

AI540 Gestaltung vertrauenswürdiger KI

Kurzbeschreibung:

Erfahren Sie, wie Sie anhand eines Prüfkataloges KI-Anwendungen vertrauenswürdig entwickeln und überprüfen können. Erlernen Sie ein Vorgehen zur Risikoanalyse anhand von Kriterien, die sich gezielt auf das Thema Vertrauenswürdigkeit beziehen. Um die Überprüfung eigener oder anderer KI-Anwendungen sinnvoll durchzuführen, erhalten Sie Zielvorgaben, die die Behandlung der zuvor identifizierten Risiken umfassen. Ihnen werden Maßnahmen für die Behandlung vorgestellt, die Sie in der Praxis umsetzen können. Mit diesem vertieften Wissen ist Ihnen auch die Entwicklung von KI-Anwendungen möglich, die diese Kriterien beachten und somit den Ansprüchen an die Vertrauenswürdigkeit gerecht werden.

Im Workshop **AI540 Gestaltung vertrauenswürdiger KI** erhalten die Teilnehmer einen strukturierten Leitfaden, anhand dessen KI-Anwendungen in Bezug auf alle relevanten Dimensionen der Vertrauenswürdigkeit evaluiert werden können. Das Hauptaugenmerk liegt auf KI-Anwendungen, die auf Maschinellern Lernen beruhen. Dabei steht der produktorientierte Prüfansatz im Fokus des Seminars.

Zielgruppe:

- Fachexperten
- IT-Fachkräfte
- Entwickler

Voraussetzungen:

Um den Lerninhalten und dem Lerntempo des Kurses gut folgen zu können, sind Vorkenntnisse im Themengebiet des Maschinellen Lernens hilfreich. Wir empfehlen vorab den Besuch folgender Workshops:

- [AI200 Einführung in Python für Data Science und KI](#)
- [AI220 Einführung in Machine Learning](#)

Sonstiges:

Dauer: 2 Tage

Preis: 900 Euro plus Mwst.

Ziele:

- Vorgehen nach Leitfaden verstehen und anwenden können
- Risikoanalysen für KI-Anwendungen durchführen können
- KPI für KI-Anwendungen kennenlernen und anwenden können
- Kriterien für vertrauenswürdige KI-Anwendungen vertieft verstehen und bei Eigenentwicklungen zielgerichtet anwenden können

Inhalte/Agenda:

- ♦ Grundlegende Konzepte und Methodik des Prüfkataloges
- ♦ KI-Steckbrief
- ♦ Dimension: Fairness
- ♦ Dimension: Autonomie und Kontrolle
- ♦ Dimension: Transparenz
- ♦ Dimension: Verlässlichkeit
- ♦ Dimension: Sicherheit
- ♦ Dimension: Datenschutz
- ♦ Dimensionsübergreifende Beurteilung der Vertrauenswürdigkeit
- ♦ Abschließende Diskussion und Q&A