Introduction to Machine Learning Safety

Sommer

(engl. Introduction to Machine Learning Safety)

Modulnummer: FIN-INF-120532

Kürzel	CP	Semester	Fachsem.	Dauer	Sprache	Niveau
MLS	6	Sommer	ab 1.	1 Semester	deutsch	Master

Link zum LSF:	LSF
Zugang beschränkt:	
Verantwortung:	Prof. Dr. Frank Ortmeier
Dozent:in:	Prof. Dr. Frank Ortmeier, Konstantin Kirchheim
Lehrveranstaltungen:	Vorlesung *Introduction to Machine Learning Safety* Übung *Introduction to Machine Learning Safety*
Verwendbarkeit:	- M.Sc. INF: Informatik - M.Sc. INGINF: Informatik - M.Sc. WIF: Informatik - M.Sc. DKE: Learning Methods and Models for Data Science - M.Sc. DE: Methoden der Informatik - M.Sc. VC: Computer Science

Angestrebte Lernergebnisse:

- Verständnis grundlegender Unterschiede zwischen klassischer Softwareentwicklung und datengetriebener Entwicklung von ML-Systemen sowie der daraus resultierenden Sicherheitsimplikationen.
- Fähigkeit, typische Fehlermodi von ML-Systemen systematisch zu analysieren und deren Ursachen (z. B. Datenverteilung, Modellunsicherheit, Verteilungsverschiebung) zu identifizieren.

- Kompetenz, geeignete Test- und Evaluationsverfahren für ML-Modelle in sicherheitskritischen Kontexten auszuwählen und anzuwenden.
- Einordnung und Bewertung von Unsicherheitsabschätzungen, Anomalieerkennung und Robustheitsmechanismen im Hinblick auf ihre sicherheitstechnische Relevanz.
- Verständnis adversarieller Bedrohungsmodelle sowie grundlegender Verteidigungsstrategien.
- Reflexion regulatorischer, ethischer und gesellschaftlicher Rahmenbedingungen beim Einsatz lernbasierter Systeme.

Inhalt:

- Grundlagen und Sicherheitskontext
- Datenqualität und Modelltestverfahren
- Robustheit und Angriffsmodelle
- Unsicherheit und Anomalieerkennung
- Alignment und gesellschaftliche Aspekte

Arbeitsaufwand:

- 20h Übungsteilnahme,
- 20h Vorlesung,
- 40h Vorlesungsvorbereitung,
- 40h Übung,
- 60h Prüfungsvorbereitung

Prüfungsvorleistungen:	Studien-/Prüfungsleistungen:	Lehrform / SWS:
Teilnahme an der Übung	Hausarbeit	Übung (2 SWS) Vorlesung (2 SWS)

Voraussetzungen nach Prüfungsordnung:	Empfohlene Voraussetzungen:
keine	Introduction to Deep Learning

Medienformen:	Literatur:

Hinweise:

Revision #1
Created 2026-10-31 09:02:42 UTC by Fakultät für Informatik
Updated 2026-10-31 09:02:42 UTC by Fakultät für Informatik