Title: Extracting Specific Anosmias as Early Indicators of Neurodegenerative Diseases Using Large Language Models

Abstract:

Olfactory dysfunction, particularly anosmia (smell loss), is a potential indicator for the early stages of neurodegenerative diseases like Alzheimer's and Parkinson's. Despite this significance, identifying specific anosmias (loss of a specific smell) in relation to these diseases remains largely manual and time-consuming in medical research. This project proposes an innovative approach to automate the extraction of olfactory-related indicators of neurodegenerative disease from scientific texts.

This pilot study will use fine-tuned LLMs, such as BERT and GPT-4, to create an automated pipeline for identifying specific anosmias linked to neurodegenerative diseases. Using Named Entity Recognition (NER) and Relation Extraction (RE), the system will extract olfactory dysfunction entities and connect them to disease entities, leading to the discovery of patterns that manual approaches may overlook. By developing a small, annotated dataset of biomedical research articles, we aim to demonstrate the system’s ability in extracting medically relevant information.

Expected outcomes include a fine-tuned LLM capable of identifying anosmia-related terms with high accuracy and a pilot dataset. This approach not only aims develops a pathway towards automated knowledge discovery, but also plans to introduce an innovative tool for advancing early diagnosis of neurodegenerative diseases.