

## AI@Work

AI@Work: Human Centered AI in Digitized Working Environments

|                                 |                                                            |                        |               |
|---------------------------------|------------------------------------------------------------|------------------------|---------------|
| <b>Programm / Ausschreibung</b> | IKT der Zukunft, IKT der Zukunft - 7. Ausschreibung (2018) | <b>Status</b>          | abgeschlossen |
| <b>Projektstart</b>             | 01.10.2019                                                 | <b>Projektende</b>     | 31.12.2020    |
| <b>Zeitraum</b>                 | 2019 - 2020                                                | <b>Projektlaufzeit</b> | 15 Monate     |
| <b>Keywords</b>                 | Human Centered AI; Working Environments; Human AI Teaming  |                        |               |

### Projektbeschreibung

Dieses Projekt greift eine zentrale Frage der aktuellen Diskussion über Human Centered KI für eine europäische KI-Strategie auf, indem es sich der Gefahr einer zunehmenden Polarisierung zwischen KI-fähiger Maschinenintelligenz und Menschen in der Arbeitswelt stellt.

Die angesprochenen Forschungsfragen sind: Wie kann die Akzeptanz von KI in Unternehmen gefördert werden und was sind die Auswirkungen auf KI-Systeme? Welche Faktoren sind in verschiedenen Arbeitskontexten besonders relevant? Welche sind relevante Bereiche der KI-Anwendung und wie ist die Position von Menschen und KI-Systemen innerhalb eines Unternehmens und seines Netzwerks? Was beeinflusst die Interaktion zwischen Mensch und System und wie können die Erwartungen des Einzelnen gestaltet werden, um die Akzeptanz der KI zu fördern?

Der rasante wissenschaftliche und technologische Fortschritt im Bereich der KI droht diese Kluft zu vergrößern, während die Menschen um ihren Arbeitsplatz fürchten. In diesem Zusammenhang konzentrieren wir uns auf Arbeitsumgebungen, die fundiertes Fachwissen und Arbeitserfahrung von Menschen erfordern, so dass datengesteuerte KI-Systeme als unterstützende Entscheidungshilfen dienen. Als exemplarische Szenarien betrachten wir KI-gestützte Systeme in verschiedenen Arbeitsbereichen, z.B. Smart Maintenance in industriellen Prozessen, Chatbot-basierter Kundenservice und KI-gestütztes Software-Engineering.

Ziel dieses Projekts ist es, eine Roadmap für ein nachfolgendes Leitprojekt zur Bewältigung der Akzeptanz- und Sensibilisierungsproblematik der Künstlichen Intelligenz (KI) in digitalisierten Arbeitsumgebungen zu erarbeiten.

Wir gehen dieses Problem mit einem interdisziplinären Ansatz an, der alle wesentlichen Aspekte an der Schnittstelle Mensch-KI-System in solchen Arbeitsumgebungen berücksichtigt. Insbesondere berücksichtigen wir a) psychologische Aspekte der Kooperationsmotivation, b) Transparenz und Vertrauenswürdigkeit des KI-Systems und c) die Nutzung von menschlichem Erfahrungswissen für KI-Systeme mittels wissensbasierter Technologien einerseits und generalisierender Techniken des maschinellen Lernens andererseits. Als Kerninnovation zielen wir darauf ab, ein Mensch-KI-Teaming-Framework zu entwickeln, das es dem KI-System ermöglicht, sein Lernen und seine Verallgemeinerung sowie seine Erklärungsmöglichkeiten zu verbessern.

Als Ergebnis erhalten wir einen systematischeren Überblick über Industriebereiche, in denen das effektive Zusammenspiel von Prozess, KI-Komponente und menschlichen Akteuren die Konzeption eines evolutionären und kontinuierlichen Prozesses

im Gegensatz zu einer vollständigen Automatisierung ohne zusätzliche menschliche Interaktion nach einer ersten temporären Einrichtungsphase erfordert. In diesem Zusammenhang erhalten wir wertvolle Einblicke in die Ist-Situation vor Ort und das Optimierungspotenzial, um KI-Initiativen effektiv umzusetzen.

Dieses interdisziplinäre Projekt wird mit nationalen und europäischen KI-Initiativen wie dem Digital Innovation Hub für KI abgestimmt und bündelt Kompetenzen aus den Bereichen KI, insbesondere maschinelles Lernen und Wissensrepräsentation, prädiktive Analytik und Instandhaltung, Arbeitspsychologie, Human Centered Computing und Produktion.

## **Abstract**

This project takes up a central question of the current discussion of Human Centered AI for a European AI Strategy by tackling the threat of increasing polarization between AI-enabled machine intelligence and people in the world of work. Research questions addressed are: How can user acceptance of AI be fostered in companies and what are the implications for AI systems? Which factors are specifically relevant in different work contexts? What are relevant fields of AI application and what is the position of humans and AI systems within an organization and its network? What influences human-system interaction and how can individuals' expectations be shaped to foster AI acceptance?

The rapid scientific and technological progress in the field of AI threatens to widen this gap while people fear for their jobs. In this context we focus on working environments which require in-depth domain knowledge and working experience provided by humans so that data-driven AI systems serve as assistive decision support systems. As example scenarios we consider AI-enhanced systems in different working domains, e.g., smart maintenance in industrial processes, chatbot based customer service and AI-aided software engineering.

The aim of this project is to work out a roadmap for a follow-up flagship project for tackling the acceptance and awareness gap of artificial intelligence (AI) in digitized working environments.

We tackle this problem by an interdisciplinary approach that considers all essential aspects at the human-AI-system interface in such working environments. In particular, we take into account a) psychological aspects regarding motivation to cooperate, b) transparency and trustworthiness of the AI system, and c) utilization of human experiential knowledge for AI systems by means of knowledge-based technologies on the one hand and more generalizing machine learning techniques on the other hand. As core innovation we aim at designing a human-AI teaming framework that allows the AI system to improve its learning and generalization as well as its explanatory capabilities.

As result we will obtain a more systematic overview of industrial areas where the effective interplay between process, AI component and human actors requires the conception of an evolutionary and continuous process in contrast to a complete automation without additional human interaction after a first temporary setup phase. In this context, we will obtain valuable insight into the as-is situation in the field and the optimization potential to adopt AI initiatives effectively.

This interdisciplinary project will be aligned with national and European AI initiatives such as the Digital Innovation Hub for AI and combines expertise from the fields of AI, in particular machine learning and knowledge representation, predictive analytics and maintenance, work psychology, machine perception, human centered computing and production.

## **Projektkoordinator**

- Software Competence Center Hagenberg GmbH

## **Projektpartner**

- APOLLO.AI GmbH
- Industrie 4.0 Österreich - die Plattform für intelligente Produktion

- Institut für Arbeitsforschung und Arbeitspolitik an der Johannes Kepler Universität Linz
- PROFACTOR GmbH
- Wirtschaftsuniversität Wien Department of Information Systems and Operations Management