LLEMA: Evolutionary Search with LLMs for Multi-Objective Materials Discovery

Nikhil Abhyankar, Sanchit Kabra, Saaketh Desai, Chandan K. Reddy




VIRGINIA TECH
Sandia National Laboratories

Materials Discovery

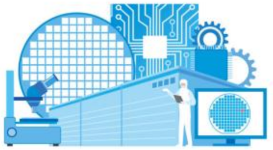
Why does materials discovery matter?

From the Bronze Age to modern day Semiconductors - advanced materials have always powered human innovation.

Advanced materials find use across aerospace, medicine, semiconductors, energy, and other domains, improving performance and overcoming limitations.

Medicine

Aerospace



Semiconductors

Construction

Energy

AI-driven inverse design of materials: Past, present and future

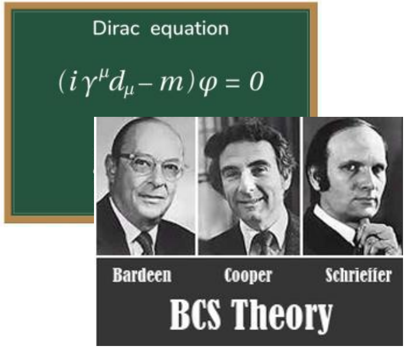
Materials Discovery Paradigms

(A) Experiment-driven

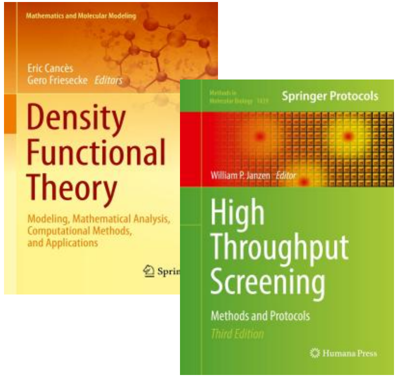
Madam Curie

Heike Onnes

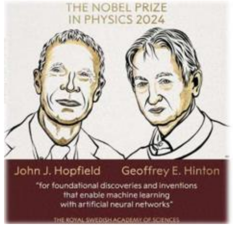
(B) Theory-driven

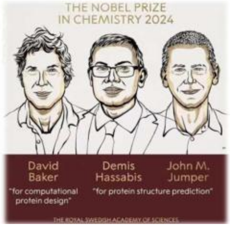


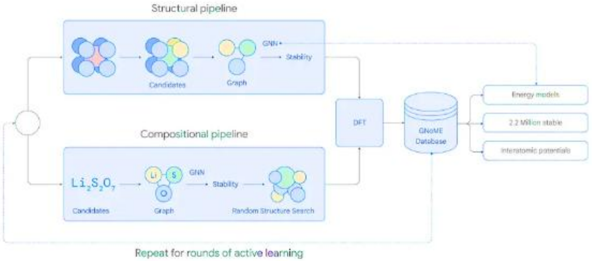
(C) Computation-driven



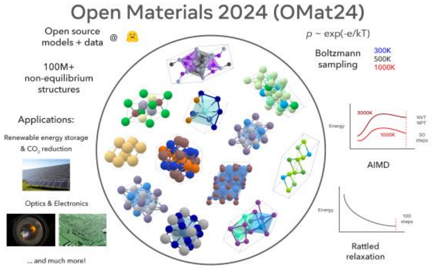
(D) AI-driven







Google GNoME



Meta OMat

AI-driven inverse design of materials: Past, present and future

AI-Driven Material Discovery

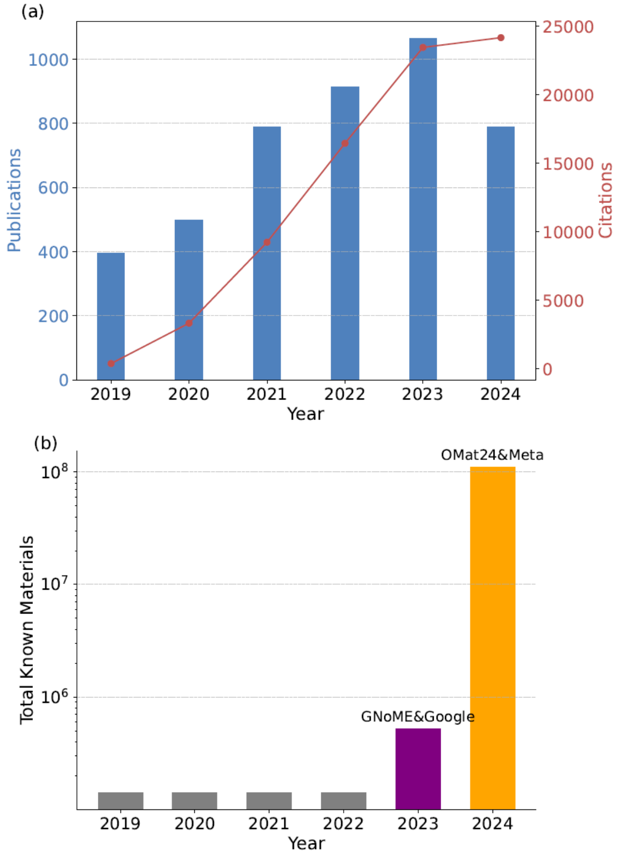
- **Traditional discovery methods** are **slow**, **costly**, and **labor-intensive**, unable to match today’s pace of innovation.

- AI enables rapid pattern recognition, uncovering hidden links between crystal structures and material properties.

- **Machine learning expedites forecasting** properties like mechanical strength and stability, dramatically speeding up development cycles.

- Consequently, the works focusing on ‘machine learning material discovery’ has been garnering escalating research interest.

- AI-driven methods like **GNoME and OMat24 has increased the pace of materials discovery process**.

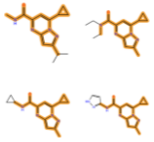


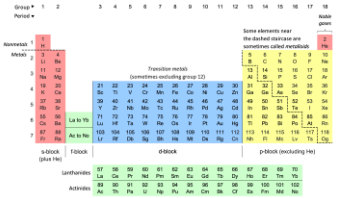
AI-driven inverse design of materials: Past, present and future

Material Discovery | Challenges

Search Space

The combinatorial explosion of chemical and structural possibilities yields a search space exceeding 10^{60} potential materials.





Resource Intensive

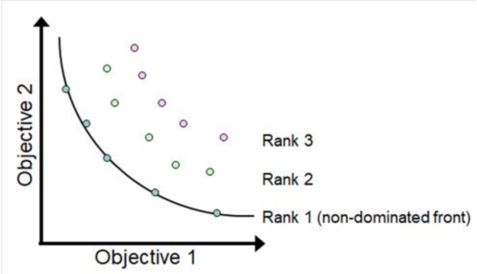
Existing methods are slow and labor-intensive relying on trial-and-error experiments.

Material Validity

Generated materials need to follow the task-dependent property constraints.

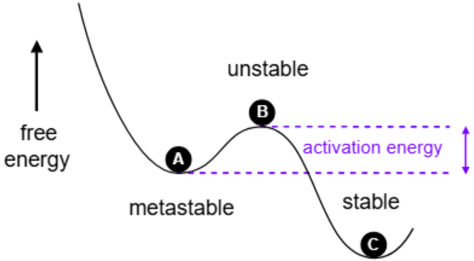
Multi-objective optimization

Most tasks require satisfying multiple conflicting property objectives. E.g., high strength and high ductility.



Thermodynamic Stability

Poor data quality can lead to inaccurate or biased models



Current Methods for Material Discovery

Does not incorporate domain knowledge

Does not generate stable materials

Generative Methods

Performs random search

- Randomly searches through the chemical space.

Lacks domain knowledge

- Does not incorporate task-specific information to generate samples.

LLM-based Methods

Uses domain knowledge

- Rely on LLM’s prior knowledge to explore the search space

Produces unstable candidates

- Produces candidates that theoretically possible but not synthesizable

- The absence of standardized multi-objective benchmarks in materials discovery hinders the systematic development and evaluation of novel baseline methods.

Using LLMs as Guided Evolutionary Optimizers

- Unlike generative methods, LLM's prior knowledge can help **focus on the relevant candidates** only and **generate meaningful materials for the given task.**

- However, existing LLM-based methods **treat material discovery as a single objective optimization task**.

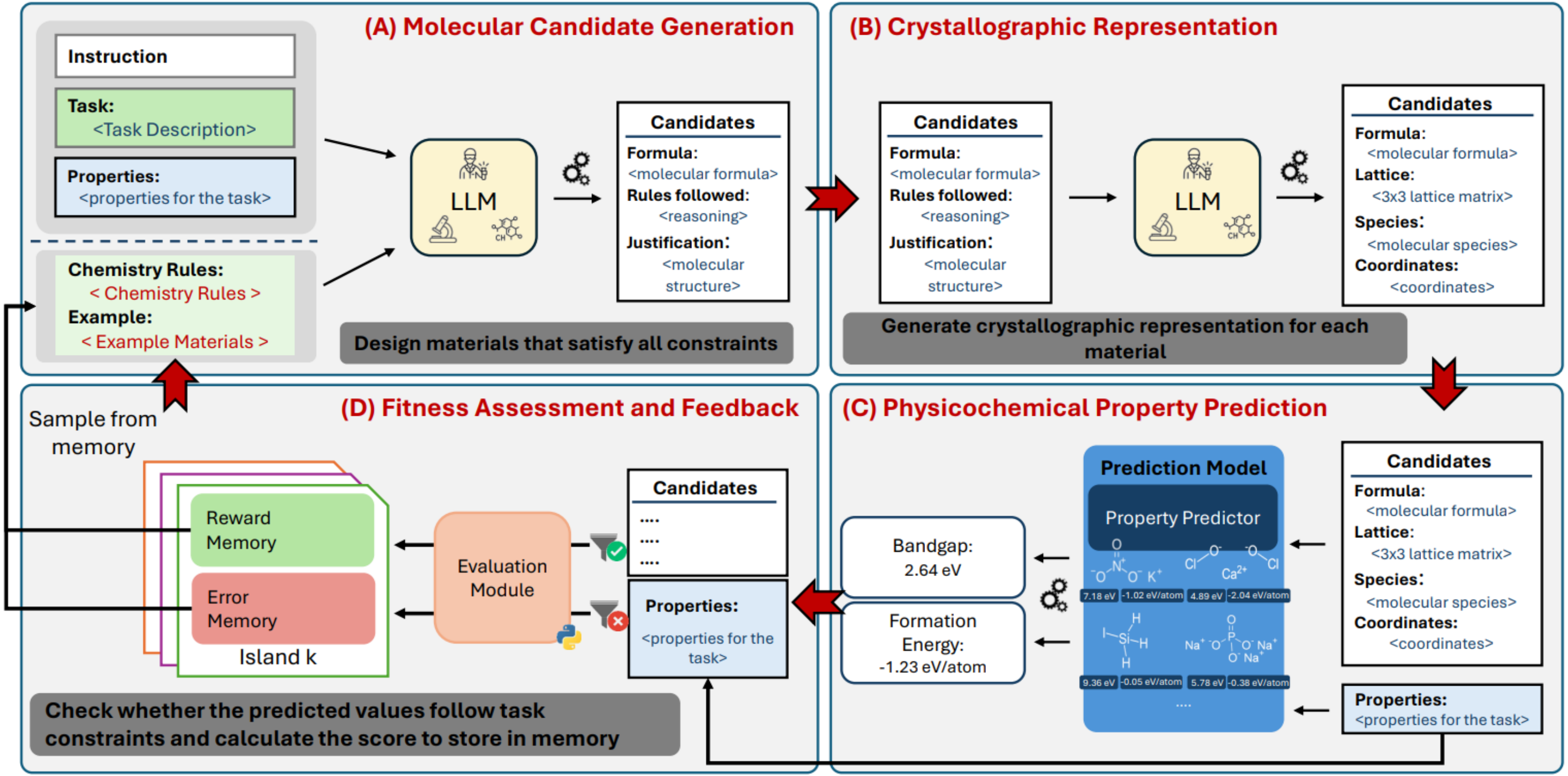
- Furthermore, they treat LLM as unguided exploratory optimizers producing theoretical compounds that are not synthesizable.

- **Question:** Can we efficiently steer the LLM toward generating valid as well as stable materials while leveraging its prior knowledge?

✓ We introduce **LLEMA**, a novel framework that **uses LLM's scientific knowledge with evolutionary search**, and **chemistry-informed design principles** to generate and refine candidates under multiple task-specific property constraints.

✓ We create **a benchmark suite comprising 14 industrially relevant material discovery tasks** designed to facilitate the development and evaluation of novel computational methodologies.

LLEMA: Evolutionary Search with LLMs for Multi-Objective Materials Discovery



(A) Molecular Candidate Generation

Instruction

Task:
<Task Description>

Properties:
<properties for the task>

Chemistry Rules:
< Chemistry Rules >

Example:
< Example Materials >

LLM

Candidates
Formula: <molecular formula>
Rules followed: <reasoning>
Justification: <molecular structure>

Design materials that satisfy all constraints

(B) Crystallographic Representation

Candidates
Formula: <molecular formula>
Rules followed: <reasoning>
Justification: <molecular structure>

LLM

Candidates
Formula: <molecular formula>
Lattice: <3x3 lattice matrix>
Species: <molecular species>
Coordinates: <coordinates>

Generate crystallographic representation for each material

(C) Physicochemical Property Prediction

Candidates
Formula: <molecular formula>
Lattice: <3x3 lattice matrix>
Species: <molecular species>
Coordinates: <coordinates>

Properties:
<properties for the task>

Prediction Model

Property Predictor

Bandgap:
2.64 eV

Formation Energy:
-1.23 eV/atom

(D) Fitness Assessment and Feedback

Candidates
....
....
....

Properties:
<properties for the task>

Evaluation Module

Reward Memory

Error Memory

Island k

Sample from memory

Check whether the predicted values follow task constraints and calculate the score to store in memory

LLEMA | Prompt Design

❶ Task Specification

Defines the role, provides task information and instructions to generate materials

❷ Demonstrations

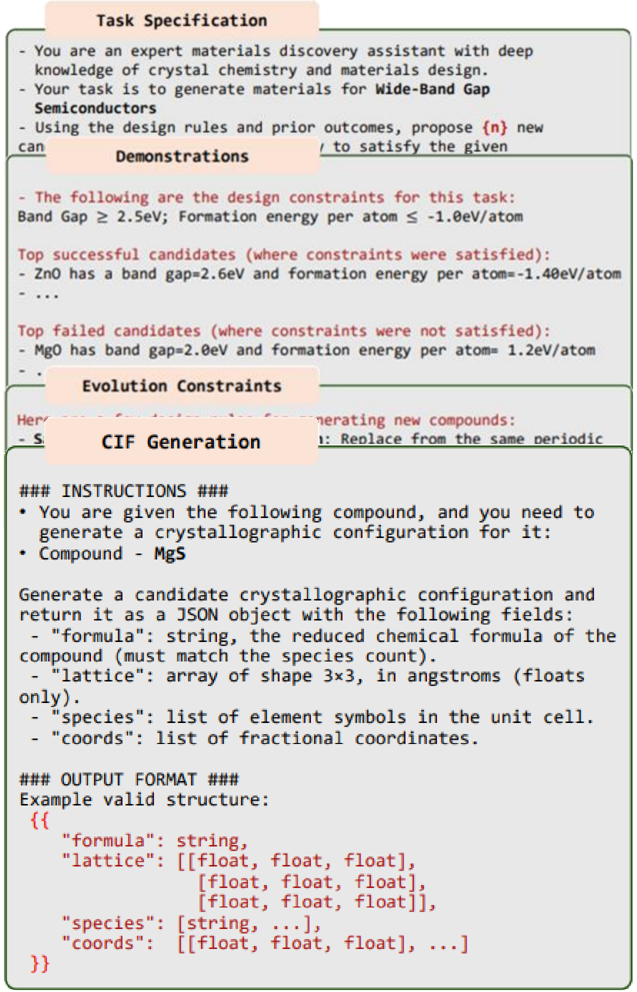
Design constraints specific to the task
Reward Memory – positive samples that follow all constraints
Error Memory – negative samples that violate constraints

❸ Chemistry-Informed Evolutionary Rules

Design rules to ensure generation of thermodynamically stable compounds

❹ CIF Generation

Generates a Crystallographic Information (CIF) File for every generated material



Task Specification

```
- You are an expert materials discovery assistant with deep
  knowledge of crystal chemistry and materials design.
- Your task is to generate materials for Wide-Band Gap
  Semiconductors
- Using the design rules and prior outcomes, propose {n} new
can                              to satisfy the given
```

Demonstrations

```
- The following are the design constraints for this task:
Band Gap ≥ 2.5eV; Formation energy per atom ≤ -1.0eV/atom

Top successful candidates (where constraints were satisfied):
- ZnO has a band gap=2.6eV and formation energy per atom=-1.40eV/atom
- ...

Top failed candidates (where constraints were not satisfied):
- MgO has band gap=2.0eV and formation energy per atom= 1.2eV/atom
- .
```

Evolution Constraints

```
Here                          nerating new compounds:
- S                         : Replace from the same periodic
```

CIF Generation

```
### INSTRUCTIONS ###
• You are given the following compound, and you need to
  generate a crystallographic configuration for it:
• Compound - MgS

Generate a candidate crystallographic configuration and
return it as a JSON object with the following fields:
 - "formula": string, the reduced chemical formula of the
compound (must match the species count).
 - "lattice": array of shape 3×3, in angstroms (floats
only).
 - "species": list of element symbols in the unit cell.
 - "coords": list of fractional coordinates.

### OUTPUT FORMAT ###
Example valid structure:
 {{
    "formula": string,
    "lattice": [[float, float, float],
                [float, float, float],
                [float, float, float]],
    "species": [string, ...],
    "coords":  [[float, float, float], ...]
 }}
```

LLEMA | Leveraging Domain Knowledge

- **Instruction:** Define the role and instruct the LLM to generate materials based on the task and its task-specific property constraints.
- **Task description:** Description of the task
  E.g., 'Wide-Bandgap Semiconductors'
- **Property Constraints:** Information of the constraints specific to the tasks.
  E.g., 'Band Gap > 2.5eV and Formation energy < -1.0 eV/atom
- **Data Samples:** Helps understand positive and negative samples to guide the generation process.
- **Chemistry-Informed Rules:** Guides the evolution toward chemically and physically possible generations using expert annotated chemistry rules.

Instruction:
You are a material science expert…

Task:
Your task is to generate materials for:
Wide-Bandgap Semiconductors

Property Constraints:
The following are the property constraints for the task:
Band gap < 2.5eV….
Top Successful Samples are:
<XY> has a bandgap 3.0eV….
Top Failure Samples are:
<AB> has a bandgap 1.5eV….

Evolutionary Rules:
1. **Same-group elemental substitution:**
   Replace each element with another from the same periodic group.
2. **Stoichiometry-preserving substitution:**
   Keep the formula ratios but replace with chemically similar elements.
3. **Oxidation state substitution:**
   Replace elements with others having the same oxidation state.

…
…

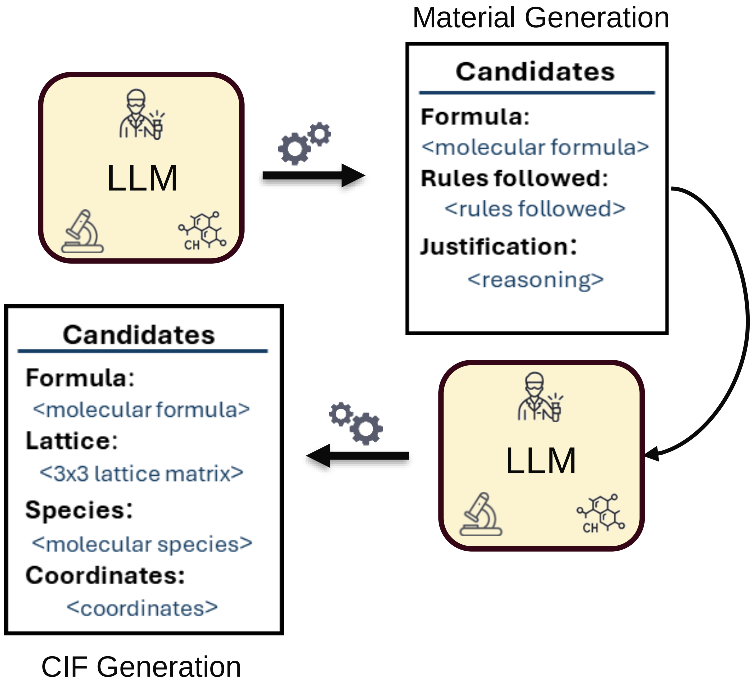
LLEMA | Material Generation

Use the LLM's **domain knowledge** and **chemistry rules** to produce constraint following task-specific CIF representation for materials.

Material generation includes:

✓ **Stage 1:**
- **Input:** Input Prompt with task-specific information
- **Output:**
  Molecular formula for the generated material
  The chemistry rule used to modify candidates
  Justification for the selected rule and modified material

✓ **Stage 2:**
- **Input:** Molecular formula for the generated material
- **Output:** Crystallographic Information File (CIF)



Material Generation

LLM

Candidates
Formula: <molecular formula>
Rules followed: <rules followed>
Justification: <reasoning>

LLM

Candidates
Formula: <molecular formula>
Lattice: <3x3 lattice matrix>
Species: <molecular species>
Coordinates: <coordinates>

CIF Generation

LLEMA | Physicochemical Prediction

(1) CIF Generation

- Generate the CIF representation for every generated material.

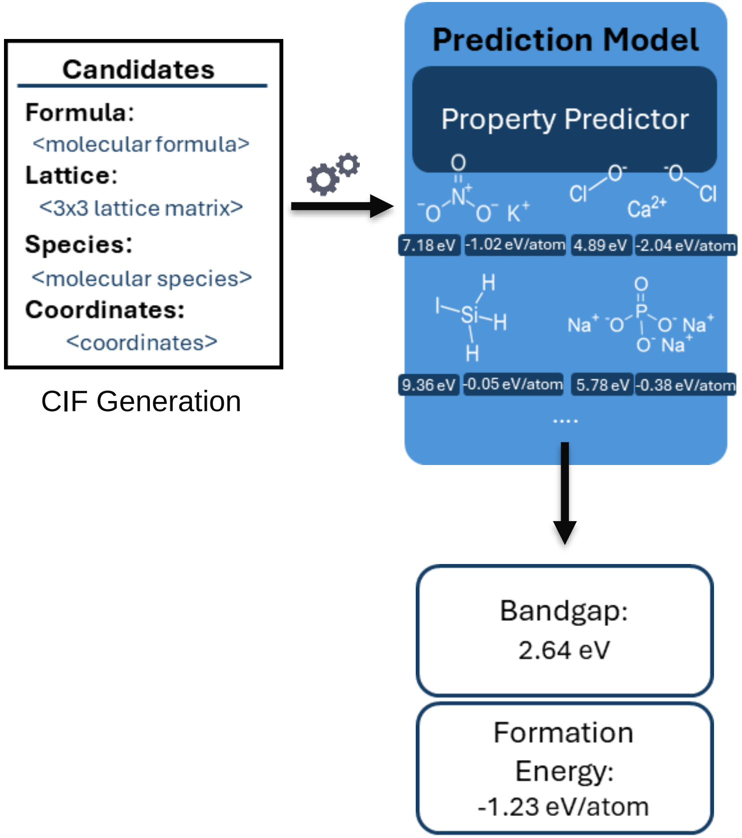
(2) Oracle Query

- Query the oracle to find the property values for the given CIF representation.

(3) Surrogate Prediction

- If the oracle does not contain the corresponding properties for the provided CIF, invoke surrogate models such as ALIGNN or CGCNN.
- We utilize property-specific pretrained files to predict the properties for the CIF representation.

Property	Model	Pretrained File
Bandgap	CGCNN	band-gap.pth.tar
Formation Energy	ALIGNN	jv_formation_energy_peratom
…	…	…



Candidates

Formula:
<molecular formula>
Lattice:
<3x3 lattice matrix>
Species:
<molecular species>
Coordinates:
<coordinates>

CIF Generation

Prediction Model

Property Predictor

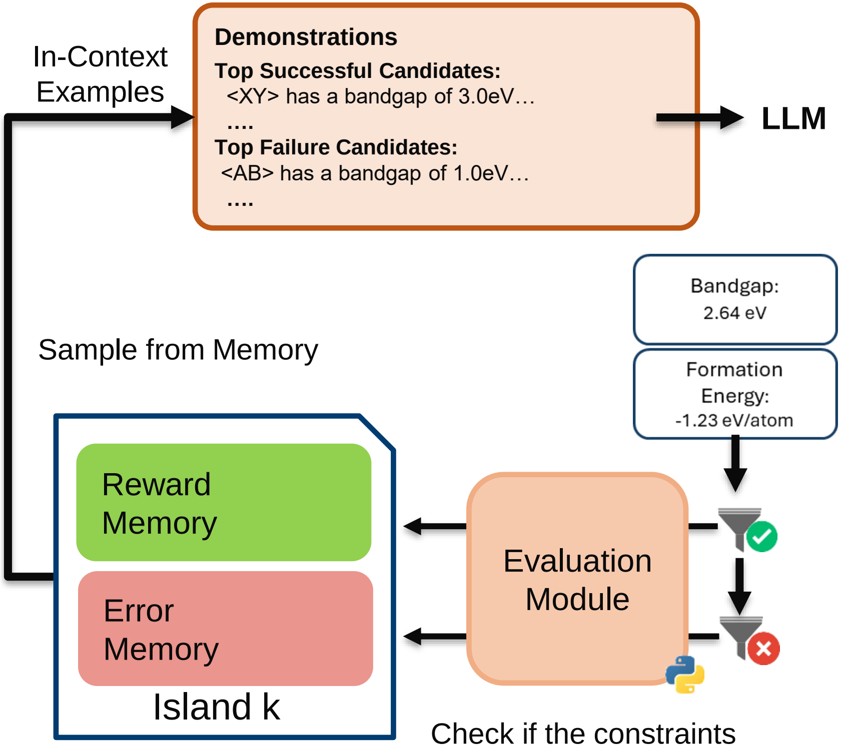
Bandgap:
2.64 eV

Formation Energy:
-1.23 eV/atom

LLEMA | Experience Management

To better navigate the search space, we store previously generated samples into **(i) Reward Memory** and **(ii) Error Memory**.

- **Store candidates in a memory buffer**: Multi-population memory (Islands) model to maintain diverse material generation.
- Reward and Error memory to store positive and negative samples to guide future generations.
- **Sample multiple examples** from both buffers within an island.
- Provide these candidates as **in-context examples** to LLM. LLM performs mutation and cross-over operations using chemistry-informed rules to generate new materials.



In-Context Examples

Demonstrations

Top Successful Candidates:
<XY> has a bandgap of 3.0eV…
….
Top Failure Candidates:
<AB> has a bandgap of 1.0eV…
….

LLM

Sample from Memory

Reward Memory

Error Memory

Island k

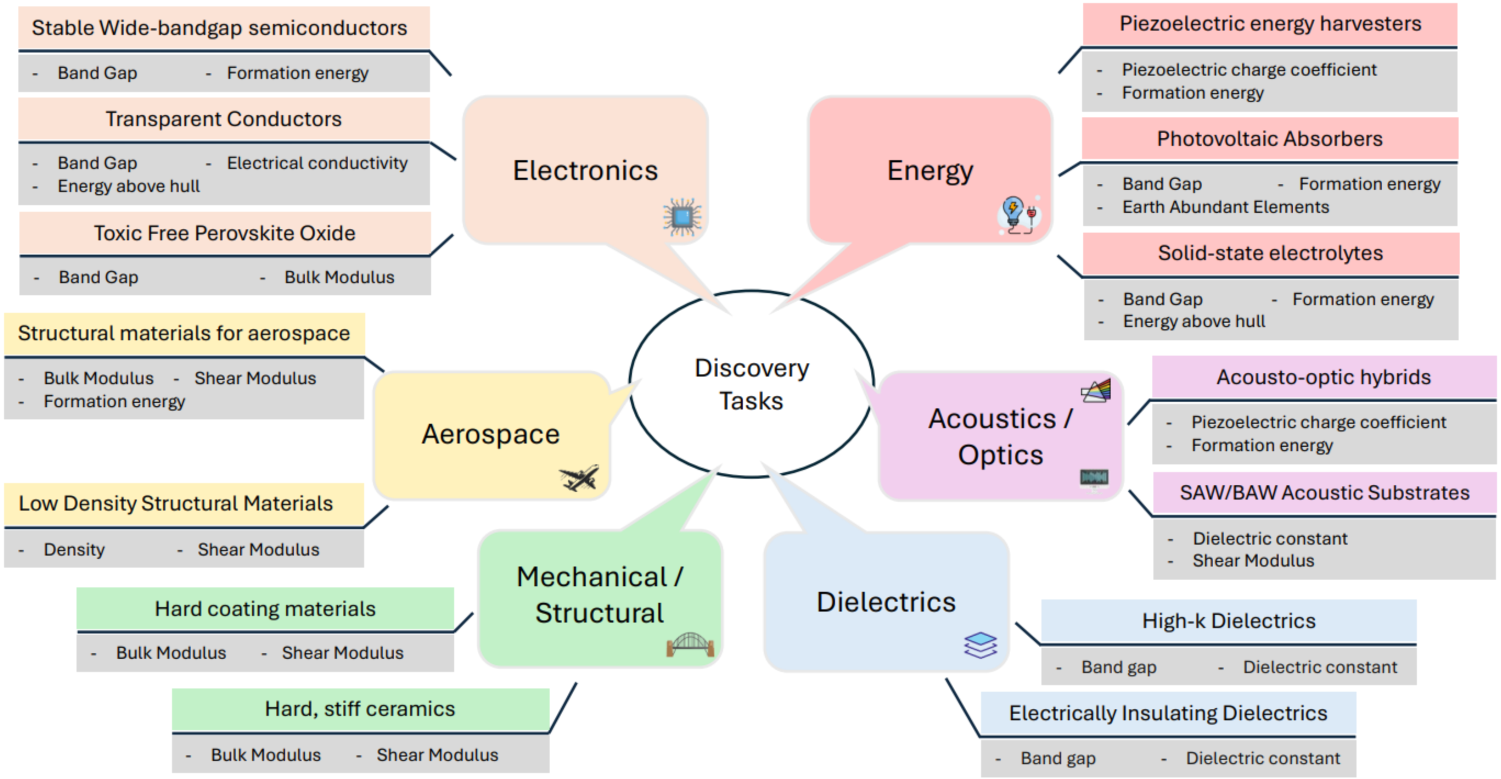
Evaluation Module

Bandgap: 2.64 eV

Formation Energy: -1.23 eV/atom

Check if the constraints are followed and score the candidates

LLEMA-Bench | Material Discovery Benchmark



Discovery Tasks

Electronics

Stable Wide-bandgap semiconductors
- Band Gap
- Formation energy

Transparent Conductors
- Band Gap
- Electrical conductivity
- Energy above hull

Toxic Free Perovskite Oxide
- Band Gap
- Bulk Modulus

Energy

Piezoelectric energy harvesters
- Piezoelectric charge coefficient
- Formation energy

Photovoltaic Absorbers
- Band Gap
- Formation energy
- Earth Abundant Elements

Solid-state electrolytes
- Band Gap
- Formation energy
- Energy above hull

Aerospace

Structural materials for aerospace
- Bulk Modulus
- Shear Modulus
- Formation energy

Low Density Structural Materials
- Density
- Shear Modulus

Acoustics / Optics

Acousto-optic hybrids
- Piezoelectric charge coefficient
- Formation energy

SAW/BAW Acoustic Substrates
- Dielectric constant
- Shear Modulus

Mechanical / Structural

Hard coating materials
- Bulk Modulus
- Shear Modulus

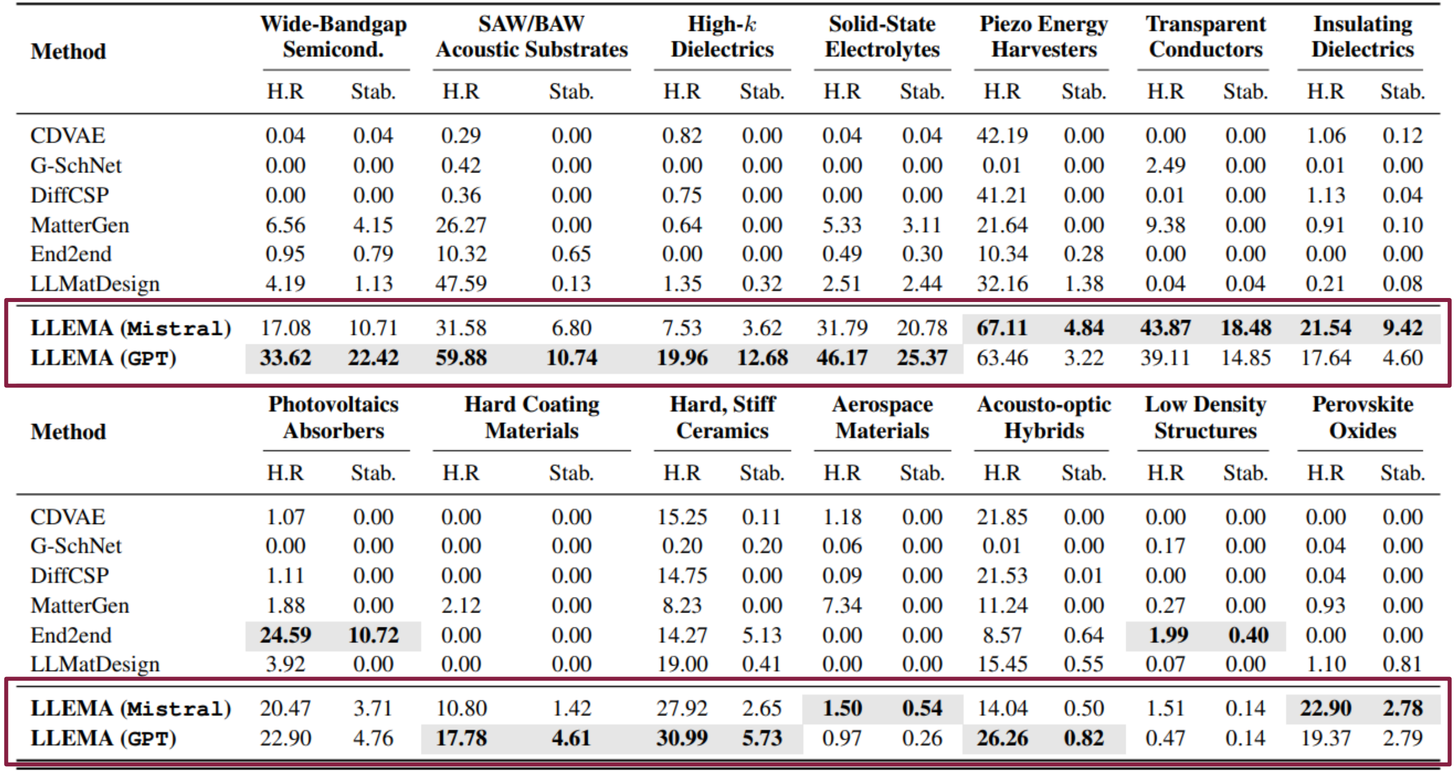
Hard, stiff ceramics
- Bulk Modulus
- Shear Modulus

Dielectrics

High-k Dielectrics
- Band gap
- Dielectric constant

Electrically Insulating Dielectrics
- Band gap
- Dielectric constant

LLEMA | Quantitative Results



Method	**Wide-Bandgap Semicond.**		**SAW/BAW Acoustic Substrates**		**High-k Dielectrics**		**Solid-State Electrolytes**		**Piezo Energy Harvesters**		**Transparent Conductors**		**Insulating Dielectrics**	
	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.
CDVAE	0.04	0.04	0.29	0.00	0.82	0.00	0.04	0.04	42.19	0.00	0.00	0.00	1.06	0.12
G-SchNet	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.01	0.00	2.49	0.00	0.01	0.00
DiffCSP	0.00	0.00	0.36	0.00	0.75	0.00	0.00	0.00	41.21	0.00	0.01	0.00	1.13	0.04
MatterGen	6.56	4.15	26.27	0.00	0.64	0.00	5.33	3.11	21.64	0.00	9.38	0.00	0.91	0.10
End2end	0.95	0.79	10.32	0.65	0.00	0.00	0.49	0.30	10.34	0.28	0.00	0.00	0.00	0.00
LLMatDesign	4.19	1.13	47.59	0.13	1.35	0.32	2.51	2.44	32.16	1.38	0.04	0.04	0.21	0.08
LLEMA (Mistral)	17.08	10.71	31.58	6.80	7.53	3.62	31.79	20.78	**67.11**	**4.84**	**43.87**	**18.48**	**21.54**	**9.42**
LLEMA (GPT)	**33.62**	**22.42**	**59.88**	**10.74**	**19.96**	**12.68**	**46.17**	**25.37**	63.46	3.22	39.11	14.85	17.64	4.60

Method	**Photovoltaics Absorbers**		**Hard Coating Materials**		**Hard, Stiff Ceramics**		**Aerospace Materials**		**Acousto-optic Hybrids**		**Low Density Structures**		**Perovskite Oxides**	
	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.	H.R	Stab.
CDVAE	1.07	0.00	0.00	0.00	15.25	0.11	1.18	0.00	21.85	0.00	0.00	0.00	0.00	0.00
G-SchNet	0.00	0.00	0.00	0.00	0.20	0.20	0.06	0.00	0.01	0.00	0.17	0.00	0.04	0.00
DiffCSP	1.11	0.00	0.00	0.00	14.75	0.00	0.09	0.00	21.53	0.01	0.00	0.00	0.04	0.00
MatterGen	1.88	0.00	2.12	0.00	8.23	0.00	7.34	0.00	11.24	0.00	0.27	0.00	0.93	0.00
End2end	**24.59**	**10.72**	0.00	0.00	14.27	5.13	0.00	0.00	8.57	0.64	**1.99**	**0.40**	0.00	0.00
LLMatDesign	3.92	0.00	0.00	0.00	19.00	0.41	0.00	0.00	15.45	0.55	0.07	0.00	1.10	0.81
LLEMA (Mistral)	20.47	3.71	10.80	1.42	27.92	2.65	**1.50**	**0.54**	14.04	0.50	1.51	0.14	**22.90**	**2.78**
LLEMA (GPT)	22.90	4.76	**17.78**	**4.61**	**30.99**	**5.73**	0.97	0.26	**26.26**	**0.82**	0.47	0.14	19.37	2.79

Table: Performance comparison of baselines on 14 multi-objective material discovery tasks

LLEMA | Qualitative Results

To assess the quality of the generated materials by LLEMA, we study its **(i) Convergence dynamics**, (ii) **Pareto front performance** and (iii) **Discovered candidates**.

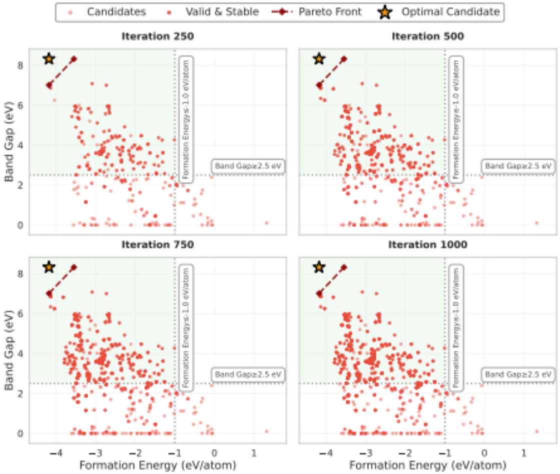


Figure: Evolution of candidates for the Stable Wide-Bandgap Semiconductor at different stages of LLEMA.

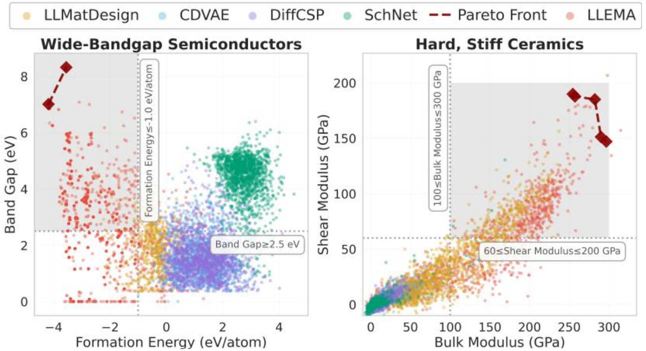


Figure: Pareto front analysis of candidate materials for two design tasks. (a) Wide-Bandgap Semiconductors; (b) Hard–Stiff Ceramics

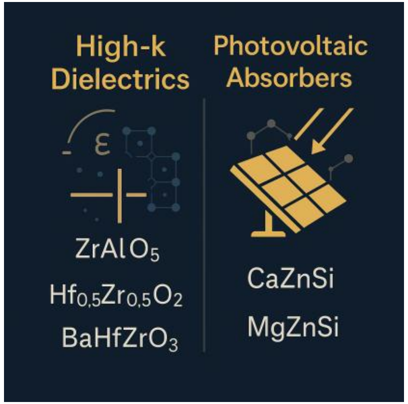
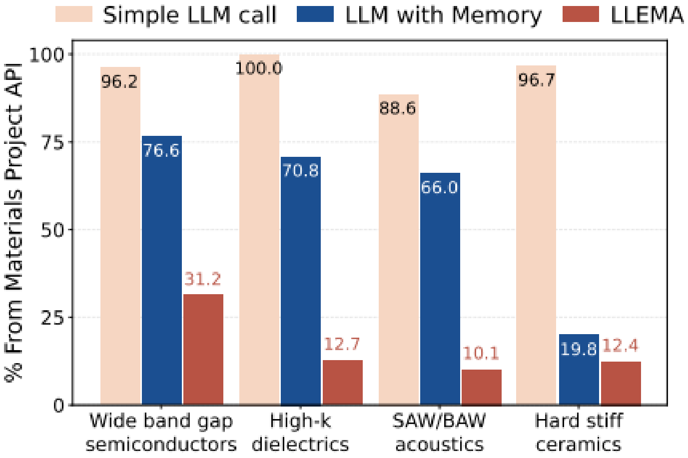
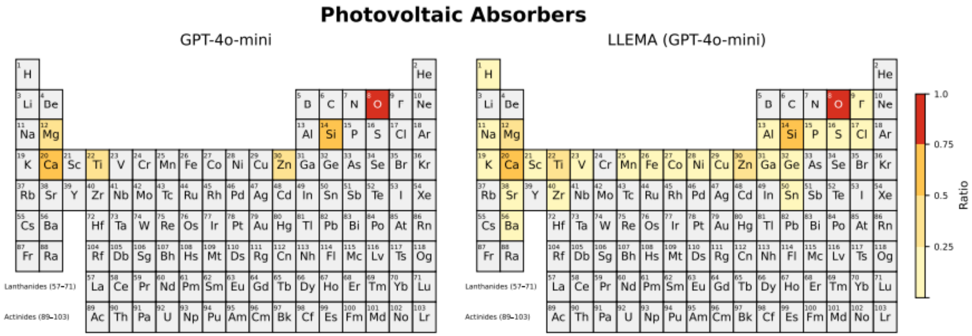


Figure: Generated materials for high-k dielectrics and photovoltaic absorbers.

LLEMA | Memorization Analysis

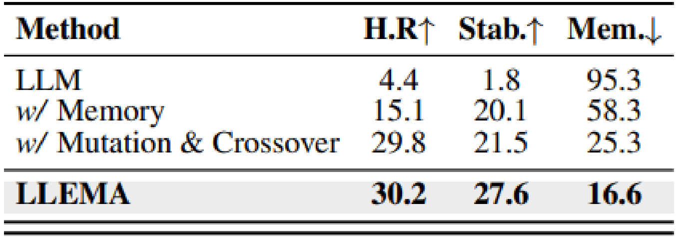




- **LLEMA mitigates memorization** by leveraging model-guided exploration
- It generates materials beyond previously known compositions.

- Unlike conventional LLM search confined to familiar chemistries, LLEMA effectively navigates the chemical space to **discover compositions with broader elemental diversity**.
- It explores multiple atomic groups, demonstrating its ability to generalize and identify unconventional yet promising materials.

LLEMA | Qualitative Analysis



Method	H.R↑	Stab.↑	Mem.↓
LLM	4.4	1.8	95.3
w/ Memory	15.1	20.1	58.3
w/ Mutation & Crossover	29.8	21.5	25.3
LLEMA	**30.2**	**27.6**	**16.6**

Table: : Comparison of hit-rate (H.R), stability (Stab.) and Memorization Rate (Mem.) across generation methods aggregated over four datasets.

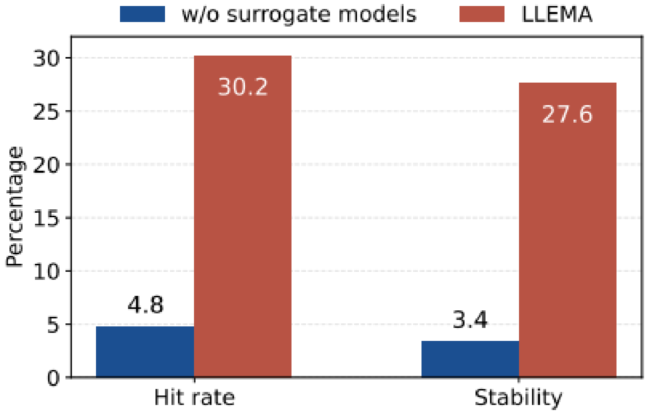


Figure: Hit Rate and Stability performance of LLEMA with and without surrogate model-based property prediction.

- LLEMA provides **higher hit rate and stability** with **lower memorization rates.**
- The integration of surrogate models helps improve the LLM feedback and results in a higher hit rate and stability. Relying only on the oracle leads to model collapse and ineffective exploration of the search space.

LLEMA | Case Study

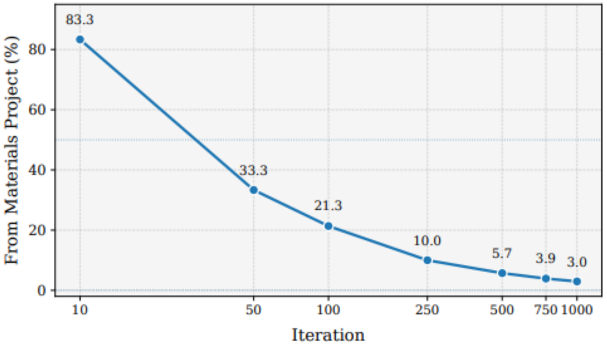


Figure: : Comparison of hit-rate (H.R), stability (Stab.) and Memorization Rate (Mem.) across generation methods aggregated over four datasets.

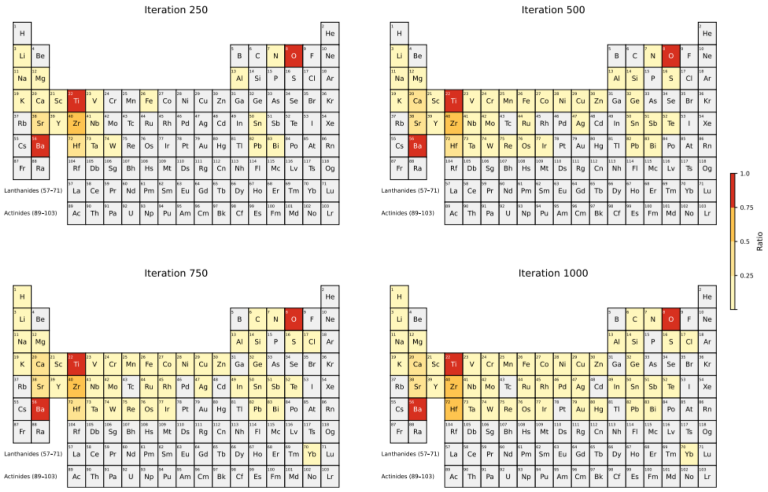


Figure: Evolution of periodic table coverage during evolution. Each panel shows the element-wise usage ratio across iterations (250, 500, 750, 1000) in the evolutionary search process.

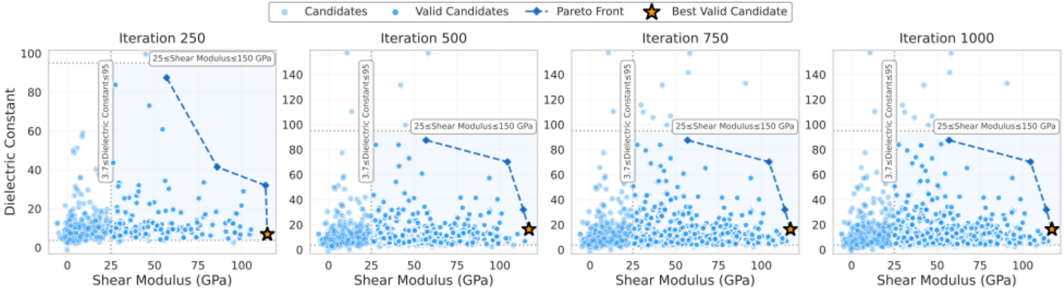


Figure: Evolution of the Pareto front during multi-objective optimization..

Summary

- LLEMA **incorporates domain knowledge and chemistry-guided rules** to generate materials that are both **novel** and **thermodynamically stable** .
- By **combining evolutionary search, data-driven optimization and chemistry-informed rules** in the discovery process, LLEMA effectively automates the discovery process.
- We **develop a benchmark of 14 industrially relevant multi-objective materials discovery tasks** across electronics, aerospace, energy and optics.
- **LLEMA achieves higher hit rates and stronger Pareto fronts** than generative and LLM-only baselines, showing improved ability to balance competing design objectives.

Future Directions

- Improving the property prediction component to provide a more robust prediction, reducing the reliance on surrogate predictors.
- Expanding the benchmark to cover more such multi-objective tasks across chemistry, biology and material science.

Thank You