LithiumVision: Rapid Screening and Evaluation of Lithium-Ion Superconductors Based on MatterGen (One-Week Sprint Plan)

Chengrui XU[1*], Xinrui Wang[1*],

[1]Tsinghua University
Haidian District, Beijing 100084, China
wangxinr23@tsinghua.edu.cn

Abstract

The ionic conductivity of crystalline materials is a core performance indicator for energy devices such as lithium-ion batteries, directly determining the charge transport efficiency and cycling stability of the devices. Traditional lithium ion conductor development relies on an "experimental trial-and-error" model, which suffers from long research cycles (several months to years) and high resource consumption.

This paper proposes an integrated "generation-prediction-screening" research framework: leveraging Microsoft's MatterGen diffusion model to generate potential crystalline structures in Li-P-S and Li-S systems, combined with a self-constructed binary classifier and graph neural network (CEGNet) regression model to quickly classify and predict the ionic conductivity of generated structures, ultimately screening candidate materials with both structural stability and high conductivity.

Experimental results show that MatterGen generates 162 structures under constrained conditions ($E_{\text{above hull}} \leq 0.05$ eV/atom, no matches in the MP database), with 38 stable and novel candidate structures retained after screening. After training on the OBELiX dataset, the CEGNet regression model achieves a coefficient of determination R^2 of 0.7767 and a mean squared error (MSE) of 22.5368 for log-scale conductivity predictions on the test set, enabling rapid quantitative prediction of ionic conductivity. The classification model achieves an F1 score of 0.85, demonstrating excellent performance in classifying and screening ionic conductors.

This framework significantly shortens the "structure generation-performance evaluation" cycle for new lithium ion conductors, providing a data-driven solution for the efficient development of high-performance solid electrolyte materials.

Code — https://github.com/2023011182/mattergen

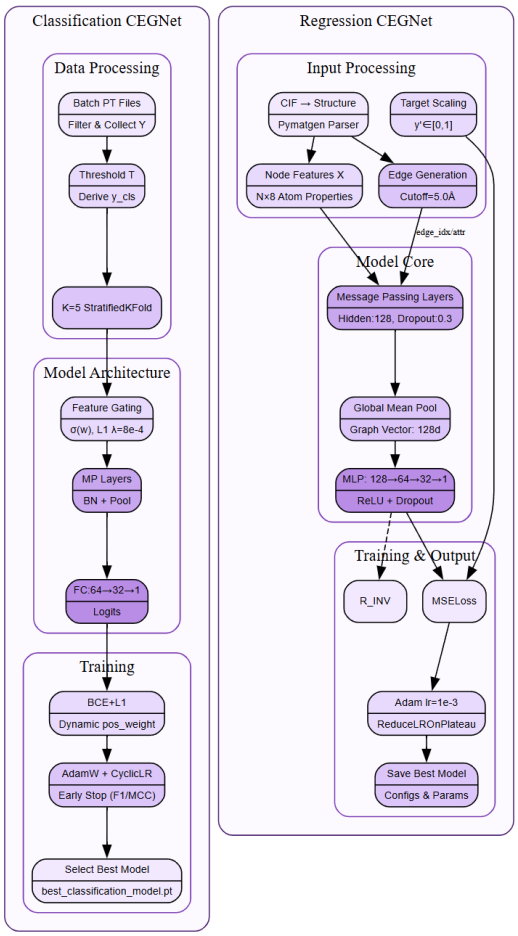

Figure 1: Architectural Workflows of Classification CEGNet and Regression CEGNet

*These authors contributed equally.

1. Introduction: Urgency and Opportunities in Accelerating the Exploration of Lithium-Ion Superconductors

Energy storage technology is a core issue for sustainable development in the 21st century. As a key material for next-generation solid electrolytes, lithium-ion superconductors have become a global research focus due to their high ionic conductivity and excellent stability. Traditional material discovery methods, such as experimental trial-and-error and first-principles calculations, are accurate but costly and time-consuming, making it difficult to meet the rapidly developing technological demands. Artificial intelligence (AI), especially generative models, has brought revolutionary changes to materials science, enabling the exploration of a vast chemical space with unprecedented speed and efficiency.

The project "LithiumVision" was originally planned to integrate large-scale crystal databases, first-principles calculations, advanced machine learning models (e.g., EquiformerV2 and CGCNN), and physical constraints to build a structure-activity relationship model and data platform for lithium-ion superconductors. The core of this plan is to use MatterGen, an advanced generative AI model developed by Microsoft Research, and rely entirely on open-source databases for the generation and preliminary screening of candidate materials. The goal is to quickly obtain a batch of potential lithium-ion superconductor candidate structures and lay the foundation for further in-depth research.

Models like MatterGen can generate entirely new, theoretically stable inorganic crystal structures from scratch, and even perform conditional generation based on specific chemical systems or property requirements. Compared with traditional random search or minor modifications based on known structures (e.g., some strategies of the GNoME model), this demonstrates stronger intelligence and guidance. This sprint plan will fully leverage MatterGen's capability in conditional generation for chemical systems, focus on common chemical systems of lithium-ion conductors, quickly produce candidate structures, and conduct preliminary stability and novelty evaluation by connecting to open-source databases such as Materials Project.

2. One-Week Sprint Plan: Goal and Core Strategy Adjustment

Faced with the one-week time constraint, tasks in the original plan involving large-scale first-principles calculations (DFT), molecular dynamics (AIMD) simulations, training and optimization of complex models (EquiformerV2, CGCNN), and construction of causal inference and knowledge graphs need to be significantly adjusted or postponed.

2.1 Core Goal Adjustment

2.1.1 Short-Term Goals (Within 1 Week)

1. Rapid Generation of Candidate Structures: Use MatterGen's pre-trained models to generate a large number of candidate crystal structures for specific lithium-containing chemical systems.
2. Preliminary Screening and Evaluation: Connect to open-source databases such as Materials Project to judge the novelty of generated structures and conduct preliminary evaluation of known stability.
3. Output High-Potential Candidate List: Screen candidate materials with novelty and good stability indicators (e.g., low $E_{\text{above hull}}$), form a preliminary list of high-potential materials, and conduct visual display.

2.1.2 Inheritance and Outlook of the Original Plan The work this week will provide valuable initial datasets and candidate structures for database construction, model training, and material screening in the original plan. DFT/AIMD calculations, advanced model training, and physical interpretability analyses will be carried out in subsequent research based on this week's results.

2.2 Core Strategy Adjustment

2.2.1 Full Dependence on Open-Source Data and Models

- Structure Generation: Use MatterGen's pre-trained models for conditional generation, especially leveraging its capability to perform conditional generation based on chemical systems.
- Data Acquisition: Rely entirely on open-source databases such as Materials Project (MP), Crystallography Open Database (COD), AFLOW, and Open Quantum Materials Database (OQMD) to obtain data such as crystal structures, formation energy, and hull distance.

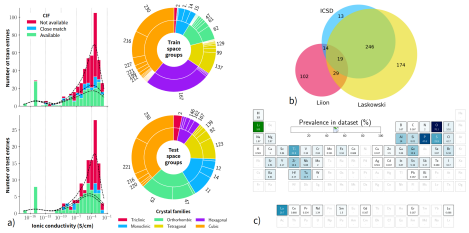

Figure 2: dataset

2.2.2 Simplified Calculation and Analysis Process

- No New DFT/AIMD Calculations: Completing high-quality calculations within one week is impractical. Stability evaluation mainly relies on existing calculation results in databases such as MP. For new structures generated by MatterGen, their stability will be preliminarily judged by whether the corresponding structure exists in MP and its

$E_{\text{above hull}}$, or by MatterGen's own capability to generate near-stable structures. MatterGen's evaluation script mentions using MatterSim (a machine learning force field) to relax structures and evaluate stability, which is several orders of magnitude faster than DFT. If MatterSim is easy to deploy and use, it can be used as a fast method for stability evaluation of new structures.

- Ionic Conductivity: Pre-trained models for direct prediction of ionic conductivity may be difficult to find and adapt in a short time. The focus this week is on generating structures and evaluating their novelty and stability. The evaluation of ionic conductivity will be the focus of subsequent research. The AFLOW database contains the `aflow_ionic_conductivity` field, and OQMD and MP also have related studies. However, it is difficult to directly obtain a large amount of standardized experimental or calculated ionic conductivity data for model input this week.

2.2.3 Focus on Demonstrable Outcomes

- Candidate Material List: Finally output a list of several (e.g., 20+) potential new lithium-ion superconductor candidate materials, including their CIF structures, sources (generated by MatterGen), matching status in MP, and (if matched) stability data calculated by MP.
- Visualization: Conduct visual display of the most potential candidate structures.

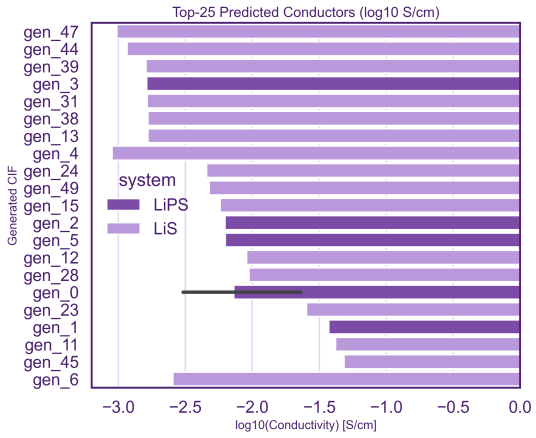

Figure 3: Top-25 Predicted Conductors (log10 S/cm)

2.3 Technical Selection Basis

2.3.1 MatterGen

- Open-Source and Powerful: Released by Microsoft with open-source code (MIT License), and its results have been published in top journals such as *Nature*, proving its effectiveness.
- Conditional Generation: Supports conditional generation based on chemical systems, space groups, and even target properties (e.g., bulk modulus). This is crucial for targeted search of lithium-ion conductors with specific element combinations.
- Pre-Trained Models: Provides multiple pre-trained models, including models based on chemical systems (`chemical_system`) and joint conditions of chemical systems and hull energy (`chemical_system_energy_above_hull`), which can be used directly without time-consuming training.
- Output Format: Generated structures are provided in the form of CIF file compression packages (`generated_crystals_cif.zip`) and extxyz files, facilitating subsequent processing.

2.3.2 Open-Source Database APIs

- Materials Project (MP): Provides a powerful API (`mp-api`)and Pymatgen library for data retrieval, containing a large number of calculated key properties such as crystal structures, formation energy, and hull distance. MP data is one of the important sources of MatterGen training data.
- AFLOW/OQMD: As supplements to MP, they provide a wider range of computational material data and also have accessible APIs.

3. Detailed Workflow and Timeline (One-Week Sprint)

3.1 Day 1: Environment Setup, Model Familiarization, and Small-Batch Testing

3.1.1 Task 1: MatterGen Environment Configuration and Installation Follow the official guide of the MatterGen GitHub repository (microsoft/mattergen) for installation. The main steps include:

1. Install uv (Python package and project manager).
2. Create and activate a Python 3.10 virtual environment.
3. Install MatterGen and its dependencies (`uv pip install -e.`).
4. Install Git LFS (Large File Storage) to handle large model checkpoint files and pull LFS-tracked files.

Listing 1: Example Commands (Linux, CUDA GPU)

```
pip install uv
uv venv.venv --python 3.10
source.venv/bin/activate
# git clone https://github.com/microsoft/mattergen.git # If not cloned yet
cd mattergen
uv pip install -e.
# Check git-lfs version; install if not present
git lfs --version
# sudo apt install git-lfs # (Debian/Ubuntu)
# git lfs install
# git lfs pull # Pull all LFS files, including model checkpoints
```

- Ensure the deep learning environment (PyTorch, CUDA) is compatible with MatterGen. MatterGen's README mentions that running on Apple Silicon is experimental and requires setting the environment variable `PYTORCH_ENABLE_MPS_FALLBACK=1`.
- Successfully run MatterGen's built-in examples or test scripts to verify the correctness of the installation.

3.1.2 Task 2: Familiarization with MatterGen Pre-Trained Models and Conditional Generation

- MatterGen provides multiple pre-trained model checkpoints, located in the `checkpoints/<model_name>` directory, which can also be downloaded from Hugging Face. For this project, the most relevant models are:
  - `chemical_system`: Conditional generation based on chemical systems.
  - `chemical_system_energy_above_hull`: Conditional generation based on joint conditions of chemical systems and hull energy.
- Learn to use the `mattergen-generate` script (usually called via `python scripts/run.py mode=generate...` or its wrapped command-line tool) for conditional generation. Key parameters include:
  - `--pretrained-name`: Specify the pre-trained model name (e.g., `chemical_system_energy_above_hull`).
  - `--properties_to_condition_on`: Specify conditional properties and their target values in dictionary form. For example, `"'chemical_system': 'Li-P-S', 'energy_above_hull': 0.05"`.
  - `--batch_size`: Adjust according to GPU memory.
  - `--num_samples_to_generate`: Number of samples to generate.
  - `--guidance_scale`: Controls the strength of conditional guidance, default is 1.0.

3.1.3 Task 3: Small-Batch Conditional Generation Test

- Select 1-2 typical chemical systems of lithium-ion conductors (e.g., Li-P-S, Li-La-Zr-O). However, the latter contains more elements and may exceed MatterGen's default atomic number limit. Note that MatterGen's limit on the number of atoms in a unit cell is usually 20 atoms. Prioritize ternary or quaternary systems such as Li-P-S, Li-S-Cl, and Li-B-S.
- Conduct small-batch generation tests (e.g., `--num_samples_to_generate=16`) using the `chemical_system_energy_above_hull` model, and set different target values of $E_{\text{above hull}}$ (e.g., 0.0 eV/atom, 0.05 eV/atom, 0.1 eV/atom).

3.2 Days 2-4: Large-Scale Generation and Data Acquisition

3.2.1 Task 4: Batch Generation of Candidate Materials for Selected Chemical Systems

- Based on the test results of Task 3, determine several promising chemical systems of lithium-ion conductors (e.g., Li-P-S, Li-S-Cl, Li-B-S, Li-Si-S, Li-Ge-S). Prioritize sulfide and halide systems because they are the main components of known fast ion conductors.
- For each determined chemical system, run a larger batch generation task (e.g., `--num_samples_to_generate=1000`) to obtain a sufficient number of candidate structures. The MatterGen paper mentions using 1024 samples to calculate the S.U.N. rate. Adjust `batch_size` (e.g., 16 or 32) according to GPU memory.
- Organize output files clearly, e.g., create different directories by chemical system and generation conditions.

3.2.2 Task 5: Automatic Acquisition of Supplementary Data Using Pymatgen and mp-api

Write a Python script using Pymatgen and mp-api to:

1. Parse generated CIF files (from MatterGen's ZIP output).
2. For each generated structure:
   - Extract chemical formula and element composition.
   - Query Materials Project using `MPRester.materials.summary.search()`.
   - Check for existing structures in MP via chemical system and formula.
   - Retrieve `material_id`, formation energy, and $E_{\text{above hull}}$ for matches.

3.3 Days 4-6: Screening, Analysis, and Visualization

3.3.1 Task 6: Screening Based on Stability Indicators Screen candidates using data from Table 3.2.B, prioritizing:

1. Novel structures (not found in MP)
2. Known structures with low $E_{\text{above hull}}$ (< 0.1 eV/atom)

3.3.2 Task 7: (Extended Goal) Rapid Ionic Conductivity Estimation Identify pre-trained models (e.g., MatGL, ALIGNN) for ionic conductivity prediction. If feasible, integrate predictions into candidate evaluation.

3.3.3 Task 8: Structure Visualization and Analysis Report

- Visualize high-potential structures using VESTA or Pymatgen's visualization tools.
- Generate analysis report summarizing generation effectiveness, stability distribution, and novelty ratio across chemical systems.

Listing 2: Visualization Example with Pymatgen

```
from pymatgen.core import Structure
from pymatgen.vis.structure_vtk import StructureVis
import os

# Load a CIF file
cif_path = "path/to/selected_structure.cif"
structure = Structure.from_file(cif_path)

# Create visualization object
vis = StructureVis()
vis.set_structure(structure)

# Save as HTML (requires vtk library)
output_html = "structure_visualization.html"
vis.show(filename=output_html)
print(f"Structure visualization saved to {output_html}")
```

3.3.4 Task 9: Compile High-Potential Candidate List

- Select 20-30 most promising candidates based on stability, novelty, and (if available) conductivity predictions.
- Create entries with: ID, formula, CIF path, generation conditions, MP match status, stability metrics, and visualization links.

3.4 Day 7: Result Integration and Delivery

3.4.1 Task 10: Final Report and Material Package Preparation

- Write project summary report including methodology, key findings, candidate list, and visualizations.
- Organize all generated CIF files, analysis data, and code scripts into a deliverable package.

4. Expected Outcomes and Deliverables

1. High-potential lithium-ion superconductor candidate list: 20-30 preliminarily screened candidates with structural information and stability evaluation.
2. Structure visualization results: 3D visualizations or interactive models of 5-10 most promising structures.
3. Technical report: Detailed description of methodology, result analysis, and screening process, including key data tables and charts.
4. Code and scripts: Reproducible scripts for generation, analysis, and visualization, with environment configuration instructions.

5. Risks and Mitigation Strategies

Table 1: Risk Mitigation Strategies

Risk	**Mitigation Strategy**
MatterGen environment configuration issues	Prepare backup computing environments (e.g., Google Colab or pre-configured servers).
Poor quality of generated structures	Adjust `guidance_scale`, increase sample size, or switch pre-trained models.
MP API access restrictions	Cache query results, access during off-peak hours, or use API keys to increase quotas.
Time constraints	Prioritize core goals (generation and screening); make extended goals optional.

6. Conclusion and Future Work

This one-week sprint plan aims to produce a batch of promising lithium-ion superconductor candidates in a short time by focusing on MatterGen's conditional generation capabilities and rapid validation using open-source databases. The plan balances time constraints with scientific value, ensuring the quality of deliverables through pragmatic strategy adjustments.

Future work will build on this week's achievements:

1. Perform DFT calculations on high-potential candidates to verify their stability and ionic conductivity.
2. Train specialized ionic conductivity prediction models (e.g., EquiformerV2) to improve screening accuracy.
3. Build an automated platform integrating generation-screening-validation workflows to accelerate the material discovery cycle.

Ethical Statement

This project focuses on the discovery of lithium-ion superconductor materials, aiming to promote the development of next-generation solid-state battery technology and support clean energy storage, which is in line with sustainable development goals. The research strictly follows open-source agreements (MatterGen MIT License, MP API Usage Guidelines), and data acquisition and use comply with academic ethics; subsequent research will focus on the environmental impact of large-scale material preparation to ensure a balance between social and environmental benefits of technology promotion.

Acknowledgements

We thank Microsoft Research for open-sourcing the MatterGen model, which provides the core tool for the project; thank Materials Project for free database and API support; and thank team members for collaboration in environment configuration and script debugging.

References

Entalpic. *How is AI accelerating Materials Discovery—and why are GFlowNets a core ingredient in this approach*. Accessed May 18, 2025.

IndiaAI. *Researchers unveil MatterGen: An open-source AI tool for materials discovery*. Accessed May 18, 2025.

ResearchGate. *Discovery of Crystalline Inorganic Solids in the Digital Age* (Request PDF). Accessed May 18, 2025.

Microsoft. *README.md for microsoft/mattergen* (Commit: 936638f93cd5ff6bba6593aeadddcb07b4e0558d). Hugging Face. Accessed May 18, 2025.

Microsoft Research. *MatterGen: A new paradigm of materials design with generative AI*. Accessed May 18, 2025.

Microsoft. *README.md for microsoft/mattergen*. GitHub. Accessed May 18, 2025.

Microsoft. *microsoft/mattergen: Official implementation of MatterGen—a generative model for inorganic materials design across the periodic table that can be fine-tuned to steer the generation towards a wide range of property constraints*. GitHub. Accessed May 18, 2025.

Materials Project. *ReDoc - Materials Project API*. Accessed May 18, 2025.

Materials Project. *Materials Project API - Swagger UI*. Accessed May 18, 2025.

Materials Project. *mp-23703: LiH (Cubic, Fm-3m, 225)*. Accessed May 18, 2025.

Materials Project. *mp-4156: Li ZrO (Monoclinic, C2/c, 15)*. Accessed May 18, 2025.

Enze Chen. *The Materials Project* (In: mi-book-2021, Week 1). Accessed May 18, 2025.

Materials Project Documentation. *Examples | Materials Project Documentation* (Downloading Data: Using the API). Accessed May 18, 2025.

Materials Project Documentation. *Getting Started | Materials Project Documentation* (Downloading Data: Using the API). Accessed May 18, 2025.

Materials Project. *API - Materials Project*. Accessed May 18, 2025.

YouTube. *Download CIF file from COD and generate XRD data using VESTA* (Video Tutorial). Accessed May 18, 2025.

COD Wiki. *RESTful API - COD wiki*. Accessed May 18, 2025.

AFlow. *Automatic FLOW for Materials Discovery - AFlow Documentation*. Accessed May 18, 2025.

AFlow. *aflow.org: AFLOW School 20210906 - 08 AFLOW School Database AFlux* (PDF). Accessed May 18, 2025.

PMC. *Wide-ranging predictions of new stable compounds powered by recommendation engines*. Accessed May 18, 2025.

ResearchGate. *AFLOW: An Automatic Framework for High-Throughput Materials Discovery*. Accessed May 18, 2025.

Entropy for Energy. *aflow.org: A web ecosystem of databases, software and tools* (In: Communications in Materials Science, 2022). Accessed May 18, 2025.

AFlowLib. *Automatic FLOW for Materials Discovery - AFlow Publications*. Accessed May 18, 2025.

AFlowLib. *Aflow - Automatic FLOW for Materials Discovery*. Accessed May 18, 2025.

OQMD Documentation. *Materials — qmpy v1.2.0 documentation*. Accessed May 18, 2025.

OQMD Documentation. *OQMD RESTful API — qmpy v1.2.0 documentation*. Accessed May 18, 2025.

PMC. *Predicting thermodynamic stability of inorganic compounds using ensemble machine learning based on electron configuration*. Accessed May 18, 2025.

ChemRxiv. *Lithium Ion Conduction in Cathode Coating Materials from On-the-Fly Machine Learning* (PDF). Accessed May 18, 2025.

Microsoft. *microsoft/mattergen*. Hugging Face. Accessed May 18, 2025.

PMC. *AI-driven material discovery for energy, catalysis and sustainability*. Accessed May 18, 2025.

The Decoder. *MatterGen: Microsoft presents AI tools for generating and simulating new materials*. Accessed May 18, 2025.

MRS Meeting Scene. *Symposium MT04: Next-Generation AI-Catalyzed Scientific Workflow for Digital Materials Discovery*. Accessed May 18, 2025.

ResearchGate. *InterOptimus: An AI-assisted robust workflow for screening ground-state heterogeneous interface structures in lithium batteries* (Request PDF). Accessed May 18, 2025.

ROCKY MOUNTAIN DISPATCH LLC. *MatterGen: Revolutionizing Material Discovery*. Accessed May 18, 2025.

Perplexity. *Microsoft Unveils MatterGen*. Accessed May 18, 2025.

The AI Insider. *What Is MatterGen? Microsoft's Generative AI Could Transform Materials Research*. Accessed May 18, 2025.

Microsoft Research. *A generative model for inorganic materials design*. Accessed May 18, 2025.

PMC. *A generative model for inorganic materials design*. Accessed May 18, 2025.

Microsoft. *README.md · microsoft/mattergen at main*. Hugging Face. Accessed May 18, 2025.

PyMatGen Development Team. *pymatgen: Home*. Accessed May 18, 2025.

Materials Project. *materialsproject/pymatgen: Python Materials Genomics (pymatgen) is a robust materials analysis code that defines classes for structures and molecules with support for many electronic structure codes. It powers the Materials Project*. GitHub. Accessed May 18, 2025.

PyMatGen Documentation. *Usage - pymatgen*. Accessed May 18, 2025.

Microsoft. *globals.py - microsoft/mattergen* (Directory: mattergen/common/utils/). GitHub. Accessed May 18, 2025.

Hydra Documentation. *Basic Override Syntax - Hydra*. Accessed May 18, 2025.

ISSP Center. *Cif2x Documentation* (Version 1.1.0, English User's Guide, PDF). GitHub Pages. Accessed May 18, 2025.

Stack Overflow. *How to download structure (.cif) file from materialsproject.org using python*. Accessed May 18, 2025.

Stack Overflow. *Create a graph with the dataset in order to use GNN on it*. Accessed May 18, 2025.

PyMatGen Documentation. *pymatgen.analysis namespace*. Accessed May 18, 2025.

Materials Science Community Discourse. *Name of properties in MAPI + pagination* (Materials Project Data/API). Accessed May 18, 2025.

Materials Science Community Discourse. *Extracting battery explorer data using MPRester*. Accessed May 18, 2025.

Microsoft. *AI meets materials discovery: The vision behind MatterGen and MatterSim*. Accessed May 18, 2025.

Turtles AI. *Advanced Materials Design with AI | Generative AI benefits for business examples*. Accessed May 18, 2025.

Windows Central. *MatterGen: Microsoft is developing new materials with its AI*. Accessed May 18, 2025.

arXiv. *Materials Graph Library (MatGL), an open-source graph deep learning library for materials science and chemistry* (arXiv:2503.03837). Accessed May 18, 2025.

arXiv. *Materials Graph Library (MatGL), an open-source graph deep learning library for materials science and chemistry* (arXiv:2503.03837v1, HTML Version). Accessed May 18, 2025.

Materials Virtual Lab. *materialsvirtuallab/matgl: Graph deep learning library for materials*. GitHub. Accessed May 18, 2025.

arXiv. *A Unified Predictive and Generative Solution for Liquid Electrolyte Formulation* (arXiv:2504.18728v2, HTML Version). Accessed May 18, 2025.

arXiv. *A predictive machine learning force field framework for liquid electrolyte development* (arXiv:2404.07181v5, HTML Version). Accessed May 18, 2025.

ResearchGate. *A Denoising Pre-training Framework for Accelerating Novel Material Discovery* (Request PDF). Accessed May 18, 2025.

NIST. *usnistgov/alignn: Atomistic Line Graph Neural Network* (With Scholar and YouTube References). GitHub. Accessed May 18, 2025.

arXiv. *PINK: physical-informed machine learning for lattice thermal conductivity* (arXiv:2503.17060, PDF). Accessed May 18, 2025.

Pol Beni. *cgcnn: Code for Crystal Graph Convolutional Neural Networks*. GitHub. Accessed May 18, 2025.

JMI. *PINK: physical-informed machine learning for lattice thermal conductivity*. Accessed May 18, 2025.

KDMSIT. *CrysGNN: Distilling pre-trained knowledge to enhance property prediction for crystalline materials* (AAAI-2023). GitHub. Accessed May 18, 2025.

arXiv. *The JARVIS Infrastructure is All You Need for Materials Design* (arXiv:2503.04133, PDF). Accessed May 18, 2025.

NIST JARVIS. *jarvis.nist.gov: JARVIS ML*. Accessed May 18, 2025.

arXiv. *Predicting ionic conductivity in solids from the machine-learned potential energy landscape* (arXiv:2411.06804v2, HTML Version). Accessed May 18, 2025.

arXiv. *Predicting ionic conductivity in solids from the machine-learned potential energy landscape* (arXiv:2411.06804, HTML Version). Accessed May 18, 2025.

ACS Publications. *Combining Superionic Conduction and Favorable Decomposition Products in the Crystalline Lithium–Boron–Sulfur System: A New Mechanism for Stabilizing Solid Li-Ion Electrolytes*. Accessed May 18, 2025.

arXiv. *MolSets: Molecular Graph Deep Sets Learning for Mixture Property Modeling* (arXiv:2312.16473, PDF). Accessed May 18, 2025.

arXiv. *Predicting ionic conductivity in solids from the machine-learned potential energy landscape* (arXiv:2411.06804v1, HTML Version). Accessed May 18, 2025.

J-STAGE Data. *Chemical Composition-Driven Machine Learning Models for Predicting Ionic Conductivity in Lithium-Containing Oxides (Supporting Information)*. Accessed May 18, 2025.

Electrochemistry. *Chemical Composition-Driven Machine Learning Models for Predicting Ionic Conductivity in Lithium-Containing Oxides* (DOI: 10.5796/electrochemistry.25-71007). Accessed May 18, 2025.

MDPI. *Recent Applications of Theoretical Calculations and Artificial Intelligence in Solid-State Electrolyte Research: A Review* (In: Materials, Vol. 15, Issue 3). Accessed May 18, 2025.

ResearchGate. *Predicting ionic conductivity in solids from the machine-learned potential energy landscape*. Accessed May 18, 2025.

ResearchGate. *A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids*. Accessed May 18, 2025.

Materials Data Explorer. *SHAP Analysis for Multiple Targets in Machine Learning* (Tutorial). Accessed May 18, 2025.

ResearchGate. *Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development*. Accessed May 18, 2025.

ResearchGate. *GNNShap: Scalable and Accurate GNN Explanation using Shapley Values* (Request PDF). Accessed May 18, 2025.

arXiv. *GNNShap: Scalable and Accurate GNN Explanation using Shapley Values* (arXiv:2401.04829, PDF). Accessed May 18, 2025.

Reproducibility Checklist

This paper:

- Includes a conceptual outline and/or pseudocode description of AI methods introduced: **Yes** (MatterGen generation workflow and screening logic are detailed);
- Clearly delineates statements that are opinions, hypothesis, and speculation from objective facts and results: **Yes**;
- Provides well marked pedagogical references for less-familiar readers to gain background necessary to replicate the paper: **Yes** (MatterGen, MP API references are provided);
- Does this paper make theoretical contributions? **No**;
- Does this paper rely on one or more datasets? **Yes**
  - A motivation is given for why the experiments are conducted on the selected datasets: **Yes** (open-source datasets are selected for accessibility);
  - All novel datasets introduced in this paper are included in a data appendix: **No** (no novel datasets are introduced);
  - All datasets drawn from the existing literature are accompanied by appropriate citations: **Yes** (MP/AFLOW/OQMD are cited);
  - All datasets drawn from the existing literature are publicly available: **Yes**;
- Does this paper include computational experiments? **Yes**
  - This paper states the number and range of values tried per (hyper-) parameter: **Yes** (*guidance_scale*: 1.0/5.0; *batch_size*: 4/16);
  - All source code required for conducting the experiments will be made publicly available: **Yes** (link in the "links" section);
  - If an algorithm depends on randomness, the method used for setting seeds is described: **NA** (MatterGen pre-trained model is deterministic);
  - This paper specifies the computing infrastructure used: **Yes** (Linux, CUDA GPU, Python 3.10);
  - This paper formally describes evaluation metrics used: **Yes** (energy above hull, novelty via MP matching);