complAI

Collaborative Model-Based Process Assessment for trustworthy AI in Robotic Platforms

Programm / Ausschreibung	Ideen Lab 4.0, Ideen Lab4.0 - Ausschreibung 2019	**Status**	abgeschlossen
Projektstart	01.02.2020	**Projektende**	31.01.2021
Zeitraum	2020 - 2021	**Projektlaufzeit**	12 Monate
Keywords	Geschäftsprozesse, Künstliche Intelligenz, Robotik		

Projektbeschreibung

Ausgangsituation, Problematik Motivation:

Um KI in Organisationen einzuführen, stehen Entscheidungsträger vor Herausforderungen die teilweisen unrealistischen Erwartungshaltungen zu bewerten, den zu erwartenden Nutzen im Vergleich zum erwarteten Aufwand abzuschätzen sowie ein Risiko im Verantwortungsbereich des Entscheidungsträgers abzuschätzen alle rechtlichen-, ethischen- und sicherheitsrelevanten Fragestellungen hinreichend zu berücksichtigen.

Aufgrund der Vielzahl verschiedener KI-Methoden mit unterschiedlichen Charakteristika – neben der präzisen aber aufwendigen symbolischen KI, und der abstrahierten aber flexiblen sub-symbolischen KI wird auch die Kombination mit der menschlichen Wissensrepräsentation im Sinne vom Wissensmanagement als intelligentes System verstanden – ist neben dem speziellen Domainwissen der Organisation auch ein tiefgreifendes Wissen über die KI notwendig um Entscheidungen verantwortungsvoll zu treffen.

Ziele und Innovationsgehalt:

Wir entwickeln deshalb ein Assistenzsystem um Organisationen bei der Verwendung von KI zu unterstützen. Relevante ethische, rechtliche und sicherheitsbezogenen Anforderungen werden durch ein modellbasiertes Risikomanagement mittels Softwaretool beurteilt. Als Sondierungsherausforderung wird die sichere, etische und strafrechtlich korrekt geprüfte Ausführung von Prozessen auf Roboterplattformen sichergestellt und die Anwendbarkeit eines solchen Systems getestet.

Die Roboterplattformen werden aufgrund der eindrucksvollen haptischen Demonstrationsmöglichkeiten gewählt um den Unterschied eines Prozesses mit und ohne KI zu verdeutlichen.

Der Innovationsgehalt ist:

• Modellbasiertes Assistenzsystem zur Entscheidungsunterstützung von KI Anwendungen

• Modellbasierte Ansteuerung von Robotern durch technische und fachliche Prozesse
• Signatur von Prozessen, die durch ein ethisches-, rechtliches- und sicherheitsspezifisches Assessment nachvollziehbar überprüft worden sind
• Die Konzeptualisierung und damit zur Verfügungstellung in einem Assistenzsystem von ethischen-, rechtlichen- und sicherheitsrelevanten Kriterienkatalogen
• Die Interpretation der Abhängigkeit von KI-Methoden und Kriterienkatalog.

Angestrebte Ergebnisse und Erkenntnisse:

• Im Sondierungsprojekt wird für den Anwendungsfall, sichere Robotik ein Modellbasiertes Assessmentsystem prototypisch entwickelt und auf ADOxx.org zur Verfügung gestellt.
• Der Mechanismus wie Prozess mittels Assessmentkriterien beurteilt und zertifiziert werden können wird erprobt.
• Roboterplattformen testen Mechanismen um ausschließlich vertrauensvolle Prozesse abzuarbeiten.
• Beispielmodelle von Assessmentkriterien und den Abhängigkeiten von Anwendungsfall spezifisch ausgewählten KI Methoden werden publiziert.

Erkenntnisse über die Machbarkeit, der Plausibilität sowie der Anwendbarkeit solcher Systeme wird im Rahmen der Demonstration zur Formulierung weiterführender Forschungsfragen verwendet.
Im Zuge dieser Sondierung wird auch das Risiko dieser Forschungsfragen speziell im Bezug auf die Möglichkeit die Assessmentkriterien zu Konzeptualisieren und somit mittels Modelle nutzbar zu machen – insbesondere die Überwindung der semantischen Lücke – bewertet.

Abstract

Starting Point, Challenges and Motivation

In order to introduce AI-based processes in organisations, the respective decision makers faces great challenges with regard to i) appropriately addressing unrealistic expectations, ii) assessing the expected added value compared to the expected effort in building and maintaining AI, and iii) the sufficient assessing the risk of legal, ethical, safety and security issues. Facing a variety of AI methods with different characteristics – the precise but expensive symbolic AI, the abstracted but flexible sub-symbolic AI as well as a combination including the human interpretation in the sense of knowledge management – requires a fundamental expertise about those AI mechanisms, chances and challenges on top of domain expert knowledge to make a traceable, transparent and informed decision.

Goal and Innovation

We develop an assistance system, which is to guide organisations in using AI. Relevant ethical, legal, safety and security issues will be assed in a model-based risk management software tool. For this investigation, we choose the safe, secure, ethical- and legal compliant execution of processes on industrial robotic platform, which perform both through certified and trustworthy processes.
The selection of the respective robotic platforms builds upon their impressive haptic demonstration potentials, which allow

the visual demonstration of a certain process with and without AI.

Innovation is identified in:

• Model-based assistance system for decision support in using AI
• Model-based control of robotic platform with technical and domain specific processes
• Using electronic signature for processes to certify that they have been ethically, legally, safety- and security-wise approved.
• Conceptualisation of assessment criteria in order to enable computer interpretation of ethical-, legal-, security- and safety issues.
• Interpretation of assessment criteria and their relation to AI methods via the assessment system.

Targeted Results and Insights

• Model-based assessment system for the investigation of using secure robotic developed on the open platform ADOxx.org
• Mechanisms how processes can be assessed using assessment criteria and, in case of traceable approval, are electronically signed.
• Robotic platform test mechanisms to ensure that exclusively trustworthy processes are executed.
• Sample models that show assessment criteria and their dependencies for application specific selection of AI methods are published.
Insights about feasibility, plausibility and applicability of such systems are used to express more detailed research questions during the demonstration of the prototypes.
The gained insights are further used to assess the risk of those research questions, especially with the possibility to conceptualise assessment criteria by considering the semantic gap that may arise in such transformation, and through this make these both computable and interpretable by software agents by considering the semantic gap that may arise in such transformation.

Projektkoordinator

- BOC Asset Management GmbH

Projektpartner

- JOANNEUM RESEARCH Forschungsgesellschaft mbH
- Universität Linz - Institut für Strafrechtswissenschaften
- Universität Wien Institut für Philosophie