

MAGIC

Machine-Assisted Guidance for Integrated Drug Discovery

Programm / Ausschreibung	IWI 24/26, IWI 24/26, Basisprogramm Ausschreibung 2024	Status	laufend
Projektstart	01.09.2024	Projektende	31.12.2025
Zeitraum	2024 - 2025	Projektlaufzeit	16 Monate
Keywords			

Projektbeschreibung

Large Language Models (LLMs) und deren einfache programmatische Ansprechbarkeit via veröffentlichten APIs (OpenAI ChatGPT, u.ä.) eröffnen eine Vielzahl von Möglichkeiten in verschiedensten Bereich. Allerdings sind diese Modelle auf allgemein verfügbaren Informationen trainiert, was sie für den konkreten Einsatz in unseren Datenanalyse und -visualisierungssysteme für die pharmakologische Forschung nicht direkt verwendbar macht.

Wir wollen einen Prototypen für ein Natural Language Interface (NLI) entwickeln, der durch tiefe Einbettung der verfügbaren LLMs in die Applikation und den darunterliegenden biomedizinischen Daten in der Lage ist, auch komplexe biomedizinische Fragestellungen zu beantworten – sofern die Antwort in den Daten enthalten ist. Wir werden evaluieren, in wie fern in der Literatur erwähnte Methoden wie Few-Shot Prompting, Chain of Thought, Tree of Thought, Retrieval-Augmented Generation verwendet und erweitert werden können, damit unsere User:innen inhaltliche Fragen an die 450 GB große biomedizinische Datenbank stellen und die Antworten in unseren Applikationen erhalten können. Wir wollen des weiteren daran forschen, wie wir unsere Forschungsergebnisse aus bisherigen FFG Projekten zum Thema Provenance Tracking dazu einsetzen können, die LLMs noch besser mit Kontext auszustatten.

Unser Ziel ist die Entwicklung und Evaluierung eines NLI-Prototypen, der in beide unsere Produkte integriert werden kann und es den Domänenexpert:innen ermöglicht, Befehle aber auch komplexe Fragestellungen an das System zu richten und dadurch die Applikation zu steuern.

Endberichtkurzfassung

Das Projekt MAGIC konnte erfolgreich demonstrieren, dass Large Language Models (LLMs) wertstiftend in Analyseplattformen für die pharmakologische Grundlagenforschung integriert werden können. Ein zentrales Ergebnis ist die Entwicklung eines Natural Language Interface (NLI), das es ermöglicht, komplexe Analysen über natürliche Sprache zu steuern. Durch eine innovative Agenten-Architektur können über 30 spezifische Analyse-Tools – von der Datenfilterung bis hin zu komplexen Visualisierungen wie Volcano-Plots – angesteuert werden. Dabei erwies sich das System als äußerst robust: Leistungsfähige Modelle erreichten in der systematischen Evaluierung eine Erfolgsquote (Exact Match Rate) von bis zu 99%.

Darüber hinaus wurde eine Lösung für komplexe, mehrstufige wissenschaftliche Fragestellungen realisiert. Durch den Einsatz von Retrieval-Augmented Generation (RAG) und einem Netzwerk spezialisierter KI-Agenten können nun autonome Literaturrecherchen durchgeführt und Antworten direkt mit der zugehörigen Evidenz und passenden Visualisierungen geliefert werden. Ein begleitendes UI/UX-Konzept stellt die notwendige Transparenz sicher, indem es die Interaktion mit der KI für die Forschenden nachvollziehbar macht. Diese Ergebnisse senken die technischen Hürden für die datengestützte Analyse in der Pharmakoforschung massiv und leisten somit einen Beitrag zur Beschleunigung der Entwicklung neuer Medikamente.

Projektpartner

- datavisyn GmbH