Bo Li

b.li.24@u.nus.edu Harrison-Li Bo Li

Research Interest: AI for materials, Inverse design, Agentic AI, Computational Chemistry, and Semiconductor

Education

National University of Singapore Singapore
Master of Science in Materials Science August 2025

Queen Mary University of London London, United Kingdom
Bachelor of Engineering in Materials Science and Engineering with Honours 2021 – 2025
First Class Honours (Top 10%)

Northwestern Polytechnical University (Joint-program) Xian, China (2021-2025)

Research Experience

Agentic optimization for molecule screening (Still in process)
Supervised by Prof. Peichen Zhong at NUS Aug 1, 2025 – Present

- Still in process.
- Agentic optimization for Proton Electron coupled transfer (PCET) reaction templates for molecular generation based on reducible molecule dataset.
- Computational workflow for conformer search, structure optimization and DFT to calculate the redox potential for PCET molecular pairs.
- Molecule optimization based on evolutionary algorithm combined with crossover and mutation of molecules.

Leverage Transfer Learning and Multi-AI Agent for Design of Self-Assembled Monolayer Molecules (*Graduation thesis*)
Supervised by Prof. Zhe Liu (NWPU), Prof. Steffi Krause (QMUL) Aug 1, 2024 – May 18, 2025

- Tokenized molecule dataset and expanded the vocabulary of a pre-trained model for full character embeddings.
- Applied fine-tuning by freezing various attention blocks and embedding layers to perform transfer learning, attaining fine-tuned model weights.
- Used fine-tuned model weights to generate molecules based on 10 predefined SAM scaffold for structure conditional generation.
- Enabled target property-guided molecule generation through parameter-efficient fine-tuning and classifier-free guidance, allowing the model to accept inputs from both property-labeled and unlabeled molecules.
- Achieved high-quality molecular generation with only 11275 samples : 0.967 validity, 0.635 Uniqueness, 0.948 Novelty and property prediction (16.3% MAPE) requiring only 0.7% of the original dataset's size (1.6 million) by parameter efficient fine-tuning approach.
- Build an AI agent platform with automatic planing and reaction by large language models to invoke tools like: generator, Unimol property predictor, and a Chemprice searcher for experimental users.

Perovskite Multi-AI agent system (In manuscript)
Supervised by Prof. Zhe Liu, NWPU *Oct 30, 2024 - Present*

- Developed a multi-AI agent to automate the materials research pipeline, integrating tools for data extraction, molecular generation, property prediction, precursor supplier search, and Retro-synthesis route plan.
- Applied DataExtractor agent to extract 319 literatures relevant to SAMs in photovoltaic area and collected them into four sub-dataset (1. Material Preparation 2. Material Property 3. Device fabrication 4. Device Performance). Extraction performance by confusion matrices (Precision, Recall, F1 Score), proving that our DataExtractor is capable of extraction with precision $> 84\%$.
- Implemented a Retrieval-Augmented Generation (RAG) to reduce hallucinations in large language models (LLMs) by indexing vectorized data from extracted datasets and abstract sections. And tested their understanding in (1. Material Preparation 2. Material Property 3. Device fabrication 4. Device Performance) compared with general-purpose LLM.
- Demonstrated the system's capability to handle multi-step tasks by designing several complex tasks, and compared their response with general-purpose LLMs.

AI-Assisted Additive Screening for High-Efficiency LED Devices by variational autoencoder
Supervised by Prof. Zhe Liu, NWPU *Jun 13, 2023 – April 18, 2024*

- Developed stratified scaffold splitting method to replace the cross-validation to improve the prediction accuracy (Averagely 4% higher).
- Applied the Variational Autoencoder (VAE) to generate the molecules for the additives.
- Calculated the molecular similarities between the generated molecules and the original additives based on Tanimoto similarity and visualized the similar parts by Rdkit API.
- Extracted the latent representations of molecules inside of the VAE and compare the latent vectors with the common molecular descriptors, Extended-Connectivity Fingerprints (ECFP), Molecular Access System fingerprints (MACCS), and Molecular graph to check the regression model (SVC, MLP, Random forest, Logistic regression) accuracy. This step aims to validate the ability to represent molecular structures by the latent space vector (73% accuracy for both ECFP and latent vectors).
- Created a multi-input Deep neural network (DNN) that applies all representations listed above together to classify the good (10.3%<PCE<22.0%) or bad (PCE<10.3%) property.

Achievements

Scholarship

- QMUL Scholarship for Outstanding Academic Progress Sep 1, 2022 – June 25, 2023
- NWPU campus second class scholarship, merit-based. Sep 1, 2022 – June 25, 2023
- NWPU campus second class scholarship, merit-based. Sep 1, 2023 – June 25, 2024

Skills

Programming skills: Python, MATLAB, Linux and LaTeX. Machine learning and Deep learning Frame: Pytorch, Scikit-learn, Tensorflow.
Computation Chemistry: Gaussian, GULP, Pyscf (Softwares and codes) and DFT (theory)
Language: Chinese(native), English: IELTS total score 6.5. Listening 6.5, Reading 7, Writing 6.5, Speaking 6.5.