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. Given the growth in the volume of research in the biomedical field, scientists and clinicians must expend much time to identify these important links between smell and neurodegenerative disease. This project proposes an innovative approach to automate the extraction of olfactory-related indicators of neurodegenerative disease from scientific texts in order to create a knowledge base of specific anosmias associated with neurodegenerative diseases.

This pilot study will use fine-tuned LLMs, such as BERT, LLaMA, and GPT, 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 such as anosmia to a specific odorant (peppermint, coffee), loss of detection of a smell category (fruity, floral), partial or distorted senses of smell (hyposmia, parosmia), and validated smell tests (UPSIT, BSIT, Scentinel, Sniffin Sticks). The system will then connect them to disease entities, leading to the discovery of patterns that manual approaches may overlook. The methods will be trained and evaluated on an annotated dataset of biomedical research articles, currently in development.

Expected outcomes include a fine-tuned LLM capable of identifying anosmia-related terms with high accuracy and a pilot dataset. In this poster presentation, I will showcase our initial annotation scheme, designed specifically for detecting anosmias linked to early-stage neurodegenerative diseases. Additionally, I will present preliminary few-shot learning results, demonstrating the model’s ability to extract these terms with minimal supervision.