Trusted AI One-Pager

Prädiktive & halluzinationssichere One-Pager-Reports – KI-generiert durch Meta-Zusammenfassungen von aktuellen Headlines

Programm / Ausschreibung	IWI 24/26, IWI 24/26, Basisprogramm Ausschreibung 2024	**Status**	abgeschlossen
Projektstart	01.06.2024	**Projektende**	31.01.2025
Zeitraum	2024 - 2025	**Projektlaufzeit**	8 Monate
Keywords			

Projektbeschreibung

Das übergeordnete Ziel ist die Entwicklung und Integration eines Verfahrens für die Generierung von automatisiert validierten & damit vertrauenswürdigen One-Pager-Reports auf Basis aktueller Nachrichtendaten aus dem firmeneigenen Media-Intelligence-System "HeadlineHunter". Die One-Pager-Reports werden durch Integration & Feintuning eines möglichst aktuellen Sprachmodells der GPT-Familie die Kernaussagen der zugrundeliegenden Headlines in Form von natürlichsprachlichen Fließtexten komprimiert zusammenfassen. Da Headlines bereits selbst sehr oft zusammenfassenden Charakter haben, sprechen wir hierbei vom neuartigen Anwendungsfall der „abstractive meta-summarization“.

Um den hohen Qualitätsanforderungen für diesen Reporting-Use-Case gerecht werden zu können, soll die Wahrscheinlichkeit der semantischen Übereinstimmung der Kerninhalte aller generierten Meta-Zusammenfassungen mit den Input-Daten bei mind. 90% liegen. Dies wird durch einen neuentwickelten Algorithmus - basierend u.a. auf Kosinus-Ähnlichkeit und Named-Entity-Informationen - gemessen. Dadurch kann der problematischen "Halluzination" von generativen KI-Modellen entgegengewirkt werden.

Die den aktuellen technologischen Fortschritten angemessene Neuentwicklung des Reporting-Systems ist wesentlich für die langfristige Sicherung & den Ausbau der Marktposition des Antragstellers. Das übergeordnete Ziel lässt sich in folgende Subziele unterteilen:

1. Implementierung einer Pre-Processing-Pipeline zum Splitting des für den jeweiligen Report bestimmten Eingabekorpus in zumindest 5 Subkorpora, gruppiert nach den enthaltenen Schlüsselkonzepten

2. Feintuning eines möglichst aktuellen Sprachmodells aus der GPT-Familie & Schnittstelle zum Output des Pre-Processings

3. Ermittlung adäquater Eingabe-Prompts

4. Automatisierte Validierung der semantischen Korrektheit der Ausgabedaten zur Sicherung der notwendigen hohen Qualität für den Reporting-Use-Case („Halluzinations-Scoring“)

5. Darstellung der Output-Daten in einem One-Pager-Report im PDF-Format und Integration in das bestehende System HeadlineHunter

Endberichtkurzfassung

Im Projekt wurde ein mehrstufiges und hoch innovatives Verfahren entwickelt und integriert, mit dem automatisiert validierte und somit vertrauenswürdige One-Pager-Reports auf Basis aktueller, sprachübergreifender Nachrichtendaten aus dem Trendanalyse-System „HeadlineHunter“ generiert werden können. Diese Reports enthalten komprimierte, natürlichsprachliche Zusammenfassungen („abstractive meta-summarizations“) der Kernaussagen mehrerer themenbezogener Headlines und basieren auf einem integrierten und speziell für diesen Anwendungsfall feinabgestimmten großen Sprachmodell. Ein eigens entwickelter Validierungsalgorithmus, der auf Kosinus-Ähnlichkeit sowie Named-Entity-Informationen basiert, gewährleistet dabei eine semantische Übereinstimmung von mindestens 90 % zwischen generierten Meta-Zusammenfassungen und den ursprünglichen Nachrichtendaten, wodurch problematische Halluzinationen des KI-Modells auf neuartige Weise effektiv reduziert werden und das Reporting-System auch in sensiblen und entscheidungsrelevanten Bereichen wie der Industrie verwendet werden kann.

Projektpartner

- Neptun Data Processing GmbH