POLIMA
Center for Polariton-driven Light-Matter Interactions

SDU
Mads Clausen Institute

PROJECT TITLE

Large language models for materials discovery.

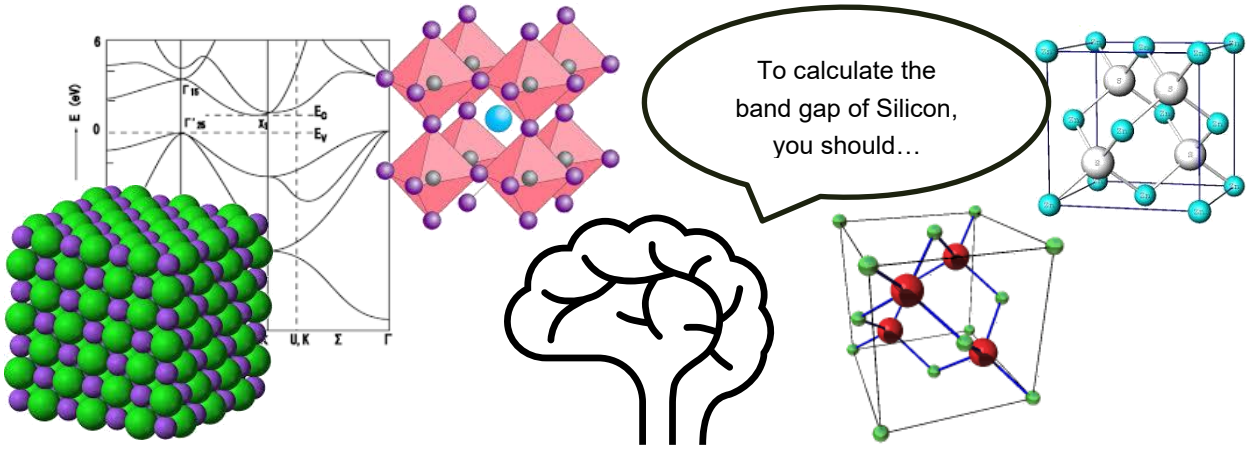
Proposed by: Line Jelver

Possible supervisor(s): Line Jelver

PROJECT DESCRIPTION

This project leverages agentic Large Language Models (LLMs) to autonomously orchestrate and optimize Density Functional Theory (DFT) workflows. By using LLMs not only as intelligent user guides but also as active agents capable of predicting optimal computational settings, the project addresses key barriers related to accessibility, sustainability, and efficiency of material discovery.

Three out of the top-ten cited papers of all time are related to the development of DFT, reflecting its foundational role within physics, chemistry, and biology [1]. DFT is a branch of computational physics that uses quantum mechanical theory to simulate material properties such as electrical conductivity, optical absorption, and chemical reactivity in solids or molecules. These calculations are often referred to as first-principles simulations because they rely solely on knowledge of the constituent atoms, rather than empirical parameters, allowing for the characterization of entirely unexplored materials and thus play a critical role in the development of a wide variety of technologies by enabling researchers to predict material behavior before engaging in costly experimental studies.



Language models are designed to understand, generate, and manipulate human language using advanced neural network architectures and enable users to communicate with and gain assistance from AI [2-5]. LLMs can thus (1) set up workflows and generate input files based on flowing text user requests, (2) through their in-context learning (ICL) capabilities, learn from prior examples and generate more reliable initial parameter guesses. In this project, you will be trained in applying density functional theory (DFT) to enable quantum mechanical simulations of materials properties, such as band gap predictions. The goal will be to tune the prompting and model parameters of different LLMs to optimize their reliability in (1) creating input files for DFT, (2) predicting optimal calculation parameters and (3) provide accurate and user-friendly instructions. The project consists solely of numerical investigations and requires experience with python programing.

[1] R. V. Noorden, Nature 640, 591 (2025.
[2] A. Vaswani et al., arXiv:1706.03762 (2017). [3]. T. B. Brown et al., arXiv:2005.14165 (2020).
[4]. R. Bommasani et al., arXiv:2108.07258 (2022). [5]. H. Touvron et al., arXiv:2307.09288 (2023).