From Generative Models to Real-World Materials: Entalpic's AI-Driven Discovery Loop

Alexandre Duval [1]

Introduction

Discovering new materials is critical to accelerating technological innovation and achieving ecological sustainability—from developing low-carbon catalysts to next-generation semiconductors and energy materials. Yet the discovery process remains painfully slow and inefficient, heavily reliant on human intuition, trial-and-error experimentation, and high-cost lab experiments or computational simulations like Density Functional Theory (DFT). These methods are not scalable enough to tackle the vastness of chemical space or the urgency of decarbonization.

At Entalpic, we are building a full stack AI-driven platform for autonomous materials discovery. Our approach integrates large-scale quantum-chemistry datasets, predictive and generative machine learning models, agentic LLMs and real-world experimental feedback into a tightly coupled active learning loop. This talk outlines our pipeline and its foundations, including key technical components and open-source initiatives like LeMaterial.

AI for Materials Discovery Loop

The chemical design space spans a quasi infinite number of potential materials, most of which have never been explored. Efficient discovery requires narrowing this space quickly while optimizing for multi-objective constraints (e.g., performance, stability, cost, availability and synthesizeability). A task that is inherently adapted to machines better than humans, which could partly or entirely automate the discovery pipeline, as they hold the ability to process all these factors while keeping in mind everything that has been tried. Overall, our proposed discovery pipeline integrates four tightly connected components:

- **Predictive Models.** We use geometric Graph Neural Networks (GNNs) such as *FAENet* (Duval et al., 2023b) and recent E(3)-equivariant architectures (Liao et al., 2023; Batatia et al., 2023) to learn structure–property relationships from atomic graphs. These models enable near-DFT accuracy at a fraction of the computational cost, and are detailed in our Hitchhiker's Guide to GNNs (Duval et al., 2023a).

- **Generative Flow Networks (GFlowNets).** Unlike diffusion models, GFlowNets generate materials step-by-step using a policy function taking actions, similarly to reinforcement learning. For e.g. selecting the crystal's space groups, its composition, Wyckoff sites, and lattice parameters, under hard-coded physical constraints. Our extension of *Crystal-GFN* (AI4Science et al., 2023) enables a diverse, controllable, interpretable exploration of crystal space, where candidate quality is guided by predictive GNN rewards (e.g., formation enthalpy or binding energies).

- **Active Learning.** To improve our ability to discover new (i.e. out-of-domain structures) structures, we augment the dataset in strategic regions of the chemical space using Density Functional Theory (DFT). And to avoid wasting computational resources on trivial or redundant candidates, models quantify uncertainty (via ensembles or embedding novelty) and prioritize new candidates for validation— either by DFT or experiments — thus improving both prediction accuracy and generative diversity over time.

- **Language Models (LLMs).** LLMs are used on scientific literature and patents to extract typical materials synthesis procedures or testing reaction conditions, helping (or suggesting) to the chemist (or robot) the route to follow to manufacture a candidate, i.e. augmenting human domain knowledge in the search loop. While full automatisation could be achieved in the long term, this is not an approach adopted by Entalpic.

Overall, this unified architecture forms a scalable, data-driven hypothesis engine that can propose, test, and refine material candidates autonomously. To achieve, a discovery, it requires two things that we will focus on: good data (to train good models), and a correlation between these simulations and experiments.

[1]Entalpic, France. Correspondence to: Alexandre Duval <alexandre.duval@entalpic.ai>.

LeMaterial: Open-Source Data, Benchmarks & Models

Machine Learning in materials science is bottlenecked not only by model quality, but by fragmented, redundant, and low-quality datasets. To address this, we launched the **LeMaterial** initiative, in collaboration with HuggingFace, to standardize and scale high-quality data access. This effort decomposes into several datasets being released, and model benchmarks about to.

- **LeMat-Bulk** (Siron et al.): Aggregates and deduplicates 5.3M structures from Materials Project, OQMD, and Alexandria using our custom **BAWL** fingerprint—combining bonding graphs, symmetry-aware Weisfeiler-Lehman hashing, and robustness to strain or noise.
- **LeMat-Traj**, **LeMat-Cat**, **LeMat-Synth** and **LeMat-Rho**: Extensions for dynamical data, catalytic surfaces, synthesis recipes and electron density.

These datasets support foundation model training, multi-fidelity learning, and enable rigorous benchmarking of downstream discovery applications—serving as a public backbone for the field.

Bridging the Gap with Experiments

Simulations and ML models provide fast approximations—but the ultimate test is experimental validation. At Entalpic, we tightly couple AI with high-throughput lab workflows to ensure that candidates are not just promising in theory, but also synthesizable and performant under realistic conditions.

- **Automated synthesis** enables parallel, reproducible fabrication of proposed materials using robotic platforms and custom workflows.
- **Characterization** (e.g., XRD, XRF, TEM) helps confirm the match between proposed and actual structures and reveals synthesis-dependent variations.
- **Performance testing** measures functional metrics such as catalytic activity, selectivity, and lifetime, in application-specific setups.

Experimental feedback is automatically reintegrated into our active learning loop—enhancing model robustness and aligning future predictions with real-world constraints.

Conclusion

Entalpic's platform falls into a new paradigm in materials science: one where AI does not just screen faster, but helps generate and validate novel compounds from first principles to deployment. By combining GNN-based property prediction, generative chemistry with GFlowNets, active learning, and automated experimentation, we move beyond trial-and-error toward a system that learns—continuously and intelligently—how to discover better materials.

References

AI4Science, M., Hernandez-Garcia, A., Duval, A., Volokhova, A., Bengio, Y., Sharma, D., Carrier, P. L., Benabed, Y., Koziarski, M., and Schmidt, V. Crystal-gfn: sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*, 2023.

Batatia, I., Benner, P., Chiang, Y., Elena, A. M., Kovács, D. P., Riebesell, J., Advincula, X. R., Asta, M., Baldwin, W. J., Bernstein, N., et al. A foundation model for atomistic materials chemistry. *arXiv preprint arXiv:2401.00096*, 2023.

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Lio, P., Bengio, Y., and Bronstein, M. A hitchhiker's guide to geometric gnns for 3d atomic systems. *arXiv preprint arXiv: 2312.07511*, 2023a.

Duval, A. A., Schmidt, V., Hernandez-Garcia, A., Miret, S., Malliaros, F. D., Bengio, Y., and Rolnick, D. Faenet: Frame averaging equivariant gnn for materials modeling. In *International Conference on Machine Learning*, pp. 9013–9033. PMLR, 2023b.

Liao, Y.-L., Wood, B., Das, A., and Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

Siron, M., Djafar, I., du Fayet, E., Rossello, A., Ramlaoui, A., and Duval, A. Lemat-bulk: aggregating, and deduplicating quantum chemistry materials databases. In *AI for Accelerated Materials Design-ICLR 2025*.