DISSERTATION ABSTRACT

"MatHypothesis": Accelerating Materials Discovery with AI-Driven Hypothesis Generation

An AI-Powered System for Generating Novel Materials Science Hypotheses from Research Literature

The exponential growth of scientific literature in materials science presents both opportunities and challenges for researchers. While the wealth of information holds the key to groundbreaking discoveries, the sheer volume of data makes it increasingly difficult for scientists to identify novel connections and generate new hypotheses. This project proposes the development of MatHypothesis, an innovative Large Language Model (LLM) system designed to accelerate scientific discovery in materials science.

MatHypothesis will integrate state-of-the-art natural language processing techniques with domain-specific knowledge to analyze vast corpora of materials science literature. The system will employ a hybrid architecture, combining a general-purpose LLM (such as GPT-4) with a specialized model trained on materials science data (like MatSciBERT). This approach will enable the system to understand complex scientific concepts, identify latent patterns across diverse studies, and generate plausible hypotheses for new materials or properties.

Key features of MatHypothesis will include:

1. Automated literature review and knowledge extraction
2. Cross-domain pattern recognition for novel insights
3. Hypothesis generation with supporting evidence
4. Ranking of generated hypotheses based on novelty and feasibility
5. Explanation of reasoning to ensure transparency and interpretability

The project will involve close collaboration between computer scientists, materials scientists, and domain experts to ensure the system's outputs are both scientifically sound and practically valuable. Rigorous evaluation will be conducted through expert review and experimental validation of selected hypotheses.

By bridging the gap between information overload and scientific intuition, MatHypothesis aims to significantly accelerate the pace of discovery in materials science. This tool has the potential to uncover unexpected relationships, suggest new research directions, and ultimately contribute to breakthroughs in areas such as energy storage, semiconductors, and advanced materials for various applications.

SUBMITTED BY:

ABDU ROUF A (PKD23CSCL01)

MTECH S3 COMPUTATUIONAL LINGUISTICS-GEC PALAKKAD(APJAKTU)