Workshop zum Thema der KI-Sicherheit im Diskurs der Ergebnisse aus dem Forschungsprojekt TAHAI – technische, organisatorische und ethische Aspekte

Ort: HTW-Berlin Campus Wilhelminenhof (Raum H 101) und MS Teams
Termin: Do. 13. März 2025
Zeit: 13:00 bis ca. 16:30 Uhr

Agenda

Eröffnung und Begrüßung zum Workshop: Erik Rodner, Andreas Schmietendorf (HTW Berlin, HWR Berlin – ProjL TAHAI)

Impuls: Andreas Schmietendorf (HWR Berlin) – 10 min

„Herausforderungen API-basierter KI-Dienste (Blackbox, Vorhersagbarkeit, XAI, Robustness, Fairness, …)“

Vortrag 1: Rainer Rumpel (Auriscon) – 25 min

„Spezielle Angriffsvektoren / -arten von KI (Bedrohungsmodellierung) „

Vortrag 2: Janek Groß (HS Mainz) – 25 min

“Towards Reliable AI / ML Testing by Systematic Assessment of Test Data Quality”

14:00 – 14:10 Uhr - Pause

Vortrag 3: Rudolf Hoffmann (HTW Berlin) – 25 min

„Robustheitsanalyse für Vision-Modelle“

Vortrag 4: Sandro Hartenstein (HWR Berlin, Universität Magdeburg) – 25 min

"Sicherheitsbewertung durch Angriff und Verteidigung: Ein KI-Benchmark mit präventiven Klassifikationsmethoden"

15:00 – 15:20 Uhr - Kaffeepause

15:20 – 16:00 Uhr

Moderierte Expertendiskussion - Identifikation der Top 3 Sicherheitsrisiken bei KI-API-Integration und Überlegungen zu konkreten Gegenmaßnahmen

Mögliche Fokusthemen für die Diskussion:

- Prompt Injection (als wachsendes Risiko)
- Datenschutz bei API-Calls (DSGVO-Konformität)
- Rate Limiting & Kostenkontrolle

16:00 Uhr

Ergebnisse und Verabschiedung

Bemerkung:

Die Teilnahme am Workshop ist kostenfrei, dennoch wird um eine Anmeldung (via E-Mail: tahai@hwr-berlin.de) zum Zweck der besseren Organisation gebeten. Ggf. notwendige Änderungen der Agenda sind vorbehalten.

Veranstalter und Unterstützer:

IFAF-Projekt TAHAI (HTW Berlin/HWR Berlin)
https://www.ifaf-berlin.de

Fraunhofer IESE Kaiserslautern
https://www.iese.fraunhofer.de

GI-FG - "Measurement & Data Science" - Arbeitskreis ESAPI
https://fg-data-science.gi.de

ceCMG - „Central Europe Computer Measurement Group"
https://cecmg.de/

Ergebnissicherung:

Im Nachgang zum Workshop werden die verwendeten Präsentationen Ergebnisdiskussionen im Internet bereitgestellt.

Überblick zum Projekt TAHAI:

Weiterführenden Informationen im Diskurs der Projektleitung TAHAI:

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