

Our research group is looking for student interested in

Master Thesis

Creating a database on perovskite photovoltaic fabrication by leveraging recent advances in Large Language Models

This project focuses on finding methods of screening the extensive knowledge available on the emerging, vibrant field of perovskite photovoltaics (PVs). Perovskite PVs are a novel semiconductor technology that could contribute significantly to next-generation photovoltaics, raising the achieved power conversion efficiency over the fundamental Shockley-Queisser limit of single junction silicon PVs. Each year, over 5000 scientific articles on this popular technology are released. The specific goal of the project is to automatically extract process parameters out of the fabrication sections of reports on perovskite photovoltaics. The project is structured according to three main steps:

- 1) Interface APIs released by common publisher of scientific articles such as Elsevier, Wiley, Springer and ACS in order to scrape publication texts corresponding to predefined search masks.
- 2) Establish a prompt engineering methodology for generating training data for successful examples of parameter extraction on a subset of the above text corpus from state-of-the-art LLMs. This task is supported by prior student works on finding relevant information in scientific literature.
- 3) Engineer a custom, comparatively small and efficient LLM finetuned on the training data from 2). The goal is to optimize precision and recall, while still maintaining resource efficiency.

The overarching goal of the project is to showcase a transferable methodology for extracting information from scientific papers with only minimal intervention by experts setting constraints of the task. The topic of perovskite photovoltaics is just a starting usecase for this purpose. With the developed model researchers could have facilitated access to state-of-the-art information on their field, using the data to infer new knowledge from data-driven predictions.

Required Skills

Following skills would be helpful in completing the tasks of this thesis.

- Programming skills in python, pytorch.
- Familiarity with version control using Git
- Experience with or motivation to learn working with high performance computational cluster
- Basic understanding of machine learning models

Application

Write us an email at thesis.sne@nano.cit.tum.de

- Attach your Transcript of Records and a CV

Technische Universität München

Professorship of Simulation of Nanosystems for Energy Conversion

Prof. Dr. Alessio Gagliardi

Hans-Piloty-Strasse 1, 85748 Garching bei München

Tel. +49 (89) 289 - 51200

<https://www.ee.cit.tum.de/sne/home/>

www.tum.de