Yi Yang

+1-412-736-3800 | yiy5@andrew.cmu.edu | linkedin.com/in/yiyang5 | github.com/Yi5817 | yi5817.github.io

Education

Carnegie Mellon University — Pittsburgh, PA
Doctor of Philosophy (Ph.D.) in Materials Science and Engineering — *Jan. 2023 – Dec. 2026 (expected)*

Carnegie Mellon University — Pittsburgh, PA
Master of Science in Computational Materials Science and Engineering — *Aug. 2021 – Dec. 2022*

China University of Geosciences, Wuhan — Wuhan, China
Bachelor of Science in Materials Science and Engineering (experimental class) — *Sept. 2017 – July 2021*

Publications

[1] Gharakhanyan, V.*, **Yang, Y.***, et al. FastCSP: Accelerated Molecular Crystal Structure Prediction with Universal Model for Atoms. *arXiv:2508.02641* (2025). [Under review, *Nature Machine Intelligence*; ∗ Equal contribution.]

[2] Gharakhanyan, V., **Yang, Y.**, et al. Open Molecular Crystals 2025 (OMC25) Dataset and Models. *arXiv:2508.02651* (2025). [Under review, *Scientific Data*.]

[3] **Yang, Y.**, et al. Genarris 3.0: Generating Close-Packed Molecular Crystal Structures with Rigid Press. *ChemRxiv* (2025). [Accepted by *J. Chem. Theory Comput.*]

[4] Nayal, K. S., **Yang, Y.**, et al. Efficient Molecular Crystal Structure Prediction and Stability Assessment with AIMNet2 Neural Network Potentials. *Crystal Growth & Design* (2025).

[5] Hunnisett, L. M. et al. The Seventh Blind Test of Crystal Structure Prediction: Structure Ranking Methods. *Acta Cryst. B* **80**, 548–574 (2024).

[6] Tom, R., **Yang, Y.**, et al. Inverse Design of Tetracene Polymorphs with Enhanced Singlet Fission Performance by Property-Based Genetic Algorithm Optimization. *Chemistry of Materials* **35**, 1373–1386 (2023).

Research Experience

Research Assistant, Artificial Intelligence for Crystal Structure Prediction — Pittsburgh, PA
Carnegie Mellon University, Advisor: Noa Marom — *Sept. 2021 – Present*

Accelerated Molecular Crystal Structure Prediction with Machine Learning Atomic Potential

- Constructed the **Open Molecular Crystals 2025 (OMC25)** dataset, over **27 million** molecular crystal structures across 12 elements, supporting large-scale training and evaluation of **machine learning interatomic potentials (MLIPs)**.
- Trained and benchmarked state-of-the-art MLIPs (**Universal Models for Atoms (UMA)**), achieving **DFT-level accuracy** and demonstrating superior generalization across targeted molecular crystal tasks.
- Co-developed **FastCSP**, an open-source **Python** crystal structure prediction (CSP) workflow integrating **Genarris 3.0** with **UMA** for high-throughput molecular crystal structure **generation** and **ranking**.
- Collaborated with the **FAIR Chemistry** team to publish and validate **MLIPs transferability** and **robustness** in large-scale FastCSP workflows on **drug-like** and other organic molecular crystals.

Genarris 3.0: Random Molecular Crystal Structure Generator | (https://github.com/Yi5817/Genarris)

- Developed **Genarris 3.0**, an open-source **Python/C** package with **MPI-based parallelism** for large-scale molecular crystal structure generation on **high-performance computing (HPC)** clusters.
- Designed and implemented **Rigid Press**, a novel **geometry optimization algorithm** employing a **regularized hard-sphere potential** to compress molecular crystal unit cells without explicit energy evaluations.
- Integrated **GPU-accelerated MLIPs** and **clustering** for efficient down-selection in the CSP workflow, accelerating exploration of molecular crystal energy landscapes.
- Validated **robustness** of CSP workflow in Genarris 3.0 across multiple **drug-like** molecular crystals and **energetic materials**, supporting future applications in **AI-driven materials discovery**.

The 7^{th} Crystal Structure Prediction Blind Test, Cambridge Crystallographic Data Center

- Pioneered early application of MLIPs for CSP, achieving near-DFT accuracy in **ranking phase** with significantly reduced computational cost.
- Developed **system-specific AIMNet2 MLIPs** trained on DFT datasets of **molecular clusters (*n*-mers)** sampled from **MD** simulations, using **active learning** to improve accuracy and outperforming other ML-based CSP methods.

Projects

Crystal Structure Prediction with Diffusion Model | *10-708, CMU* *Sept. 2024 - Dec. 2024*

- Explored **autoregressive diffusion model** to enhance **conformer generation** accuracy for complex molecular systems.

AI Ramanujan: Discovery of Formula Equivalence | *11-785, CMU* *Oct. 2022 – Dec. 2022*

- Trained **Transformer** encoders to identify equivalent mathematical formulas.
- Designed a **translator** to convert mathematical formulas from Maple to Mathematica format.
- Enhanced the encoders with **cross-attention** to capture correlations between embeddings.
- Employed **contrastive learning** for model training to improve performance.

JPX Tokyo Stock Exchange Prediction *May 2022 – July 2022*

- Engineered key financial features from JPX data for trend forecasting.
- Developed a **Long short-term memory (LSTM)** model with PyTorch for predicting JPX stock trends, ranking top **13%** on Kaggle.

Honors & awards

MLCM Rising Star Award, Los Alamos National Lab	May 2025
Frontera Computational Science Fellowship, TACC	Apr. 2025
Award for Research Excellence in the Master Program, CMU	May 2023
ATK-Nick G. Vlahakis Graduate Fellowship, CMU	Feb. 2023
Outstanding Undergraduate Award, CUG	June 2021

Technical Skills

Programming Languages: Python, C, R
Frameworks & Libraries: PyTorch, TensorFlow, ASE, Pymatgen, RDKit, OpenBabel, PySpark
Systems & Tools: Linux, High-Performance Computing (HPC), MPI, AWS
High Performance Computing: Bridges2 (PSC), Frontera (TACC), Vista (TACC), Perlmutter (NERSC)
Scientific Software: FHI-aims, VASP, Orca
Machine Learning Models: Transformer, Graph Neural Network (GNN), LSTM
Domains: Scientific Computing, Computational Materials, Drug Discovery
Languages: English (Fluent), Chinese (Native Speaker)