

# Erklärbare und sichere KI

Sommer

(engl. Explainable and safe AI )

**Modulnummer: FIN-INF-110496**

**Kürzel**  
XAI

**CP**  
5

**Semester**  
Sommer

**Fachsem.**  
ab 3.

**Dauer**  
1 Semester

**Sprache**  
deutsch

**Niveau**  
Bachelor

**Link zum LSF:**

[LSF](#)

**Zugang beschränkt:**

**Verantwortung:**

Sebastian Stober

**Dozent:in:**

Sebastian Stober

**Lehrveranstaltungen:**

- Vorlesung *Erklärbare und sichere KI*
- Übung *Erklärbare und sichere KI*

**Verwendbarkeit:**

- B.Sc. INF: Informatik - Wahlpflicht
- B.Sc. INF: Studienprofil: Künstliche Intelligenz
- B.Sc. CV: Informatik - Wahlpflicht
- B.Sc. INGINF: Informatik - Wahlpflicht
- B.Sc. WIF: Gestalten und Anwenden - Wahlpflicht
- B.Sc. INF (bilingual): Informatik - Wahlpflicht

## Angestrebte Lernergebnisse:

Die Studierenden ...

- können Methoden zur Erklärung von Entscheidungen und internen Repräsentationen, sowie Prozessen von Modellen des maschinellen Lernens sicher anwenden
- haben die Fähigkeit, die Qualität von Modellen abseits von Genauigkeitsmetriken einschätzen zu können
- können Modelle zielgerichtet verbessern

## Inhalt:

Einordnung unterschiedlicher Kategorien von Erklärbarkeits-Methoden für Modelle des

maschinellen Lernens sowie Anforderungen an Nutzer:innenzentrierte Erklärungen. Darstellung direkt interpretierbarer Modelle (ante-hoc) sowie Analyse- und Visualisierungs-Methoden für Blackbox-Modelle (post-hoc). Zudem werden grundlegende Post-hoc Erklärungstechniken mit modell-agnostischen und modell-spezifischen Ansätzen vorgestellt.

### Arbeitsaufwand:

- 56h Präsenzzeit (Vorlesung + Übung)
- 94h selbstständige Arbeit (Vor- und Nachbearbeitung von Vorlesung (OER) und Übung, Bearbeiten von Übungs- und Programmieraufgaben)

| Prüfungsvorleistungen: | Studien-/Prüfungsleistungen:                                                                     | Lehrform / SWS:                                                                             |
|------------------------|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
|                        | Klausur 120 Minuten Ankündigung der notwendigen Vorleistungen in der ersten Veranstaltungswoche. | <ul style="list-style-type: none"><li>• Vorlesung (2 SWS)</li><li>• Übung (2 SWS)</li></ul> |

| Voraussetzungen nach Prüfungsordnung: | Empfohlene Voraussetzungen:  |
|---------------------------------------|------------------------------|
| keine                                 | Deep Learning für Ingenieure |

| Medienformen: | Literatur: |
|---------------|------------|
|               |            |

### Hinweise:

Revision #1

Created 2026-10-01 07:03:09 UTC by Fakultät für Informatik

Updated 2026-10-01 07:03:10 UTC by Fakultät für Informatik