bis 15:00 Uhr Kaffeepause

15:00 – 16:30 Poster-Session mit einführenden Impulsvorträgen (jeweils 5 min.)

Sandro Hartenstein (Vertrauen in KI-Web-APIs)

Julius Schinschke (Vertrauen Datenquellen)

Lukas Scholz (Explainable Artificial Intelligence)

Daniel Krohmer (KI Security Testing)

Lisa Jöckel (Testing AI Systems)

Dr. Michael Kläs (Safe AI)

17:00 Uhr Ende des Workshops

Anmeldungen zum Workshop bitte über Herrn Sandro Hartenstein (Sandro.Hartenstein@hwr-berlin.de) bzw. unter www.cecmg.de realisieren.

Bemerkung: Änderungen der Agenda vorbehalten!

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, Bank-Verlag GmbH Köln
F. Victor, TH Köln	C. Wille, TH Bingen	G. Gurczik BMDV
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

Partner

Fraunhofer-Institut für Experimentelles Software Engineering (Gastgeber)
https://www.iese.fraunhofer.de

Central Europe Computer Measurement Group (Sponsoring)
https://cecmg.de

Arbeitskreis Software-Qualität und -Fortbildung e.V. (ASQF)
www.asqf.de

SIGS DATACOM GmbH (Medienpartner)
https://www.sigs-datacom.de

Shaker Verlag GmbH Düren (Medienpartner)
https://www.shaker.de