

# OBARIS

Ontology-Based ARtical Intelligence in the Environmental Sector

<b>Programm / Ausschreibung</b>	IKT der Zukunft, IKT der Zukunft, IKT der Zukunft - 7. Ausschreibung (2018)	<b>Status</b>	abgeschlossen
<b>Projektstart</b>	01.02.2020	<b>Projektende</b>	31.10.2022
<b>Zeitraum</b>	2020 - 2022	<b>Projektlaufzeit</b>	33 Monate
<b>Keywords</b>	semantics, knowledge graphs, AI, auditability, semantic AI, interoperable micro services		

## Projektbeschreibung

In den vergangenen Jahren hat eine Renaissance der Künstliche Intelligenz (KI) Forschung eine Reihe von Durchbrüchen ermöglicht, welche auch zunehmend in industriellen Anwendungen wie etwa intelligente Assistenten oder selbstfahrende Fahrzeuge resultieren. Gleichzeitig sind aus Forschungen im Bereich des Semantic Web neue Technologien im Bereich von symbolischen, semantischen Systemen entstanden. Diese Entwicklung hat auch zu einer zunehmenden Anwendung von Wissensgraphen geführt, welche es ermöglichen komplexes Wissen in wiederverwendbar Form abzubilden und daraus neues Wissen abzuleiten. Außerdem arbeitet die Forschung an neuen Möglichkeiten zur Lösung von Problemen um latentes Wissen auf Basis schwacher Signale zu erfassen.

Bedeutende Fortschritte wurden dabei sowohl im Bereich symbolischer, als auch im Bereich subsymbolischer KI-Forschung gemacht; die Kombination beider Paradigmen zur Nutzung derer komplementären Stärken bietet aber noch großes Potential. In diesem Zusammenhang mangelt es gegenwärtig an (i) einem systematischen Verständnis der Kategorien semantischer KI Systeme und (ii) technischen Werkzeugen zur Implementierung solcher Systeme. Schließlich kommt dem Aspekt der Auditierbarkeit, welcher eine grundlegende Voraussetzung für Transparenz darstellt, sowohl im KI-Bereich im Allgemeinen, als auch in vielen praktischen Anwendungsfällen große Bedeutung zu.

OBARIS zielt darauf ab, den Stand der Forschung und Praxis im Bereich semantischer KI-Systeme voranzutreiben. Der Fokus liegt dabei auf zwei Anwendungsfällen im Umweltbereich, die sich dadurch auszeichnen, dass sie komplexe Analysen von heterogenen Daten aus unterschiedlichsten Quellen erfordern. Konkret bestehen die Projektziele darin, (i) den Stand des Verständnisses von Typologien von Semantischen KI-Systemen zu verbessern; (ii) Technologien zur Umsetzung dieser Systeme zu entwickeln; (iii) Leitlinien und einen Rahmen für die Umsetzung von auditierbaren, transparenten, und vertrauenswürdigen Systemen zu schaffen und (iv) die Projektergebnisse durch reale Anwendungsfälle zu validieren. Zu diesem Zweck verfolgt das Projekt einen multidisziplinären Ansatz welcher Forschung zu Wissensgraphen, Linked Data, Machine Learning, Auditierbarkeit, und Umweltinformatik vereinigt.

Das Projekt wird eine Reihe von konkreten, innovativen Ergebnissen hervorbringen, insbesondere (i) eine Taxonomie semantischer KI-Systeme, welche das Verständnis solcher Systeme sowohl in der Forschung, als auch bei potentiellen industriellen Anwendern vertiefen wird; (ii) einem Semantischen KI-Technologie-Stack, welcher die Erstellung hybrider semantischer Machine Learning Pipelines aus intelligenten Bausteinen für die Datenakquise, -integration, und -analyse

erlaubt; (iii) Leitlinien, einen Rahmen, und einen Forschungsprototypen zur Entwicklung auditierbarer semantischer KI-Systeme; und (iv) Demonstratoren, welche die entwickelten Methoden und Komponenten in zwei Anwendungsfälle im Umweltbereich übertragen. Die Beteiligung des Umweltbundesamtes als Anwendungsfallpartner und zentrale Kompetenzstelle der österreichischen Bundesregierung in Umweltbelangen wird dabei die Schlagkraft des Projektes sicherstellen, was zu einer Stärkung der KI-Agenda in Österreich beitragen kann.

## Abstract

In recent years, we have seen a renaissance in Artificial Intelligence (AI) research and the diffusion of AI into industry-strength applications, such as intelligent question answering systems or self-driving vehicles. On the one hand, technologies in the area of symbolic, semantic systems have emerged from Semantic Web research. This development culminated in the industrial adoption of Knowledge Graphs, which help to capture complex world knowledge in a reusable manner and makes it possible to infer new knowledge from it. On the other hand, machine learning research has been extended to solve problems where it is important to capture latent knowledge by learning from weak signals.

While advances have been made in both symbolic and subsymbolic AI research, further opportunities lie in the combined use of these two paradigms in ways that benefit from their complementary strengths, especially when applied in real-life settings that exhibit characteristics addressable by both paradigms. At this stage, there is still a lack of (i) a systematic understanding of categories of such semantic AI systems and (ii) technology stacks that enable implementation of such system. Finally, auditability, which is a necessary prerequisite towards providing transparency, is a major concern in AI in general and in many practical application contexts in particular.

In this setting, OBARIS aims to advance the state of the art in semantic AI systems by investigating both conceptual aspects of these systems and by developing a technology stack that facilitates transposing these system types into concrete settings. In particular, we focus on two use cases from the environmental domain characterized by the need to perform complex analytics on collections of heterogeneous data sources. Concretely, the project's goals are to (i) advance the understanding of typologies of semantic AI systems; (ii) establish a technology stack for creating such systems; (iii) produce guidelines and frameworks that enable building auditable, transparent and trustworthy systems and (iv) to validate the project findings by means of real-world use cases. To that end, the project follows a multidisciplinary approach that brings together research on Knowledge Graphs, Linked Data, Machine Learning, Auditability, and Environmental Informatics.

The project goals will materialize in a number of concrete, innovative results including (i) a taxonomy of semantic AI systems that will foster the understanding of such system typologies by the research community and industry adopters alike; (ii) a semantic AI framework offering a technology stack for creating hybrid semantic - machine learning data processing pipelines constructed from intelligent building blocks for data acquisition, integration, and analytics; (iii) guidelines, frameworks and a research prototype that enable building and monitoring auditable, transparent and trustworthy Semantic AI Systems; and (iv) demonstrators that apply the developed methods and components within two use cases in the environmental domain.

The involvement of Umweltbundesamt (U), the environmental specialist institution of the Austrian Federal Government, as a use case partner will enable a strong impact of the project, which should ultimately contribute towards the general AI agenda of Austria.

## Projektkoordinator

- Technische Universität Wien

## **Projektpartner**

- Semantic Web Company GmbH
- Umweltbundesamt Gesellschaft mit beschränkter Haftung (UBA-GmbH)