

Relevance Agent

Relevant Data Aggregation Agent for Deep Analytics – Enabling Self-Service and Data Distribution in HeadlineHunter

Programm / Ausschreibung	IWI 24/26, IWI 24/26, Basisprogramm Ausschreibung 2025	Status	laufend
Projektstart	10.06.2025	Projektende	09.06.2026
Zeitraum	2025 - 2026	Projektaufzeit	13 Monate
Projektförderung	€ 259.910		
Keywords			

Projektbeschreibung

Gesamtziel des Projekts ist die Entwicklung eines LLM-basierten Multi-Agenten-Systems zur vollautomatisierten, nachvollziehbaren Aggregation relevanter Datensätze aus dem firmeneigenen HeadlineHunter Data Lake - einer dynamisch wachsenden Datenbank mit über 500 Mio. Nachrichtendaten, laufend aktualisiert durch den eigenen Web-Crawler. Nutzer:innen sollen die gewünschten Datensätze künftig möglichst intuitiv im Self-Service auf Basis natürlichsprachlicher Beschreibungen anfordern und für weiterführende tiefgehende Analysen, sogenannte Deep Analytics, verwenden können - etwa für die vollautomatisierte Auswertung durch den in einem vorangegangenen FFG-Basisprogramm erfolgreich entwickelten AI One-Pager Report, der bereits von namhaften B2B-Kund:innen genutzt wird. Aktuell ist der gesamte Prozess nur mit großem manuellen Aufwand durch Training und Validierung individueller Suchmodelle seitens neptun.ai möglich.

Um die notwendige Qualität und Transparenz der vom Agentensystem generierten Datensätze für nachgelagerte Analysen – seien es Reports, Trendanalysen oder Training von ML-Modellen – sicherzustellen, wird ein Datenvalidierungssystem mit nachvollziehbarer Entscheidungslogik entwickelt. Denn Ziel ist die Generierung von Datensätzen, wie sie in dieser Qualität bisher nur durch aufwendige manuelle, bei großen Datenmengen praktisch nicht durchführbare Expert:innenkuratorierung möglich wäre – nun vollständig automatisiert, skalierbar, im Self-Service nutzbar und mit transparenter, nachvollziehbarer Herleitung der Zusammensetzung des Datensatzes versehen. Dies stellt einen erheblichen Innovationssprung gegenüber klassischen Ansätzen zur Datenselektion dar und stärkt die europäische Datenhoheit im KI-Umfeld.

Die Besonderheit des Systems liegt in der kooperativen Architektur spezialisierter Softwareagenten, die – gesteuert durch moderne LLMs und Chain-of-Thought-Techniken – erstmals zentrale Aufgaben wie Relevanzbewertung, Datenprüfung, Optimierung und Konsolidierung vollständig automatisiert übernehmen. Das Gesamtziel gliedert sich in folgende Subziele:

1. Entwicklung eines Self-Service-Eingabeassistenzmoduls mit LLM-gestützter Interpretation und Vervollständigung natürlichsprachlicher Datasat-Anfragen durch Nutzer:innen
2. Aufbau einer latenzarmen Schnittstelle zur performanten und flexiblen Anbindung des Agentensystems an den

HeadlineHunter Data Lake

3. Entwicklung eines Multi-Agenten-Systems zur intelligenten, vollautomatischen Aggregation relevanter Subdatasets aus dem HeadlineHunter Data Lake
4. Entwicklung eines agentenbasierten Validierungs- und Konsolidierungsmoduls zur Qualitätsprüfung und Zusammenführung der Subdatasets zu einem finalen Ausgabedatensatz auf Expert:innenqualitätsniveau für weiterführende Deep Analytics

Projektergebnisse:

1. Forschungsjahr: Die Architektur des agentenbasierten Systems ist vollständig spezifiziert, das Eingabeassistenzmodul ist entwickelt und erfüllt die KPIs, die Data-Lake-API ist implementiert und erfüllt die KPIs.
2. Forschungsjahr: Das agentenbasierte System zur relevanzbasierten Datenselektion ist entwickelt & validiert, das agentenbasierte System zur Datenprüfung, Optimierung & Konsolidierung ist entwickelt & validiert und erfüllt die KPIs, der Deep Analytics Data Aggregation Agent ist in das Produktivsystem HeadlineHunter integriert.

Projektpartner

- Neptun Data Processing GmbH