FAIR-AI

Fostering Austria's Innovative Strength and Research excellence in Artificial Intelligence

Programm / Ausschreibung	AI AUSTRIA Initiative, AI Austria 2022 (Vertrag), AI AUSTRIA Leitprojekt	**Status**	laufend
Projektstart	01.01.2024	**Projektende**	31.12.2026
Zeitraum	2024 - 2026	**Projektlaufzeit**	36 Monate
Keywords	Artificial Intelligence, Trustworthy AI		

Projektbeschreibung

FAIR-AI adressiert die Forschungslücke, die durch den Umgang mit gesellschaftsbezogenen Risiken in der Anwendung von KI entsteht. Insbesondere stehen dabei die Anforderungen des kommenden europäischen KI-Gesetzes und die Hindernisse bei seiner Umsetzung in der täglichen Entwicklung und Verwaltung von KI-basierten Projekten und seiner KI-gesetzkonformen Anwendung im Zentrum des Interesses. Diese Hindernisse sind vielschichtig und ergeben sich aus technischen Gründen (z. B. intrinsische technische Risiken des aktuellen maschinellen Lernens wie Datenverschiebungen in einer nicht-stationären Umgebung), technischen und Management-Herausforderungen (z. B. der Bedarf an hochqualifizierten Arbeitskräften, hohe Anfangskosten und Projektrisiken auf Projektmanagementebene) und soziotechnischen applikationsbezogenen Faktoren (z. B. die Notwendigkeit eines Risikobewusstseins bei der Anwendung von KI, einschließlich menschlicher Faktoren wie kognitive Verzerrungen bei der KI-gestützten Entscheidungsfindung). In diesem Zusammenhang betrachten wir die Erkennung, Überwachung und, wenn möglich, die Antizipation von Risiken auf allen Ebenen der Systementwicklung und -anwendung als einen Schlüsselfaktor. FAIR-AI verfolgt dabei eine Methodologie zur Entflechtung dieser Risikotypen. Anstatt eine allgemeine Lösung für dieses Problem zu fordern, verfolgt unser Ansatz eine Bottom-up-Strategie, indem wir typische Fallstricke in einem spezifischen Entwicklungs- und Anwendungskontext auswählen, um eine Sammlung von lehrreichen, in sich geschlossenen Use-Cases, welche in Research Modulen umgesetzt werden, zu erstellen, um die intrinsischen Risken zu veranschaulichen. Wir gehen über den Stand der Technik hinaus und erforschen Möglichkeiten der Risiko-Entflechtung, -vorhersage und deren Integration in ein Empfehlungssystem, das in der Lage ist, aktive Unterstützung und Anleitung zu geben.

Abstract

FAIR-AI addresses the research gap created by dealing with society-related risks in the application of AI. In particular, it focuses on the requirements of the upcoming European AI law and the obstacles to its implementation in the day-to-day development and management of AI-based projects and its AI law-compliant application. These obstacles are multi-faceted and arise from technical reasons (e.g., intrinsic technical risks of current machine learning such as data shifts in a non-stationary environment), technical and management challenges (e.g., the need for a highly skilled workforce, high initial costs, and project management-level risks), and socio-technical application-related factors (e.g., the need for risk awareness

in the application of AI, including human factors such as cognitive biases in AI-assisted decision making). In this context, we consider the detection, monitoring, and, when possible, anticipation of risks at all levels of system development and application as a key factor. In this regard, FAIR-AI follows a methodology to disentangle these types of risks. Rather than demanding a general solution to this problem, our approach takes a bottom-up strategy by selecting typical pitfalls in a specific development and application context to create a collection of instructive, self-contained use cases, which are implemented in research modules, to illustrate the intrinsic risks. We go beyond the state of the art to explore ways of risk disentanglement, prediction, and their integration into a recommender system capable of providing active support and guidance.

Projektkoordinator

- AIT Austrian Institute of Technology GmbH

Projektpartner

- Fachhochschule St. Pölten ForschungsGmbH
- Brantner Österreich GmbH
- Fabasoft R&D GmbH
- DIBIT Messtechnik GmbH
- onlim GmbH
- Fachhochschule Technikum Wien
- Research Institute AG & Co KG
- Siemens Aktiengesellschaft Österreich
- Österreichische Akademie der Wissenschaften
- Fachhochschule Salzburg GmbH
- JOANNEUM RESEARCH Forschungsgesellschaft mbH
- Wirtschaftsuniversität Wien
- eutema GmbH
- Semantic Web Company GmbH
- Women in Artificial Intelligence Austria (Frauen in Künstlicher Intelligenz Österreich), kurz: Women in AI Austria, WAI Austria
- MD meinDienstplan GmbH
- Austrian Standards International - Standardisierung und Innovation
- Technische Universität Wien
- Hochschule für Angewandte Wissenschaften Burgenland GmbH
- SCHEIBER Solutions GmbH
- Software Competence Center Hagenberg GmbH