

Large Language Models for Materials Discovery

S. Zollner, B. Kiefer, I. Vasiliev, M. Zhou, L. Miao, New Mexico State University

We used the Microsoft MatterGen framework and a pre-trained bandgap model to establish a workflow that utilizes ChatGPT style large language models for predicting new materials with constraint properties.

- Ported MatterGen successfully from GitHub to NMSU compute resource, Discovery.
- Tested implementation on Discovery, generated 1000 materials with constrained bandgap: $3 \text{ eV} < E_g < 5 \text{ eV}$, a range that is of interest for quantum sensing and quantum computing technologies.
- Timing information shows that it takes $\sim 18 \text{ s}$ to generate one material.

This result will enable us to expand research into two directions:

- Train models for custom properties.
- Better understand trade-offs between training set size and prediction error.

By undergraduate students
Daniel Reyes and Gavin Campos.

