Computational Detection of Logical Fallacies in Political Discourse Using Natural Language Processing

Objective: To develop and evaluate computational methods for automatically identifying logical fallacies in political discourse, with the goal of enhancing media literacy and promoting more informed democratic participation.

Activities:

- State-of-the-art mapping and existing dataset analysis
- Comprehensive literature review of computational fallacy detection and argument mining research
- Survey and evaluation of available datasets containing annotated political debates and fallacious arguments
- Assessment of current benchmarks, evaluation metrics, and methodological approaches in the field
- Study of methodologies for data analysis and augmentation
- Investigation of natural language processing techniques for argument structure analysis and fallacy identification
- Creation of a silver standard dataset by aggregating and harmonizing multiple existing fallacy detection datasets
- Development of data augmentation strategies to address dataset limitations and class imbalance issues
- Design of robust annotation frameworks for creating high-quality labeled data from political debate transcripts; Experimentation with Large Language Models
- Fine-tuning and adapting large language models for multi-class fallacy detection tasks using the silver standard dataset
- Evaluation of prompt engineering and in-context learning approaches for zero-shot and few-shot fallacy identification
- Comparative performance analysis of different model architectures and training strategies for political argument analysis