ALAIT

Austrian Lab for AI Trust

Programm / Ausschreibung	Digitale Technologien, Digitale Technologien, Digitale Schlüsseltechnologien: Ausschreibung 2023	**Status**	laufend
Projektstart	01.09.2024	**Projektende**	28.02.2027
Zeitraum	2024 - 2027	**Projektlaufzeit**	30 Monate
Keywords	KI Trust; Technikfolgenabschätzung; Technikbewertung; Trustworthy AI; Participatory Design; KI Monitoring		

Projektbeschreibung

Das Projekt „KI Trust Monitor“ (2024 – 2027) entwickelt und erprobt mit wissenschaftlichen Methoden eine neuartige, interdisziplinäre Form von Technologiebewertung für KI-Technologien, welche sich an den Grundsätzen von Trustworthy AI orientiert und pilotiert diese in der Anwendung direkt in der Gesellschaft: Es erprobt im Sinne eines Participatory Designs ein KI Trust Labor mit Dialogformaten und Feedback-Loops zur Technologiebewertung. Damit adressiert das Projekt ein wesentliches sozialen Setting für die Entstehung von gesellschaftlichem Vertrauen und Governance in KI, jenem der Branchen und Berufsgruppen, welche KI anwenden, wie z.B. Bildung, Gesundheit, Medien oder Sachgüterproduktion.
Das Projektkonsortium besteht aus den vier Partnern winnovation, leiwand.ai, TU Wien und APA, welche große wissenschaftliche Expertise in KI-Technologien, ihrer Qualitätsbewertung und gesellschaftlichen Anforderungen an KI sowie innovative methodische Ansätze und eine hohe Disseminierungskraft durch eine starke Verankerung im österreichischen KI-Ökossystem aufweisen.
Konkret entstehen im F&E-Projekt u.a. folgende Ergebnisse:
- neuartige Technology Impact Assessments für aktuelle KI-Technologien
- KI-Technologie-Profile in visualisierter Form z.B. zu KI-Chatbots oder KI-Bilderkennung, welche wesentliche Chancen und Risken aufzeigen
- Branchen White Papers zu KI-Anforderungen der Unternehmen im Zusammenhang mit der Umsetzung von Normen
- Ein neuartiges KI-Dialogformat für Berufsgruppen/Branchen, das KI Trust Labor
- Zum Zwecke der Skalierung ein umfassenden KI-Trust-Labor-Netzwerk an KI-interessierten Organisationen (Berufsverbände, Unternehmen), welche selbst KI Trust Labore durchführen

Der „KI Trust Monitor“ ermächtigt durch eine breite Disseminierungs- und Verwertungsstrategie Schlüssel-Akteursgruppen in der Gesellschaft, informiert mit aktuellen KI-Technologien umzugehen und selbst sozio-technische Normen für einen ethischen und qualitätsvollen Einsatz von KI in ihren jeweiligen Branchen und Bereichen zu entwickeln.
Das Projekt trägt somit in drei wesentlichen Dimensionen zum Aufbau von Vertrauen in KI bei: Schaffung von
1. Transparenz im Sinn von Zugänglichkeit von kompakter, umfassender Information über Wirkungen von KI-Technologien,
2. Governance für den KI-Einsatz und ihrer direkten Einbettung und Übersetzung von Normen in die Gesellschaft und
3. KI-Literacy im Sinn des Erwerbs von Wissen und Kompetenzen für den Umgang mit aktuellen Technologien.

Abstract

The "AI Trust Monitor" project (2024 - 2027) uses scientific methods to develop and test a new, interdisciplinary form of technology assessment for AI technologies that is based on the principles of trustworthy AI and pilots its application directly in society: it tests an AI Trust Lab with dialog formats and feedback loops for technology assessment in the sense of a participatory design. The project thus addresses an essential social setting for the development of social trust and governance in AI, that of the industries and professional groups that use AI, such as education, health, media or material goods production.

The project consortium consists of the four partners winnovation, leiwand.ai, TU Vienna and APA, which have great scientific expertise in AI technologies, their quality assessment and social requirements for AI as well as innovative methodological approaches and a high dissemination power due to a strong anchoring in the Austrian AI ecosystem.

Specifically, the R&D project will produce the following results, among others:

- novel technology impact assessments for current AI technologies

- AI technology profiles in visualized form, e.g. on AI chatbots or AI image recognition, which show the main opportunities and risks

- Industry white papers on companies' AI requirements in connection with the implementation of standards

- A new type of AI dialog format for professional groups/industries, the AI Trust Lab

- For the purpose of scaling, a comprehensive AI Trust Lab network of AI-interested organizations (professional associations, companies) that conduct AI Trust Labs themselves

Through a broad dissemination and exploitation strategy, the "AI Trust Monitor" empowers key stakeholder groups in society to deal with current AI technologies in an informed manner and to develop their own socio-technical standards for the ethical and high-quality use of AI in their respective industries and sectors.

The project thus contributes to building trust in AI in three key dimensions: Creation of

1. transparency in the sense of accessibility of compact, comprehensive information on the effects of AI technologies

2. governance for the use of AI and its direct embedding and translation of standards into society and

3. AI literacy in the sense of acquiring knowledge and skills for dealing with current technologies.

Projektkoordinator

- winnovation consulting gmbh

Projektpartner

- APA - Austria Presse Agentur eG
- leiwand AI gmbh
- Technische Universität Wien