

Vorhaben	<b>SensAI</b> <b>Selbstorganisierende persönliche Wissensassistenten in evolutionären Unternehmensgedächtnissen</b>
Titel	<b>Abschlussbericht Teil I: Kurzbericht</b>
Förderkennzeichen	01IW20007
Zuwendungsempfänger	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH Trippstadter Straße 122, D-67663 Kaiserslautern
Ausführende Stelle	DFKI GmbH - Smarte Dienste & Wissensdienste
Projektleiter	Prof. Dr. Prof. h.c. Andreas Dengel, Christian Jilek
Bewilligungszeitraum	01.08.2020 - 31.10.2023
Autoren	Christian Jilek, Heiko Maus, Markus Schröder, Michael Schulze, Mahta Bakshizadeh, Desiree Heim, Federico Raue, Andreas Dengel
Erstellungsdatum:	01.03.2024

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IW20007 gefördert. Dem Muster in Anlage 2 zu Nr. 4.1 NKBF 2017 folgend, handelt es sich bei diesem Dokument um Teil I, einen Kurzbericht des Vorhabens. Die Teile II und III liegen als gesonderte Dokumente vor. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

**Vision & Herausforderungen:** Die Vision für SensAI war ein sich selbst anpassender persönlicher Assistent, der in die Arbeitsumgebung der Nutzer eingebettet ist. Zur Erreichung dieser Vision wurden konkret drei Ziele verfolgt: 1. Erforschung und Entwicklung von Human-in-the-Loop-Ansätzen zur Konstruktion von persönlichen und Unternehmens-Wissensgraphen aus unübersichtlichen („messy“), spärlichen und vielfältigen Unternehmensdaten, 2. Erfassung und Realisierung sich entwickelnder und selbstorganisierender Kontexte basierend auf Benutzeraktivitäten, dabei insb. Erforschung wissensbasierter Deep-Learning-Ansätze für das SensAI-Szenario und 3. Realisierung von SensAI als eine anpassbare Ökonomie von Wissensdiensten eingebettet in die Arbeitsumgebung der Nutzer. Herausforderungen hierbei ergaben sich aus den oben beschriebenen Eigenschaften der Datenbestände, dem Aufwand für den Aufbau und die Pflege der sich entwickelnden Wissensgraphen unter Verwendung dieser heterogenen Daten. Neben diesen unternehmensbedingten Herausforderungen erforderte der verfolgte, menschenzentrierte Ansatz zur Einbindung der Nutzer und zur Berücksichtigung ihrer persönlichen Sichten auf den Informationsraum des Unternehmens den Umgang mit ihren persönlichen Daten und Interaktionen, sowie Unterstützung bei der sich entwickelnden Komplexität des Unternehmensgedächtnisses und die Notwendigkeit, Vorschläge von Wissensdiensten zu erklären.

**Evolutionäre Wissensgraphen (AP1):** Im ersten Arbeitspaket wurden verschiedene Human-in-the-Loop-Verfahren zur Erzeugung von Wissensgraphen aus unübersichtlichen, spärlichen und vielfältigen Unternehmensdaten entwickelt, insb. aus Kurztexten wie Dateinamen oder Tabellen. Hierzu gab es eine ganze Reihe von Veröffentlichungen und eine Dissertation wurde abgeschlossen. Ein untersuchtes Spezialgebiet in diesem Bereich war Purchase-to-Pay, zu dem neben Lösungen und auch eine eigene Ontologie vorgestellt wurde. Mit *Contextifier* wurde ein Werkzeug zur retrospektiven Erkennung („Mining“) von Nutzerkontexten vorgestellt. Ebenso wurden kontextbezogene Empfehlungen und die Unterstützung der Evolution von Nutzervokabular untersucht.

**Selbstorganisierende Kontexte (AP2):** In AP2 wurden Experimente zu Link Prediction und Disambiguierung durchgeführt. Hinsichtlich wissensbasierter Deep-Learning-Verfahren wurden insb. Kontext-Embeddings und der Ansatz des Kontinuierlichen Lernens (Continual Learning) zur Kontextzuordnung untersucht. Zudem wurden mehrere Verfahren zur Erkennung neuer Kontexte und Zuordnung von Nutzeraktivitäten zu bestehenden Kontexten entwickelt. Ebenso wurden prototypische Lösungen zur Unterstützung des gesamten Lebenszyklus eines Kontexts, d.h. die Vision selbstorganisierender Kontexte, realisiert. Dies umfasst das Erzeugen bzw. Zurücknehmen eines Kontexts (bei Re-Evaluation einer Situation), ggf. Aufteilen oder Verschmelzen nach einiger Zeit, sowie das Kondensieren und das ganz oder teilweise Ausblenden (Vergessen) von Kontexten. Hierzu wurde eine Dissertation abgeschlossen und generell zu allen Themen des SensAI-Projekts eine umfassende, retrospektive Übersicht veröffentlicht.

**SensAI Wissensdienste (AP3):** Diverse Entwicklungen im Projekt flossen in ein gemeinsames Assistenz-Framework. In mehreren Studien wurde dabei der Umgang von Nutzern mit Kontexten und insb. selbstorganisierenden Kontexten untersucht und dabei insb. auf die Quantifizierung von Vorteilen dieser Nutzung abgestellt. Beispielsweise konnten erleichterte Aufgabenwechsel beobachtet werden. Ebenso stellten sich aus der Psychologie bekannte Cognitive-Offloading-Effekte ein, d.h. Nutzer, die durch das System unterstützt wurden, hatten nachweislich eine geringere mentale Belastung und zeigten so bspw. bei Folgeaufgaben bessere Leistungen. Weitere untersuchte Aspekte waren Nutzervertrauen in solch autonome Assistenzsysteme, deren Einsatz insb. in der Buchhaltung und im Bereich Digitaler Zwillinge im Unternehmen zur Unternehmensmodellierung.

**Entstandene Datensätze & Werkzeuge:** Im Projekt wurden mehrere Datensammlungen durchgeführt. Hierbei ist insb. *RLKWiC* zu nennen, was für „reale Wissensarbeit im Kontext“ steht. RLKWiC ist der erste öffentlich verfügbare Wissensarbeits- bzw. Information-Management-Datensatz seiner Art, da er vier Aspekte vereint, die bisher in keinem Datensatz gemeinsam verfügbar waren. Zudem wurden mehrere Datensatzgeneratoren entwickelt und veröffentlicht. Hier ist insb. *KnoWoGen* zu nennen, die Vision eines Generators, der Multi-Agenten-Systeme und große Sprachmodelle (Large Language Models) kombiniert, um so auf umfassende Weise Wissensarbeit zu simulieren und authentische Dokumente und Datenspuren zu erzeugen. KnoWoGen hat das Ziel, mit vollständig synthetischen aber täuschend echt aussehenden Daten die üblichen Probleme beim Sammeln von Nutzerdaten wie Datenschutz, Privatsphäre, Zensur und Datenselektion endgültig zu lösen. Obwohl sich KnoWoGen noch in einem frühen Stadium befindet, wurden bereits erste vielversprechende Ergebnisse erzielt. Außerdem wurden im Projekt diverse Werkzeuge zur Erzeugung von und dem Umgang mit Wissensgraphen und semantischen Technologien entwickelt.

**Publikationen & Kooperationen:** Insgesamt wurden durch Kernmitglieder des SensAI-Teams 29 Publikationen (darunter zwei Dissertationen) und sechs Datensätze bzw. Datensatzgeneratoren veröffentlicht, neun Open-Source-Projekte gegründet (zu weiteren beigetragen) und zwölf studentische Abschlussarbeiten betreut. Neben anderen Forschungsbereichen und Transfer Labs<sup>1</sup> des DFKI kooperierte das Team insb. mit Forschern der Universität Trier, PH Ludwigsburg, Leibniz Universität Hannover & L3S, HS RheinMain, HS Bonn-Rhein-Sieg, Ostschweizer FH (OST), TU Eindhoven, Universität Mannheim und der W3C Knowledge Graph Construction Community Group. Kooperierende Industriepartner waren insb. die envia Mitteldeutsche Energie AG (enviaM) und die b4value.net GmbH.

---

<sup>1</sup>Wissens- und Technologietransferprojekte u.a. mit der Deutschen Bundesbank, sowie dem Bundes- und Landes kriminalamt Rheinland-Pfalz