Lupine Science

2026-07-09

The 0.2% Synthesis Problem

https://lupine.science/articles/the-02-percent-synthesis-problem/

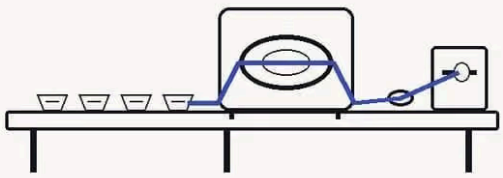
CLIMATE PARTNERSHIPS PROOF PACK

RESEARCH NOTE

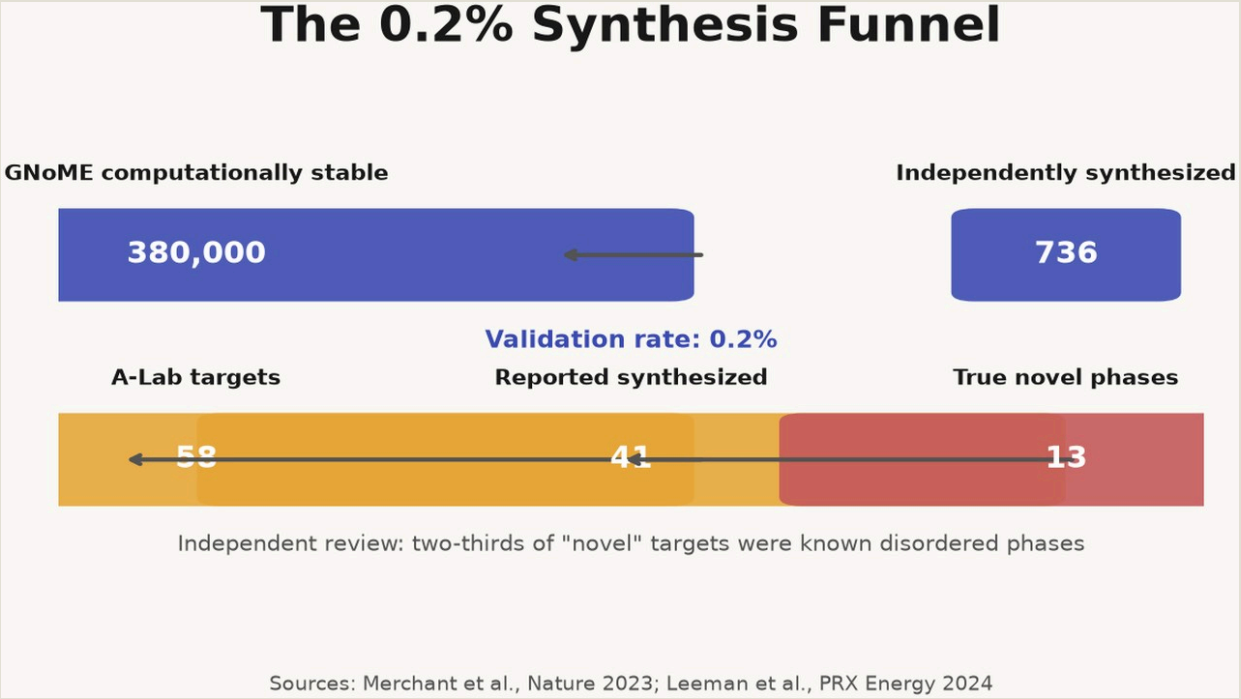
The 0.2% Synthesis Problem

Why most computationally predicted materials never reach a synthesis vessel, and why that matters for climate-critical materials.

July 9, 2026 · DRAFT



A long laboratory bench with many empty ceramic sample boats entering a small tube furnace and one finished pellet in a coin-cell clamp — the narrow furnace-to-test route makes the physical synthesis bottleneck visible.



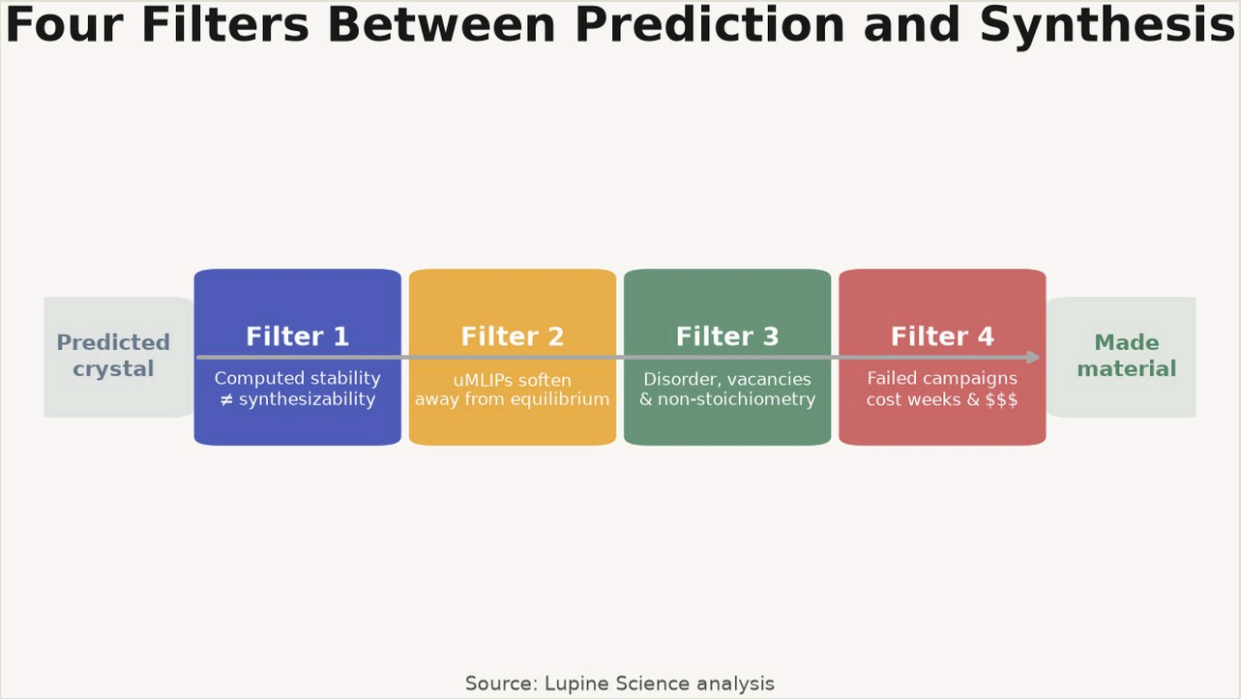
> Of 380,000 computationally stable structures reported by GNoME, only 736 had been independently synthesized by late 2023 — a 0.2% validation rate.

Computational materials science is producing candidates faster than laboratories can validate them. Google DeepMind's GNoME reported 380,000 computationally stable inorganic structures, yet only 736 had been independently synthesized by late 2023 — a 0.2% validation rate[1]. The A-Lab Author Correction (Nature 650:E1, 2026) records 36 confirmed of 57 eligible targets, with 4 inconclusive and one compound removed for training-data leakage; a separate independent review disputed the novelty classification of many targets[2].

The headline numbers are symptoms, not accidents. They point to a consistent set of bottlenecks between a predicted crystal and a made material. If those bottlenecks remain unaddressed, the climate cost will be measured in gigatonnes of CO_2 and years of missed deployment windows. Batteries alone are directly linked to roughly 20% of the CO_2 reductions required by 2030 and indirectly to another 40%[3]. Clean-energy hardware investment must rise from $1.8 trillion in 2023 to roughly $4.5 trillion per year by the early 2030s if net-

zero targets are to stay within reach[4]. Materials discovery is not moving too slowly in absolute terms; it is moving too slowly relative to the scale and urgency of the problem.

Lupine Science occupies the gap between prediction and synthesis. Closing that gap requires a correction and verification layer rather than a larger generator.



> Four structural filters separate a predicted crystal from a made material.

Why most predicted crystals disappear

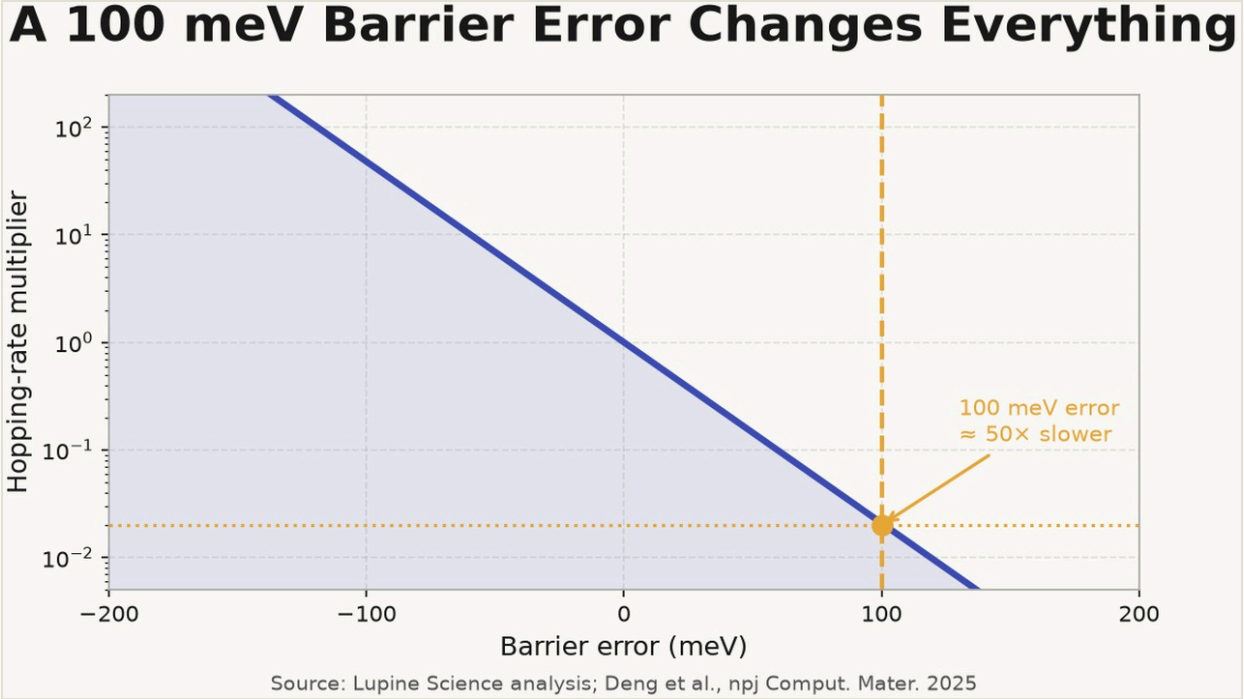
A recent study that tracked the fate of 736 predicted inorganic materials found that only 0.2% were made and structurally confirmed. The low number is not primarily a failure of synthesis skill. It is the result of four filters that structure generators and fast predictors routinely fail to apply.

Filter 1: Computed stability is not synthesizability

A structure can be on the convex hull of thermodynamic stability and still be unreachable by any practical chemical route. Most screens rank candidates by DFT energy or a machine-learned surrogate of it, but the real question is whether a precursor path exists that converges to the target rather than to a competing phase, an amorphous intermediate, or decomposition. Real synthesis routes routinely produce metastable phases that convex-hull-only screening eliminates prematurely.

Filter 2: Force-field accuracy away from equilibrium

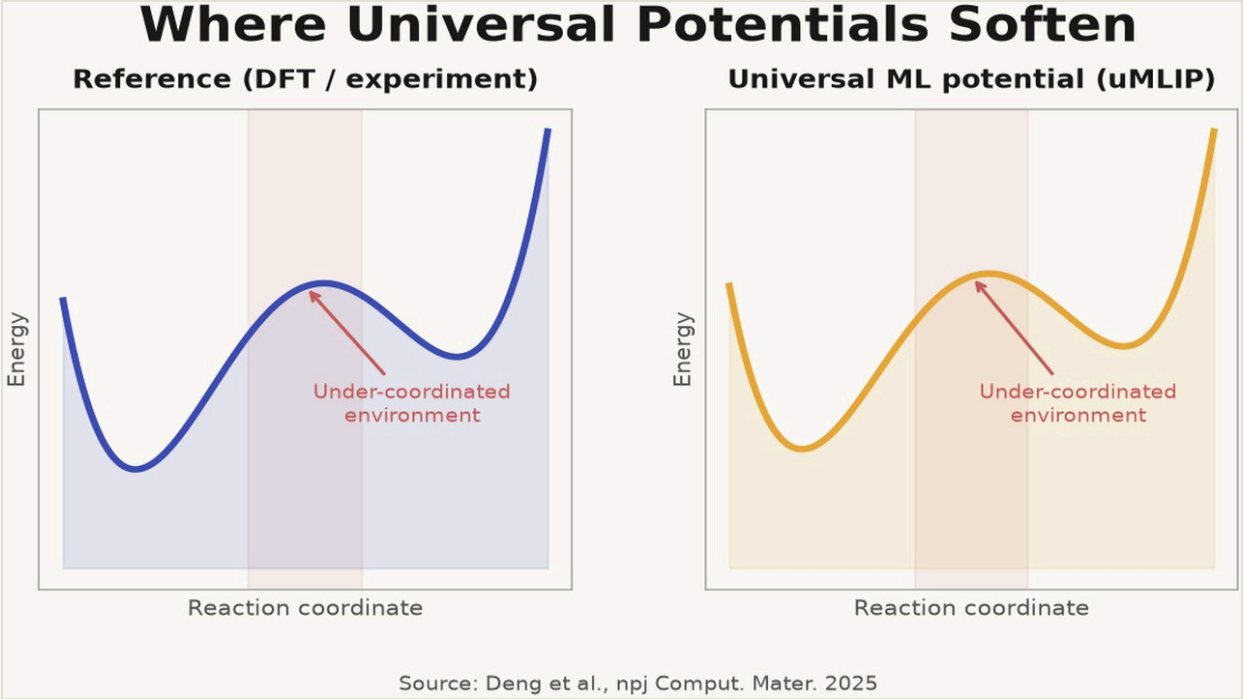
Most screening uses universal machine-learning interatomic potentials (uMLIPs) that have never seen the reactive precursors or the real furnace environment. These models are trained on near-equilibrium bulk configurations and systematically soften the potential energy surface in under-coordinated environments such as surfaces, vacancies, and transition states[5]. The consequence is defect and migration-barrier errors that can invert the ranking of candidate materials. A 100 meV barrier error changes the hopping rate by roughly 50× at room temperature, enough to misclassify a fast-ion conductor as an insulator.



> A 100 meV barrier error changes ionic hopping rates by roughly 50× at room temperature — enough to misclassify a conductor as an insulator.

Filter 3: Disorder and non-stoichiometry

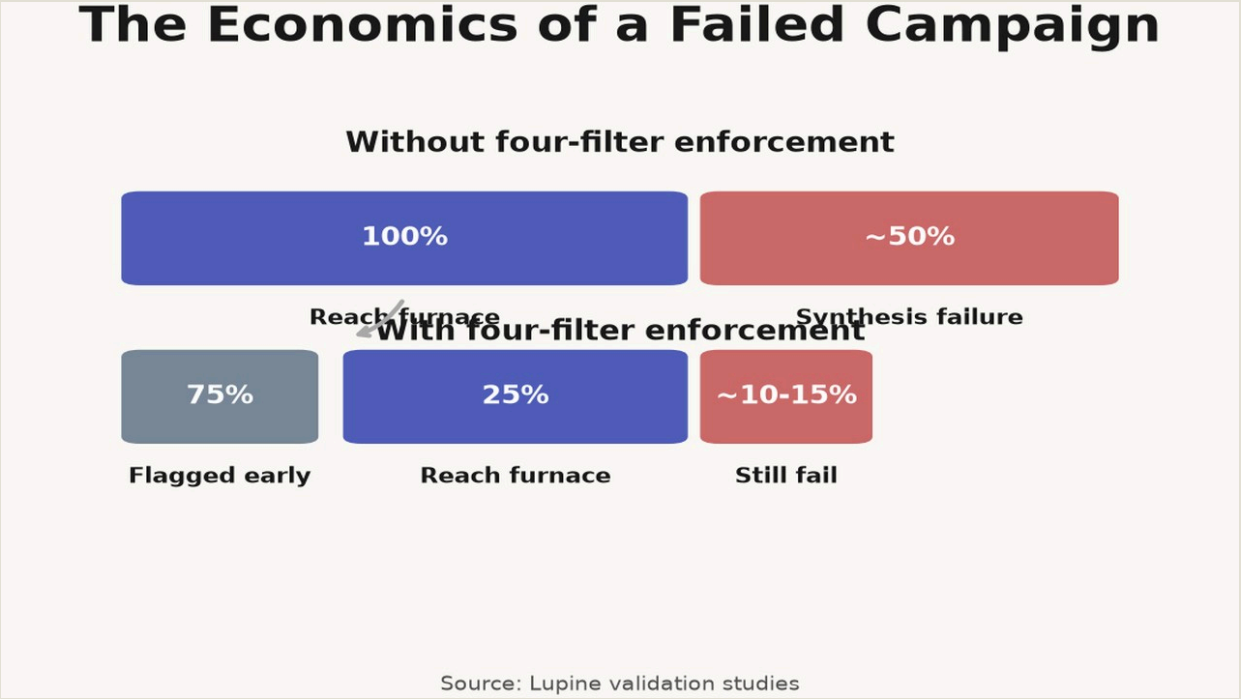
Many real materials are not perfectly ordered crystals. Cation disorder, oxygen vacancies, stacking faults, and amorphous surface layers dominate functional behavior, especially in battery cathodes and solid electrolytes. Structure generators usually emit ideal unit cells; predictors usually evaluate them as written. The resulting candidates fail when the synthesized powder is not the predicted crystal.



> Universal machine-learning interatomic potentials systematically soften the potential energy surface away from equilibrium.

Filter 4: The cost of a failed campaign

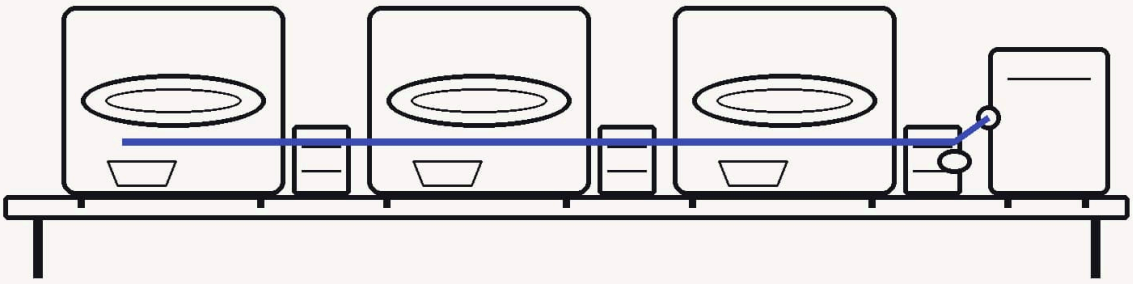
Each filter compounds the next. A candidate that passes an energy screen but fails in synthesis wastes weeks of lab time and thousands of dollars. In internal validation studies the failure rate is closer to 40–60% than 99.8%, but even that rate is too high for high-throughput screening to be economically rational. The estimate is that roughly three-quarters of all predictions would drop out before reaching a factory if these four filters were enforced globally.



> Enforcing the four filters would flag roughly three-quarters of predictions before they reach a furnace.

The climate math

The cost of leaving the bottleneck unsolved is not academic. Cobalt-free cathodes are the highest-leverage near-term target because cobalt supply is concentrated: the Democratic Republic of Congo produces roughly 70% of global cobalt[6]. Any battery chemistry that removes cobalt while preserving energy density and cycle life removes a geopolitical and ethical constraint on electrification.



> Cobalt-free cathodes remove a supply-chain chokepoint, while clean-energy investment must more than double this decade.

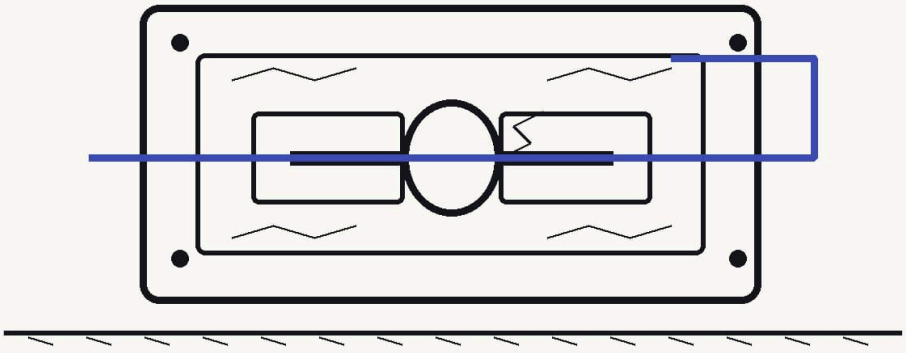
More broadly, the International Energy Agency estimates that the clean-energy technologies required for net-zero emissions must abate tens of gigatonnes of CO_2 by 2050[7]. Internal climate analysis suggests that the five material areas under active pursuit could, if fully deployed, abate 5–12 Gt CO_2 yr^{-1}—roughly 10–25% of what the IEA says clean-energy hardware must remove by 2050.

The window is narrow. A material discovered in 2035 can still shape 2040 deployment; one discovered in 2045 cannot. The difference is measured in cumulative emissions that no later innovation can recover.

> A material discovered too late cannot recover cumulative emissions already locked in.

What correction looks like

Closing the gap requires moving from “stable on paper” to “synthesizable in practice.” The approach is to learn an error field around each environment rather than to train a bigger model. The field measures how a uMLIP deviates from reference data as a function of local atomic coordination. That correction is applied at runtime with analytic forces, so molecular dynamics and structure relaxations follow proper gradients. The result is a ranked list in which the top candidates are far more likely to survive synthesis, and in which candidates that cannot be rescued are flagged before a furnace is turned on.



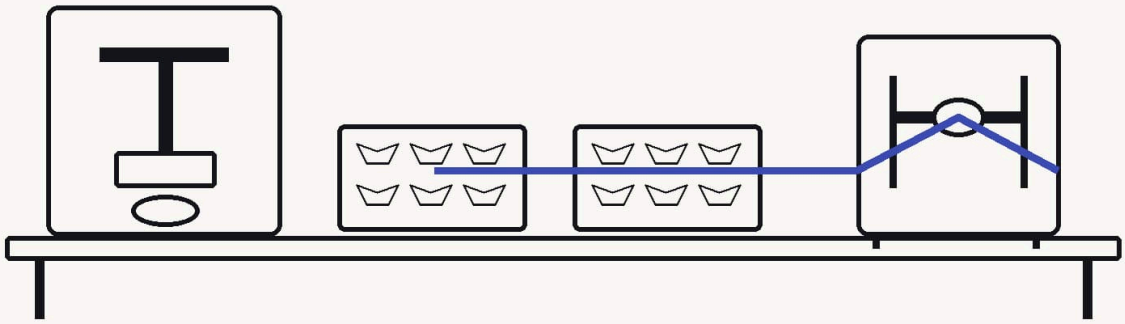
> A local environment error field is applied at runtime so molecular dynamics follows corrected gradients.

The approach is already being validated by the market. POSCO Future M has completed development of lithium-manganese-rich cathode materials and is preparing mass production in 2025[8]. General Motors and LG Energy Solution aim to begin commercial production of prismatic lithium-manganese-rich cells by 2028[9]. Solid Power is supplying BMW with automotive-scale solid-state cells for qualification testing[10]. Factorial Energy has delivered 100+ Ah quasi-solid-state cells to Mercedes-Benz, which is now road-testing a modified EQS sedan with the technology[11]. These are not laboratory curiosities; they are the endpoints that a corrected screening pipeline must feed.

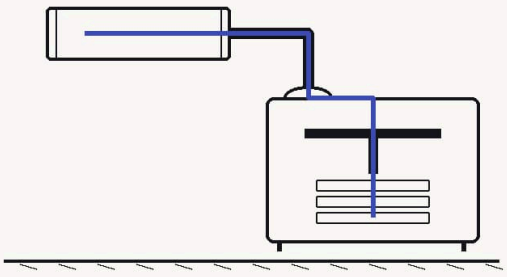
> Corrected screening must feed partners who are already scaling cobalt-free cathodes and solid-state cells.

From predictions to partners

Predictions are necessary but not sufficient. The path from a corrected energy landscape to a commercial material runs through named experimental collaborators who can synthesize, characterize, and scale the top candidates. The next article in this series turns from the diagnosis to the method: how the environment error field is measured, why the field is measured rather than learned, and how machine-checked proof prevents false positives from propagating through an autonomous pipeline.



> Corrected predictions must flow into synthesis, characterization, and scale partners — not into another generation cycle.



FOOTNOTES

1. A. Merchant *et al.*, “Scaling deep learning for materials discovery,” *Nature* 624, 80–85 (2023). https://doi.org/10.1038/s41586-023-06735-9 ↩

2. N. J. Szymanski *et al.*, “An autonomous laboratory for the accelerated synthesis of novel materials,” *Nature* 624, 86–91 (2023); J. Leeman *et al.*, “Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis,” *PRX Energy* 3, 011002 (2024). https://doi.org/10.1038/s41586-023-06734-7; https://doi.org/10.1103/PRXEnergy.3.011002 ↩

3. International Energy Agency, *Batteries and Secure Energy Transitions*, IEA, 2024. ↩

4. International Energy Agency, *World Energy Investment 2024*, IEA, 2024. ↩

5. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025). https://doi.org/10.1038/s41524-024-01500-6 ↩

6. International Energy Agency, *The Role of Critical Minerals in Clean Energy Transitions*, IEA, 2022. ↩

7. International Energy Agency, *Net Zero by 2050: A Roadmap for the Global Energy Sector*, IEA, 2021. ↩

8. POSCO Future M, “POSCO Future M to lead entry-level and standard EV markets with LMR cathode materials,” press release, June 2025. ↩

9. General Motors, “Why LMR batteries will change the outlook for the EV market,” GM Newsroom, May 2025; LG Energy Solution / GM, Battery Innovation of the Year, The Battery Show North America, October 2025. ↩

10. BMW Group / Solid Power, joint development agreement and Series B investment announcement, 2021; BMW i7 ASSB demo-vehicle road testing reported 2025. ↩

11. Factorial Energy / Mercedes-Benz, solid-state battery cell development and EQS road-test announcements, 2023–2025. ↩

Lupine Science

2026-07-09

A Field, Not a Neural Net

https://lupine.science/articles/a-field-not-a-neural-net/

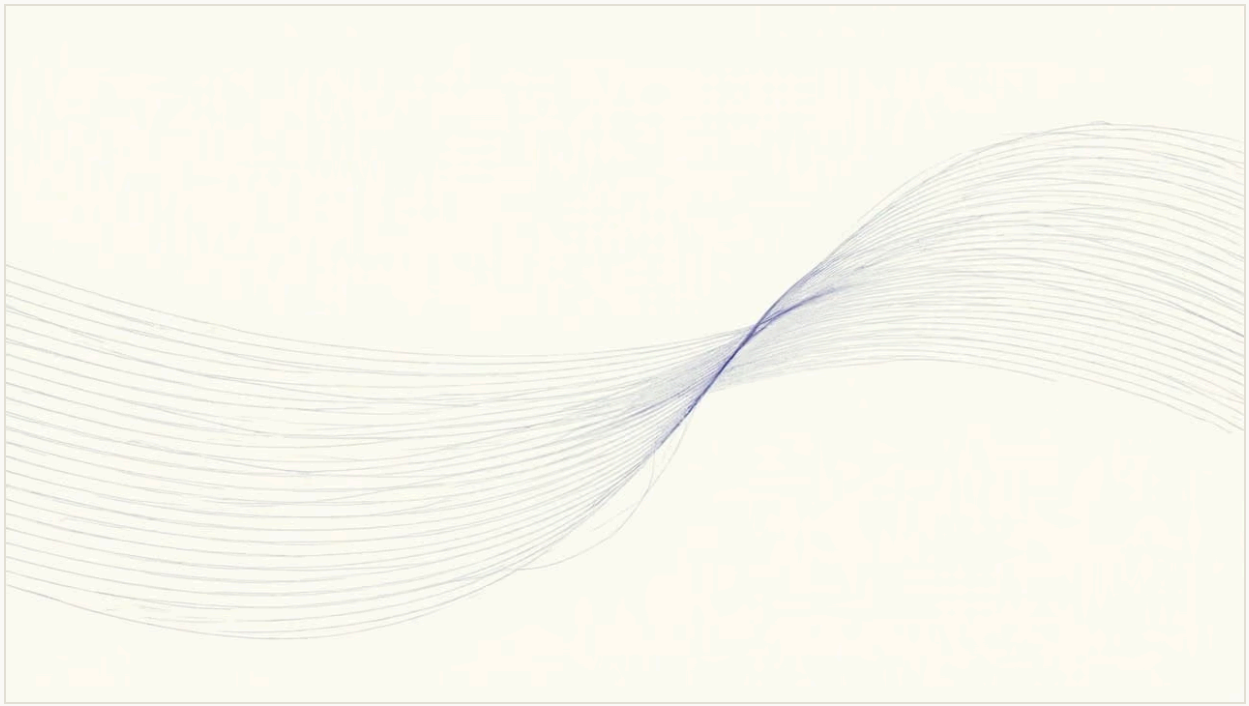
CLIMATE PARTNERSHIPS PROOF PACK

RESEARCH NOTE

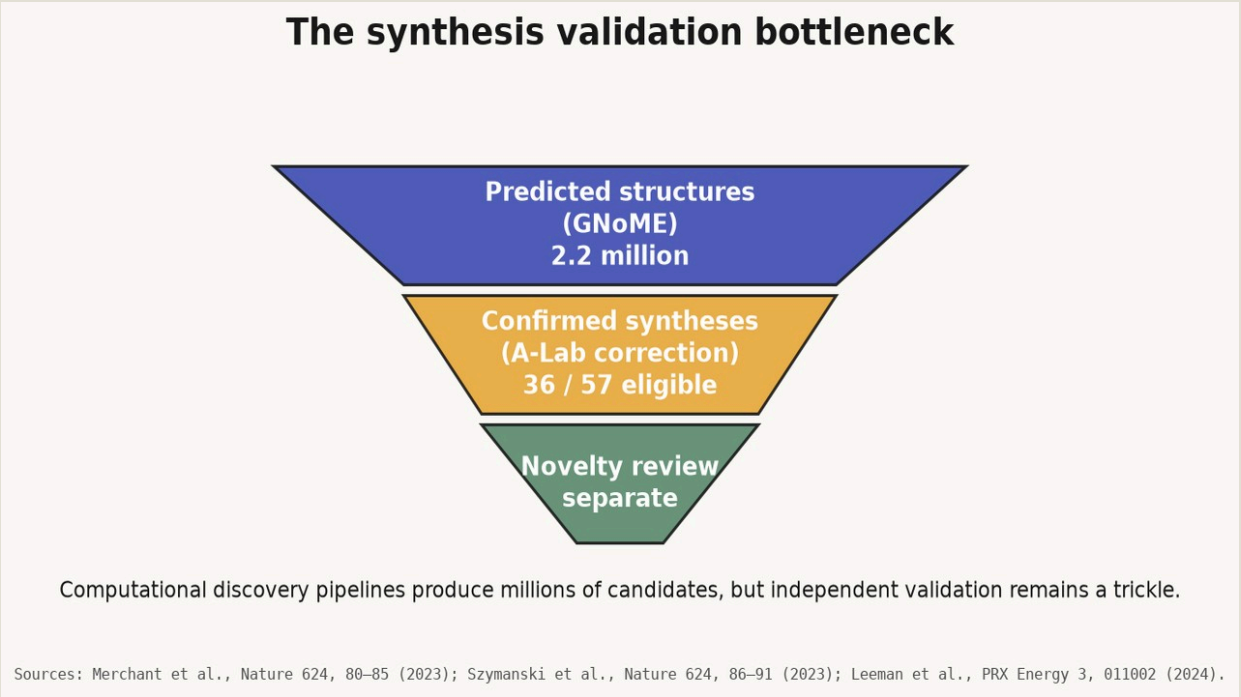
A Field, Not a Neural Net

The environment error field as a measured physical correction layer for universal machine-learning interatomic potentials.

July 9, 2026 · DRAFT



A smooth field gradient over a sparse lattice: the measured error geometry of universal machine-learned interatomic potentials.



> Predicted materials vastly outrun verified discoveries: GNoME's 2.2 million crystals produced only 736 confirmed syntheses, while A-Lab's corrected record confirms 36 of 57 eligible targets. Sources: Merchant et al., Nature 624, 80–85 (2023); Szymanski et al., Nature 624, 86–91 (2023); Author Correction, Nature 650:E1 (2026); Leeman et al., PRX Energy 3, 011002 (2024).

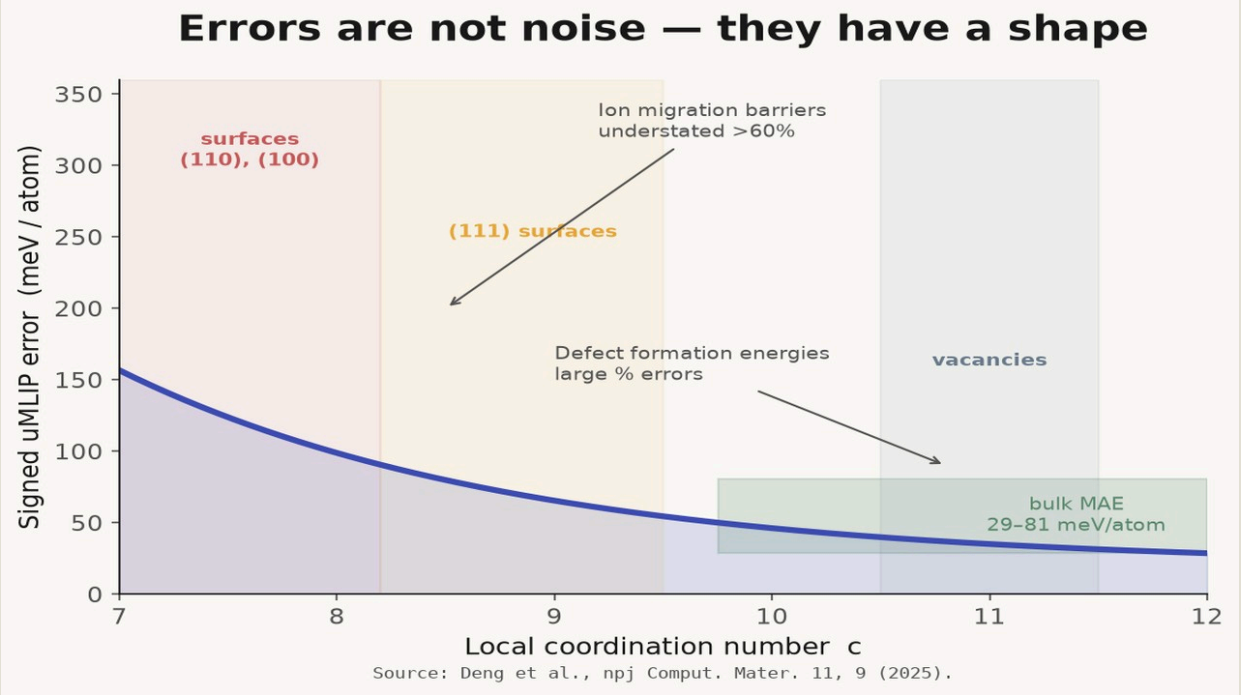
The dominant narrative in computational materials science is that bigger models trained on more data will close the gap between prediction and synthesis. The evidence points elsewhere. Google DeepMind's GNoME predicted 2.2 million crystals; only 736 had been independently synthesized by late 2023, a 0.2% validation rate[1]. The A-Lab Author Correction (Nature 650:E1, 2026) records 36 confirmed of 57 eligible targets, with 4 inconclusive and one compound removed for training-data leakage; a separate critique disputed the novelty classification of many targets[2]. The problem is not that the models are too small. It is that their errors have a shape — a systematic, low-dimensional geometry in the space of local atomic environments — and the field has been treating that shape as noise.

The present approach starts from the opposite premise. The wrongness of universal machine-learning interatomic potentials (uMLIPs) is not random. It is a smooth function of coordination number, measurable from a handful of anchor observables, and correctable at runtime with analytic forces. The environment error field is the focus: what it is, how it is measured rather than learned, and why the distinction matters for climate-critical materials discovery.

The Field Hypothesis

A uMLIP such as CHGNet, MACE-MP-0, or M3GNet evaluates atomic energies in milliseconds, enabling screening campaigns that would be economically impossible with density functional theory (DFT). On bulk formation energies the best models reach mean absolute errors of 29–81 meV/atom, close to the intrinsic accuracy of the DFT functionals on which they were trained[3]. But functional materials do not operate as bulk crystals. Ion conduction, catalytic turnover, and hydrolytic stability all depend on under-coordinated environments — vacancies, surfaces, and transition states — where uMLIPs fail most severely: defect formation energies carry large percentage errors, and ion migration barriers are underestimated by more than 60%[4].

The error is not an inescapable property of machine learning. It is a predictable consequence of training data. Databases such as the Materials Project trajectories, OMat24, and Alexandria are dominated by near-equilibrium, closed-shell bulk structures in which every atom sits close to its high-symmetry coordination. A smooth regressor trained on that distribution interpolates accurately at the bulk coordination but extrapolates with correlated bias into under-coordinated regimes: surfaces, vacancies, transition states, and grain boundaries[5].



> Where atoms are under-coordinated — surfaces, vacancies, and transition states — universal potentials drift predictably away from reference energies, not randomly. Source: Deng et al., npj Comput. Mater. 11, 9 (2025).

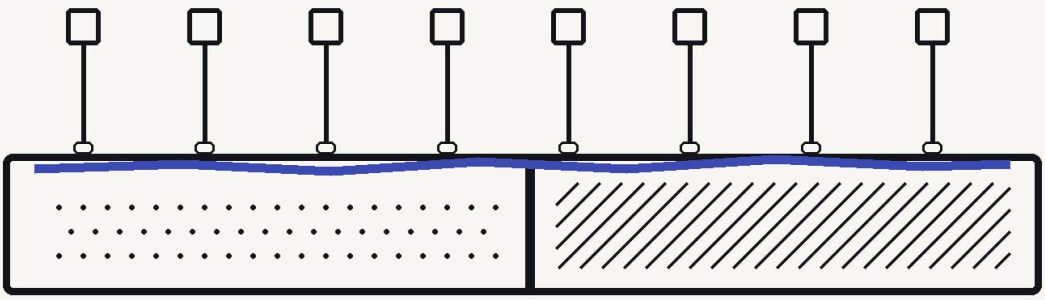
This observation is formalized as the environment error field. For a given configuration, the total systematic energy error of a uMLIP is approximated by a sum over atoms of a per-atom error that depends only on the atom's local coordination number c:

$$E^{\text{model}} - E^{\text{ref}} \approx \sum_i P(c_i).$$

For a perfect face-centered-cubic (fcc) bulk atom, $c = 12$, and the field is defined to be zero — consistent with the empirical fact that uMLIPs are already accurate in the bulk. The field $P(c)$ is therefore not a universal constant fitted across all materials; it is a material-specific and model-specific measured quantity, anchored to the physics of bond counting.

Measured, Not Learned

The critical difference between this field and previous correction strategies is that it is measured, not learned. Delta-ML trains a separate model to learn the discrepancy between a fast baseline and an accurate target; it requires per-system retraining and abundant reference data[6]. Fine-tuning updates uMLIP weights on target-system data; it works when curated training sets exist, which they do not for unexplored composition spaces[7]. Both approaches treat the error as an arbitrary function to be discovered from data.



The method imposes the functional form. A cubic spline through three standard anchor observables — the (100) surface probing $c = 8$, the (111) surface probing $c = 9$, and the vacancy formation energy probing $c = 11$ — is fixed by the boundary condition $P(12) = 0$. Linear continuation below $c = 8$ predicts the error at $c = 7$, which corresponds to the (110) surface energy. That observable is never fitted during field construction; it is predicted blind.

> The error field is measured from three standard observables and fixed to zero in the bulk; everything below the lowest anchor is a true blind prediction.

Each anchor samples a different coordination deficit relative to the bulk reference. Because the same coordination numbers recur across materials that share a crystal structure family, the spline transfers across properties within that family — surface energies, vacancy formation energies, and migration barriers — without additional DFT oracle calls. The correction is measured in the sense that geometry dictates the functional form and only the three knot values are empirical.

The r = 0.906 Result

The empirical test is strict blind prediction. Across 36 independent (model, material) combinations, the field predicts the signed error of the never-fitted (110) surface energy with Pearson correlation $r = 0.906$ ($p = 10^{-4}$, 95% confidence interval [0.82, 0.96]), using zero adjustable parameters.

> Across 36 independent combinations, the measured field predicts blind (110) surface-energy errors with r = 0.906, surviving a structurally aware permutation null.

The per-model breakdown is informative. CHGNet v0.4.2 achieves $r = 0.86$ ($p < 0.01$). MACE-MP-0 small achieves $r = 0.90$ ($p < 0.01$). MACE-MPA-0, trained on the larger OMat24 non-equilibrium dataset, achieves $r = 0.96$ ($p < 0.001$). MACE-MP-0 medium is the one non-significant case at $r = 0.47$ ($p = 0.10$). That exception matters: it bounds the field's domain rather than refuting it. The OMat24-trained model performs best because its training distribution

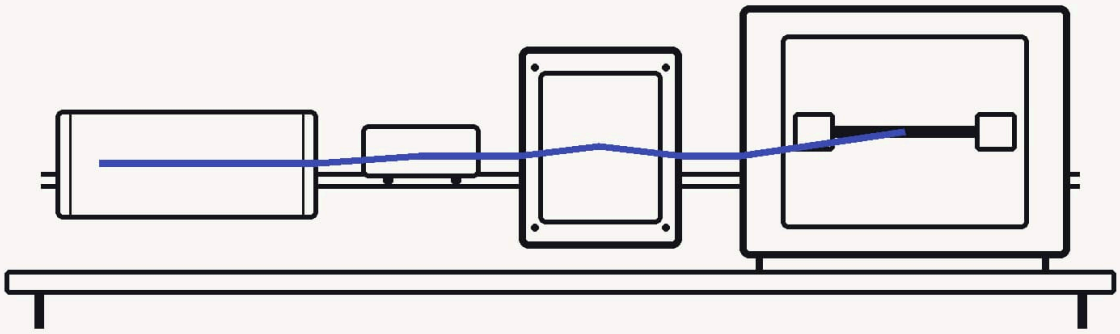
already samples more under-coordinated configurations, supporting the hypothesis that the field captures a real training-data bias rather than an accidental correlation.

An honest null is essential. Because (100), (111), and (110) facets of the same material are structurally correlated, a naïve permutation null would be too permissive. The reported null distribution, with mean $r = 0.44$, is generated by permuting within models across 10,000 draws, preserving the structural correlation. The measured $r = 0.906$ survives this conservative test.

At integer-scaled precision of 10^{-4} J/m^2, 26 of 36 cells improve strictly over the predict-zero baseline, with median residual dropping from 0.104 to 0.066 J/m^2. The number is 26, not 27 — a distinction that becomes important when the same pipeline is subjected to machine-checked proof.

Runtime Correction

The field inverts naturally into an additive correction energy:



$$E_{\mathrm{corr}} = -\sum_i P(c_i).$$

Because the field is negative where the model underpredicts — the typical case for defects and surfaces — the correction raises under-coordinated configurations toward the reference energy. Forces follow analytically from the chain rule, $F_{\mathrm{corr}} = -\nabla E_{\mathrm{corr}} = \sum_i P'(c_i)\nabla c_i$, so molecular dynamics conserves energy and structure relaxations follow proper gradients. Force accuracy was verified against numerical differentiation to 10^{-6} eV/Å on rattled slabs.

> The correction layer sits beside any existing uMLIP, adds analytic forces, and keeps molecular dynamics conservative while adding only single-digit overhead.

The correction layer deploys beside any existing uMLIP calculator without retraining its weights. Coordination computation is a local neighbor-list operation, so the overhead is small: 15.6% on CHGNet steps in the current Python implementation. A compiled pair-style overlay in LAMMPS would reduce that below 1%, because coordination reduces to a few floating-point operations per neighbor pair. Even with the present overhead, a corrected uMLIP remains many orders of magnitude faster than DFT.

Validation confirms selective correction. The fitted observables recover exactly in closure tests, demonstrating that the bond-counting field survives real structural relaxation. The blind (110) facet improves 6.5× for Ni (9.7% error to 1.5%) and 2.0× for Cu (28.0% to 13.7%). Bulk lattice constants are unchanged, because the correction vanishes identically at $c = 12$.

> Selective correction cuts surface-energy errors several-fold while leaving bulk lattice constants untouched, preserving uMLIP speed at near-DFT defect accuracy.

Machine-Checked Proof

Statistical validation has a well-known limitation: it cannot catch bugs in the analysis code, and it applies only to the samples tested. A formal verification layer is added. Every quantitative claim — ordering relations, correction bounds, and impossibility results — is encoded as a Lean 4 theorem over integer-scaled data carrying SHA-256 provenance of the source files. The formalization currently comprises 899 build-locked theorems and zero `sorry` proofs — Lean's escape hatch for unproven claims. The count is generated directly from the open Lean source at build time.

The theorems fall into three categories. Ordering claims are inequality chains verified by the Lean kernel at fixed precision, not statistical summaries. Isotonic correction theorems implement the log-affine correction as a Lean function whose outputs the kernel checks against expected ordering properties. Impossibility theorems prove, with concrete counterexample witnesses, where no monotone correction can reconcile a model's ordering with the reference ordering.

The kernel-rejected claim episode illustrates why this matters. A statistical filter initially accepted that 27 of 36 cells improved strictly. When submitted to Lean, the `decide` tactic refused to prove $813 < 813$; at 10^{-4} J/m^2 integer precision, one cell's improvement margin was exactly zero. The corrected count is 26 of 36. A p-value does not know about integer rounding; the proof kernel does. In a pipeline where each false positive can cost weeks of synthesis effort, the distinction between statistical confidence and mechanical certainty is economically decisive.

Where Correction Fails, the Framework Proves Impossibility

A verification layer is only as useful as its boundary conditions. No claim is made that the field works everywhere; the framework proves where it cannot work. Three boundary conditions are formally established.

> Instead of silent failures, the verification layer returns machine-checked impossibility proofs for ranking inversions, noise-floor cells, and out-of-domain structures.

First, ranking inversions. When a model's ordering disagrees with the reference ordering, a monotonicity impossibility lemma proves that no monotone correction can map both to their references simultaneously. The

experimentalist receives not a vague uncertainty estimate but a machine-checked statement that the candidate cannot be rescued within the method.

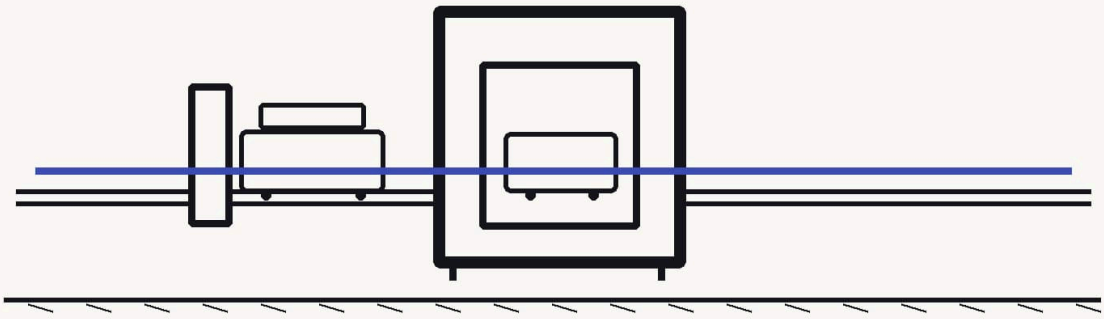
Second, already-converged cells. Where the raw uMLIP error sits at the noise floor of the anchor observables — for instance, MACE-MPA-0 on certain surfaces — correction adds no value and is refused. The proof flags the redundancy before computational effort is wasted.

Third, outside the field's domain. Environments that require second-shell structure or explicit electronic-structure treatment — planar faults, charged defects, strongly correlated oxides — are outside the first-shell coordination approximation. The framework proves the domain violation and reports the witness.

These impossibility proofs are not admissions of weakness. They are the feature that prevents the A-Lab pattern: systematic false positives propagating through an autonomous pipeline because no layer knew how to say no.

The Six-Step Loop

The field measurement, runtime correction, and formal verification layers close into a six-step discovery cycle:



1. Simulate — run a uMLIP through the standard property pipeline: equation of state, surface slabs, vacancy supercells.
2. Identify — compare predictions to the 228-value provenance-annotated reference database.
3. Validate — check whether errors follow the environment error field; use the three anchors to predict a fourth blind observable.
4. Generate — construct the cubic spline $P(c)$ and deploy it as an additive correction with analytic forces.
5. Verify — submit every quantitative claim to the Lean 4 kernel; ordering claims become inequalities, correction bounds become isotonic theorems, failures become impossibility lemmas.
6. Improve — run the corrected calculator through the full pipeline. If correction succeeds, report corrected values with theorem certificates. If it fails, the impossibility proof tells the experimentalist why.

> Field measurement, runtime correction, and machine-checked proof close into a six-step loop that terminates in either a certified candidate or a provable reason to stop.

These certificates are no longer only a publication artifact. The field-domain gate, ranking-inversion gate, and barrier-underestimation conservatism check are now wired into the runtime policy engine and the promotion gate. When a prediction carries first-shell coordination data, the system checks whether every atom lies inside the measured field's domain; out-of-domain atoms trigger a `skip_correction` action backed by `FieldDomain.refusal_has_witness`. When a candidate model is promoted, adjacent pairs in its reference and predicted rankings are checked for monotone rescuability; an inverted pair is downgraded automatically because `inversion_defeats_monotone` proves no monotone recalibration can rescue it. Barrier claims are checked against `softened_barrier_underestimates` and `softening_never_hides_conductor`. The mathematics that was proved in Lean now governs what reaches the lab.

Against DFT, the loop offers comparable defect-energy accuracy at many orders of magnitude higher speed. Against raw uMLIPs, it replaces uncontrolled systematic bias with a bounded, measured correction. Against delta-ML and fine-tuning, it eliminates per-system retraining because the spline transfers within a crystal-structure family. The practical consequence: a 10^6-structure screening campaign needs no DFT calls, yet every candidate carries a provable accuracy boundary.

From Geometry to Climate Targets

The field is not an abstract mathematical device. It maps directly onto the defects, surfaces, and transition states that determine climate-critical material performance.

> The defects that break climate-critical materials — ion hops, surface catalytic sites, hydrolysis nodes — are exactly the under-coordinated environments the field measures.

In cobalt-free lithium-manganese-rich cathodes, voltage fade is transition-metal migration from octahedral to tetrahedral sites — a hop whose barrier is underestimated when the transition state is under-coordinated. In earth-abundant halide solid electrolytes, ionic conductivity depends exponentially on the Li^+ migration barrier; a 60% barrier error changes predicted conductivity by roughly three orders of magnitude. In MOFs for direct air capture, humidity stability is governed by metal-linker hydrolysis at under-coordinated nodes. In electrochemical ammonia catalysts, $N{\equiv}N$ cleavage occurs at surface defect sites. In lead-free perovskite absorbers, Sn^{2+} oxidation and vacancy formation set the stability ceiling.

Every one of these properties is a coordination-dependent defect property. Every one is therefore within the domain where a measured error field can correct the prediction, or where an impossibility proof can flag the candidate as unsupported. The next part turns from the method to the targets: five material classes where correction plus verification could unlock 5–12 $GtCO_2$/year.

> The measured field and formal verification replace the arms race for bigger models, eliminating per-system retraining while raising the standard of evidence.

FOOTNOTES

1. A. Merchant *et al.*, “Scaling deep learning for materials discovery,” *Nature* 624, 80–85 (2023). https://doi.org/10.1038/s41586-023-06735-9 ↩

2. N. J. Szymanski *et al.*, “An autonomous laboratory for the accelerated synthesis of novel materials,” *Nature* 624, 86–91 (2023); J. Leeman *et al.*, “Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis,” *PRX Energy* 3, 011002 (2024). https://doi.org/10.1038/s41586-023-06734-7; https://doi.org/10.1103/PRXEnergy.3.011002 ↩

3. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025). https://doi.org/10.1038/s41524-024-01500-6 ↩

4. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025). ↩

5. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025); see discussion of training-data composition and bias toward near-equilibrium bulk structures. ↩

6. F. Ramprasad *et al.*, “Machine learning in materials informatics: recent applications and prospects,” *npj Computational Materials* 3, 54 (2017). https://doi.org/10.1038/s41524-017-0057-0 ↩

7. B. Deng *et al.*, “CHGNet: pretrained universal neural network potential for charge-informed atomistic modelling,” *Nature Machine Intelligence* 5, 1031–1041 (2023). https://doi.org/10.1038/s42256-023-00716-3 ↩

Lupine Science

2026-07-09

Five Materials That Could Unlock 5–12 $GtCO_2$/Year

https://lupine.science/articles/five-materials-for-5-to-12-gtco2-year/

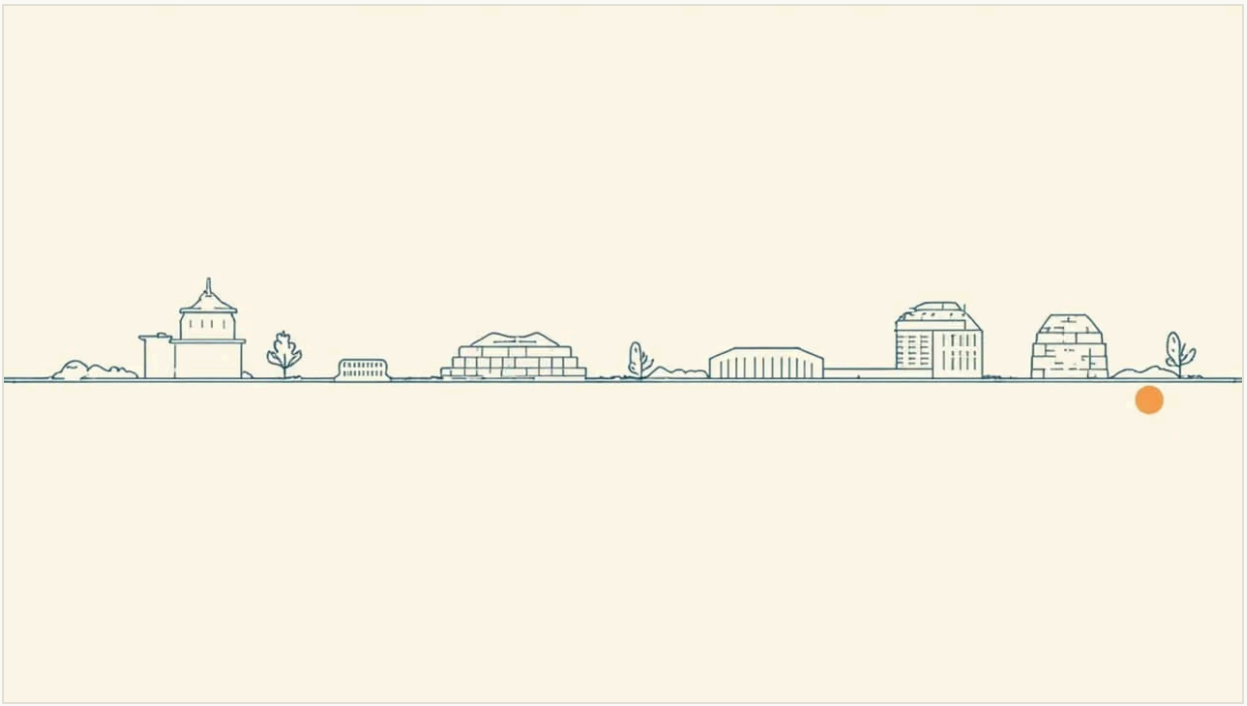
CLIMATE PARTNERSHIPS PROOF PACK

RESEARCH NOTE

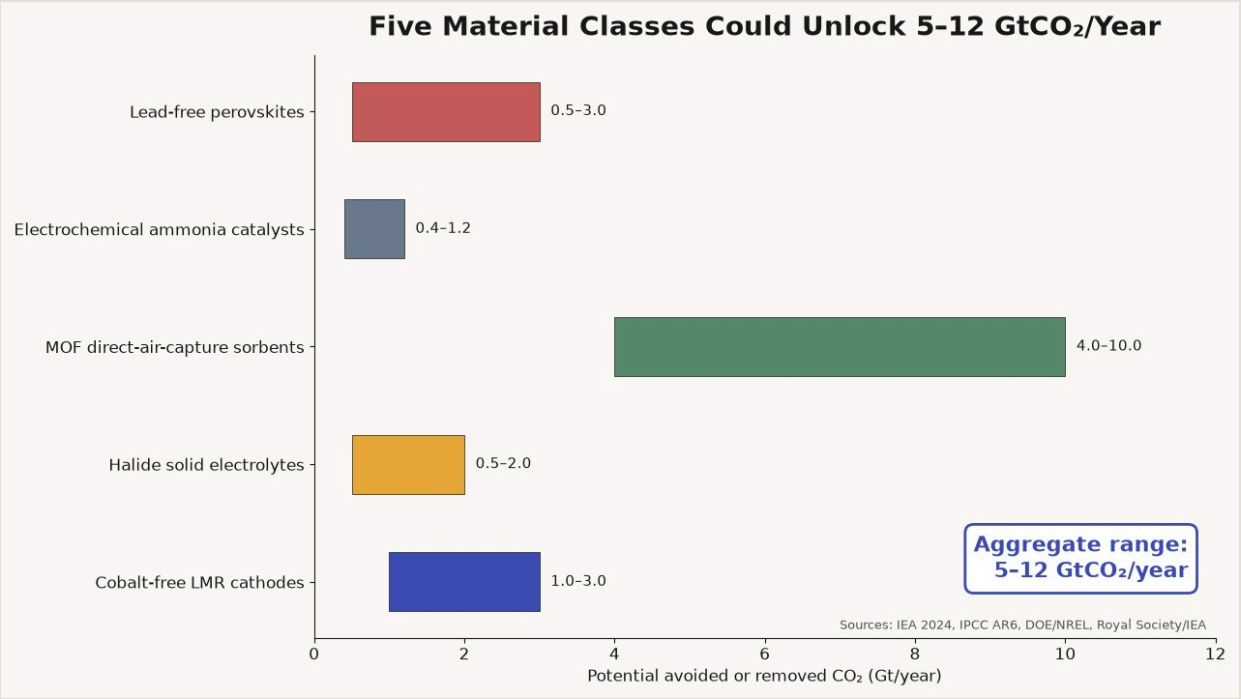
Five Materials That Could Unlock 5–12 $GtCO_2$/Year

A portfolio-level view of five material classes targeted for gigatonne-scale climate impact.

July 9, 2026 · DRAFT



Five grounded infrastructure vignettes — battery storage, air-capture contactor, ammonia plant, solar field, and solid-state cell — connected by a quiet indigo verification line.



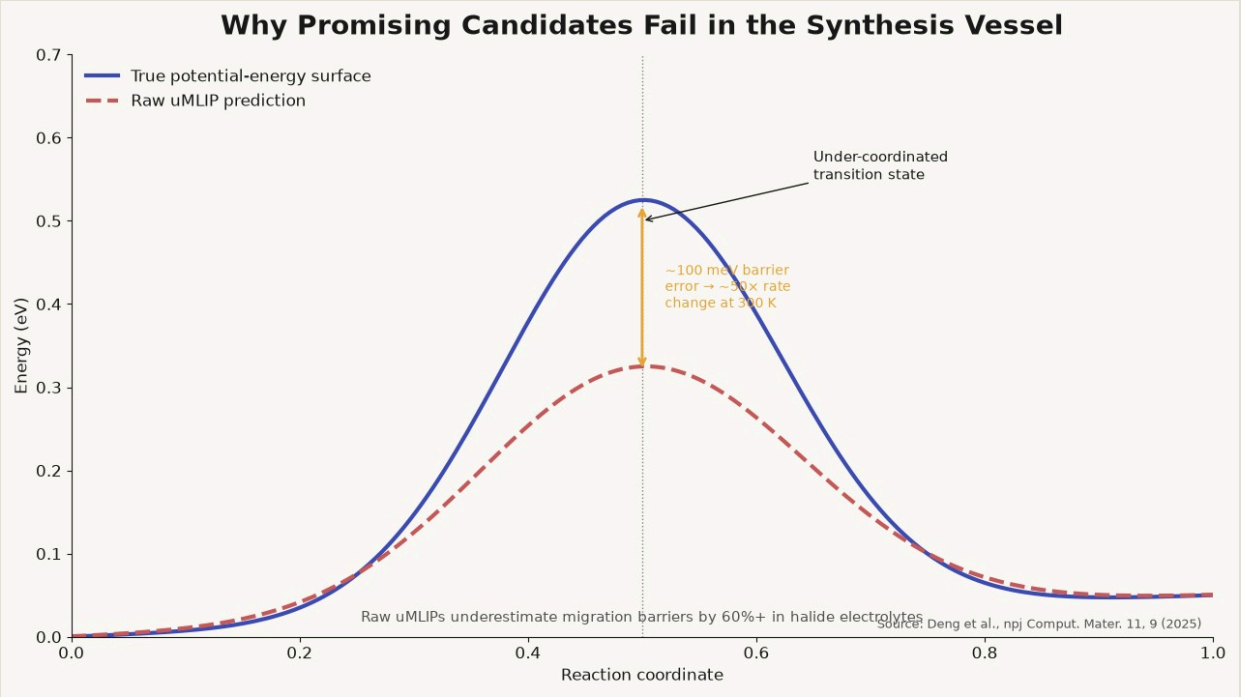
> Five independently sized material targets sum to an aggregate climate potential of 5–12 $GtCO_2$/year.

The difference between a climate model and a climate outcome is usually a material. Batteries need cathodes that survive a thousand cycles. Solid-state cells need electrolytes that conduct lithium without conducting dendrites. Direct air capture needs sorbents that bind CO_2 at 400 ppm and survive humidity. Ammonia synthesis needs catalysts that split dinitrogen without the Haber-Bosch furnace. Solar needs absorbers that rival lead perovskites without the lead. None of these materials exists at the required scale and price.

The objective is not to invent these materials one at a time. The focus is a correction-and-verification layer that makes computational discovery trustworthy enough to screen the multi-component spaces where these materials live. The result is a five-target portfolio with a combined climate potential of 5–12 $GtCO_2$/year.

Each target was chosen for gigatonne-scale climate impact, computational hardness rooted in defect-mediated failure, and a precise correction mechanism. The unifying methodological idea—measuring systematic error as a

physical field over local atomic environments—translates into progress across all of them. The aggregate number is not a fundraising flourish; it is the sum of five independently grounded impact estimates, each contingent on solving a specific defect-chemistry problem.



> Universal machine-learning potentials systematically soften the energy surface at under-coordinated sites, inverting the rankings that determine which candidates reach the lab.

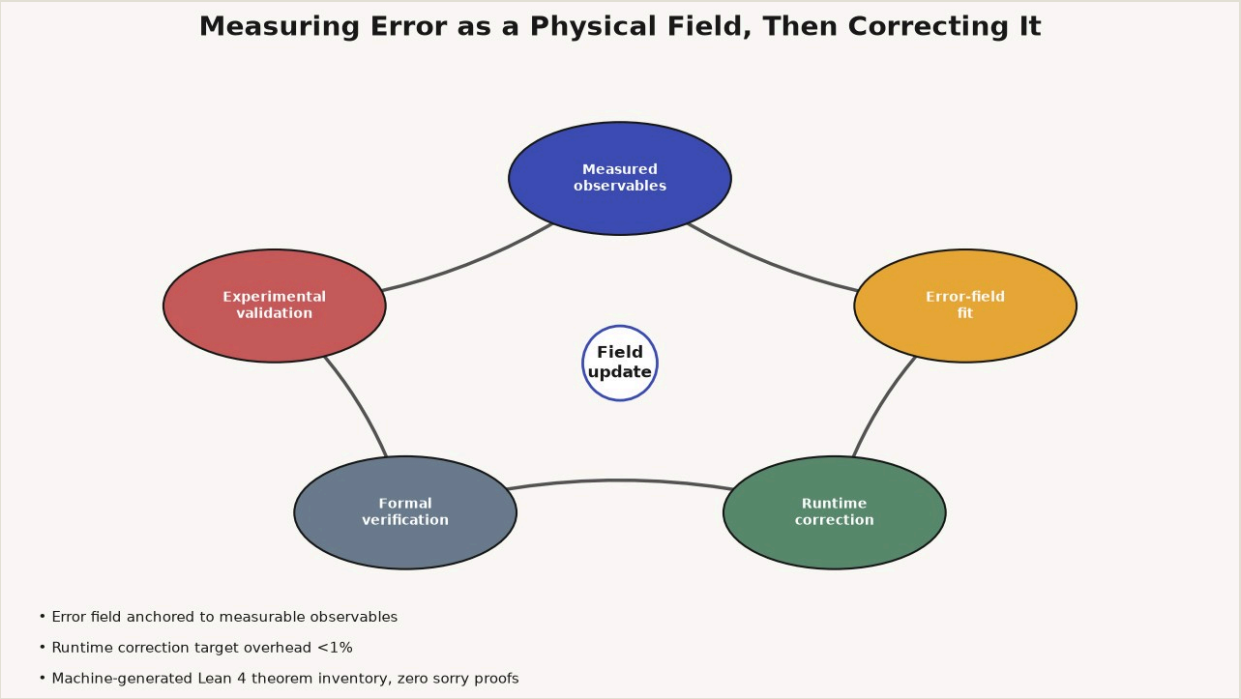
How the targets were selected

Each target satisfies three filters.

First, gigatonne-scale climate impact. The International Energy Agency estimates that batteries alone are directly linked to roughly 20% of the CO_2 reductions required by 2030, and indirectly to another 40%[1]. Two of the five targets are battery materials for exactly that reason; the other three address industrial emissions, legacy carbon removal, and solar deployment.

Second, discovery difficulty that has defeated existing methods. In every case, the property that determines performance—ion migration, defect formation, catalytic barrier, oxidation resistance—occurs at under-coordinated atomic environments where universal machine-learning interatomic potentials (uMLIPs) systematically soften the potential energy surface[2]. Raw uMLIPs are fast but wrong where it matters; DFT is right but economically impossible at screening scale. The result is a generation of predicted materials that look stable on paper and fail in the synthesis vessel or the device.

Third, a precise correction mechanism. Each target maps a known failure mode onto a specific correction or proof. LMR cathodes require corrected transition-metal migration barriers. Halide electrolytes require corrected Li^+ hop barriers. MOFs require corrected hydrolysis energies and impossibility proofs for unsynthesizable frameworks. Ammonia catalysts require corrected N_2 dissociation barriers and selective flags for scaling-relation breakers. Lead-free perovskites require corrected Sn vacancy formation energies and provable boundaries for metastable phases.



> Systematic error is measured as a physical field over local environments, corrected at runtime, and verified through machine-checked proof rather than p-values.

The targets are not independent scientific bets. They are five instances of the same defect-mediated problem. This is deliberate: a platform that solves one of them credibly has a path to solving all of them, because the correction layer transfers across any material family whose error field can be anchored to measurable observables.

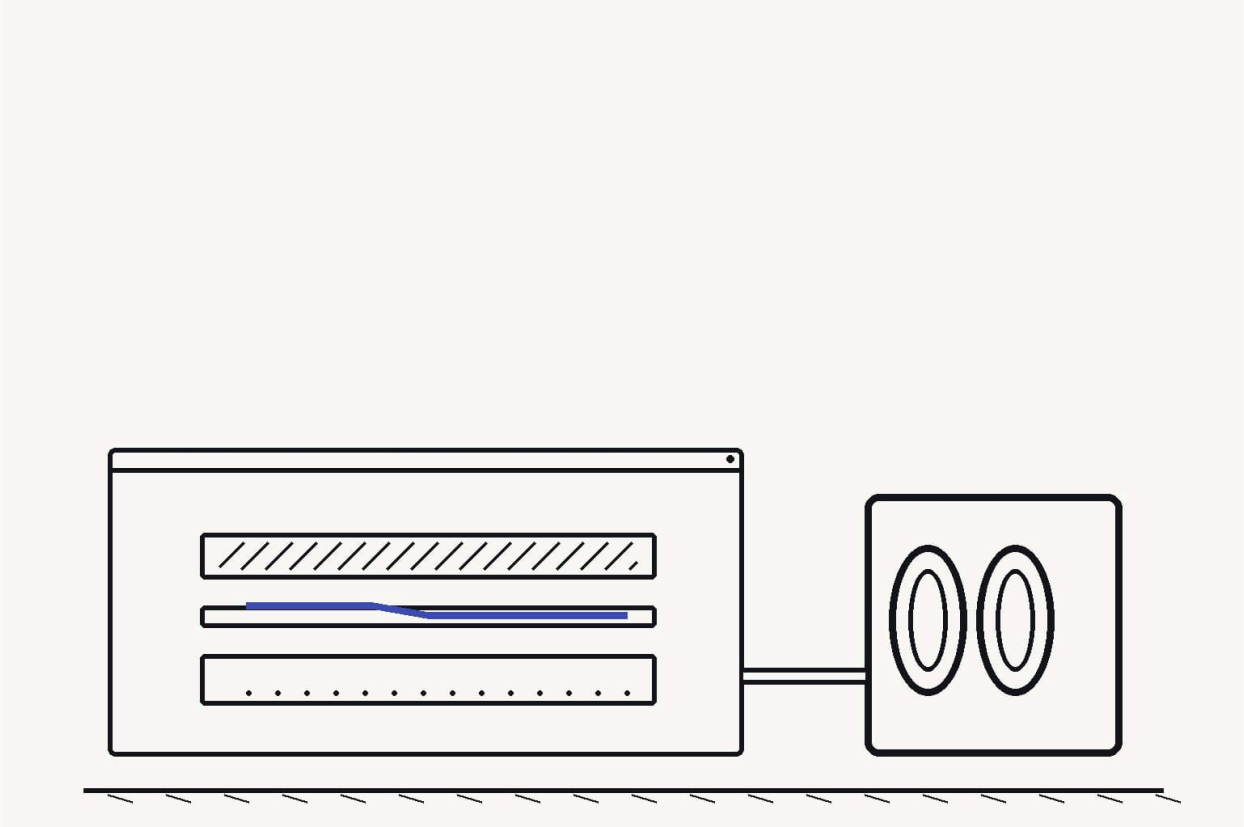
The Same Defect-Mediated Failure, Five Different Materials

Material	Defect property	uMLIP failure	Lupine correction	Experimental outcome
Cobalt-free LMR cathodes	Transition-metal migration barrier	Softens under-coordinated TM path	Corrected migration barriers + ordering proof	>300 Wh/kg, >1,000 cycles
Halide solid electrolytes	Li^+ hop barrier	Underestimates barrier 60%+	Corrected hop barrier transferred across Li–M–Cl	>10 mS/cm, moisture tolerant
MOF DAC sorbents	Metal-linker hydrolysis energy	Softens bond dissociation	Hydrolysis correction + impossibility proof	>2 mmol/g at 400 ppm, 40–70% RH
Electrochemical ammonia catalysts	N_2 dissociation barrier	Underestimates binding on stepped sites	Barrier correction + scaling-relation flag	>60% energy eff. >300 mA/cm²
Lead-free perovskites	Sn vacancy formation energy	Softens under-coordinated Sn neighbors	Vacancy correction + metastability proof	>20% PCE, 25-year stability

> Each target is a different expression of the same defect-mediated problem, and each is addressed by the same runtime correction and verification layer.

Batteries: cathodes and electrolytes as a pair

Cobalt-free lithium-manganese-rich cathodes



The highest-leverage near-term target is a cobalt-free lithium-manganese-rich (LMR) cathode exceeding 300 Wh/kg cell-level energy density, cycling stability above 1,000 cycles, and cost below $80/kWh. The EaCAM consortium at Argonne National Laboratory has demonstrated cobalt-free LMR systems at approximately 270 Wh/kg and ~$80/kWh[3].

The remaining barrier is voltage fade. LMR cathodes derive excess capacity from oxygen redox, but oxygen redox triggers oxygen loss, transition-metal migration from octahedral to tetrahedral sites, and surface reconstruction into spinel or rock-salt phases that block lithium diffusion[4]. Removing cobalt—essential for supply-chain security, since the Democratic Republic of Congo produces roughly 70% of global cobalt[5]—exacerbates the instability.

The compositional space is enormous. When concentration gradients, coatings, and doping profiles are included, the candidate count exceeds 10^6. uMLIPs systematically underestimate the migration barriers that govern voltage fade because the transition state involves under-coordinated metal ions far from the bulk configurations in their training data. The result is candidate rankings that can be wrong by orders of magnitude in ionic mobility. A 100 meV barrier error changes the hopping rate by roughly e^(E/kBT) ≈ 50× at room temperature, enough to invert the ranking of candidate compositions by ionic mobility.

The environment error field corrects those barriers at runtime. The correction is smooth over coordination-number space, so chemically similar compositions receive similar corrections and the true best candidate is not buried by ranking inversion. Formal verification then checks that predicted voltage profiles remain ordered after correction, preventing false-positive synthesis attempts.

> Before correction, the highest-performing candidates can be ranked below also-rans; the correction layer recovers the true ordering and prevents wasted synthesis runs.

Earth-abundant halide solid electrolytes

The second battery target is a halide solid electrolyte in the Li–Zr–Cl or Li–Fe–Cl family with ionic conductivity above 10 mS/cm, electrochemical stability against lithium metal, and moisture tolerance. Solid-state batteries with lithium metal anodes are the most credible path to >400 Wh/kg and the elimination of thermal runaway. Industry analysts project the solid-state battery market to grow by an order of magnitude or more over the next decade.

Current halide electrolytes such as Li_3InCl_6 and Li_3YCl_6 achieve 1–12 mS/cm[6], but indium and yttrium are critical raw materials. Replacing them with Zr, Fe, Al, or Mg could cut cost by roughly an order of magnitude while preserving the moisture tolerance that sulfides lack.

The computational challenge is the same shape as the cathode problem, but the property is Li^+ hop barrier rather than transition-metal migration. Raw uMLIPs underestimate migration barriers by 60%+ due to PES softening at under-coordinated transition states[7]. Because ionic conductivity depends exponentially on barrier—$\sigma \propto e^{(-E_a/kBT)}$—a large barrier error changes predicted conductivity by many orders of magnitude at a typical 300 meV barrier and room temperature. Fast-ion conductors are discarded as insulators before an experimentalist ever sees them.

The correction layer corrects barriers on a representative halide such as Li_2ZrCl_6 and transfers the field across the Li–M–Cl space because all members share the same close-packed anion sublattice. The formalization now includes a dedicated rocksalt/halide anchor layout — a single c=5 vacancy anchor with bulk pin at the rocksalt coordination c=6 — so the halide family has the same two-tier measured/anchored field structure as the fcc, bcc, and diamond families. Grain-boundary screening, which requires 500+ atom supercells where DFT is prohibitively expensive, becomes feasible with corrected uMLIPs. The false-negative elimination that plagues raw uMLIP screening is replaced by a ranked list whose top entries have DFT-level barrier accuracy.

Carbon removal: MOFs for direct air capture

Direct air capture is the only technology that can address legacy emissions regardless of source sector. The IPCC AR6 estimates cumulative carbon dioxide removal needs of 100–1,000 $GtCO_2$ by 2100, with annual rates approaching 10 $GtCO_2$/year by mid-century[8]. The material bottleneck is the sorbent.

The target is a metal-organic framework (MOF) with CO_2 working capacity above 2 mmol/g at 400 ppm, stability under 40–70% relative humidity, and scalable synthesis cost below $50/kg. Amine-functionalized MOF-808 has

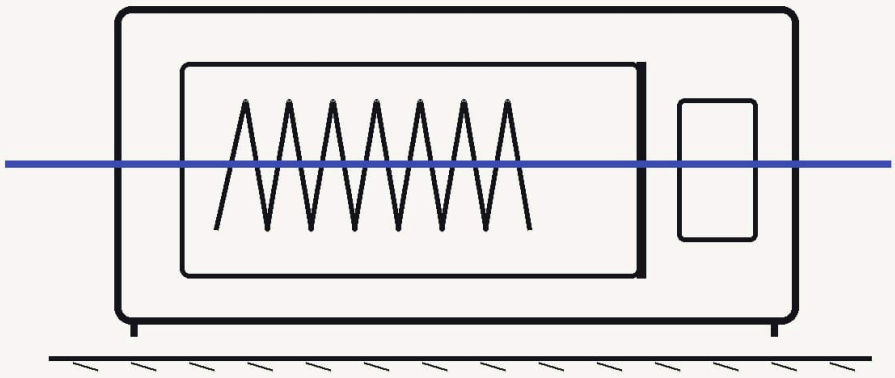
reached 1.2 mmol/g at 400 ppm and 50% relative humidity[9]. At gigatonne scale, each \$10/tCO_2 cost reduction saves billions of dollars annually.

The MOF design space spans trillions of structures, yet only a small fraction are experimentally synthesized. Stability is the harder problem than capacity. Water competes with CO_2 for binding sites and hydrolyzes metal-linker bonds; predicting humidity-stable frameworks requires evaluating defect formation energies at metal-linker bonds—exactly the under-coordinated environments where uMLIPs are least accurate[10].

The correction addresses metal-linker bond dissociation energies, the key determinant of hydrolytic stability. Its impossibility proofs flag frameworks where mixed-metal nodes create coordination environments outside the measured field, separating candidates worthy of experimental investment from computationally unsupported ones. Ranking preservation then keeps the best humidity-stable frameworks at the top of a large screen, rather than scattering them beneath false-positive high-capacity candidates.

Industrial decarbonization: electrochemical ammonia catalysts

Ammonia synthesis via the Haber-Bosch process consumes 1–2% of global energy and emits more than 450 MtCO_2/year, primarily through steam methane reforming[11]. Electrochemical synthesis at ambient conditions, powered by renewable electricity, could eliminate those emissions while enabling distributed fertilizer production.



Lithium-mediated electrochemical nitrogen reduction has achieved >90% Faradaic efficiency at ambient conditions, but energy efficiency is stuck near 28% because lithium plating requires very negative potentials, dissipating more than 70% of input energy as heat[12]. The U.S. Department of Energy target is >60% energy efficiency at current densities above 300 mA/cm²[13].

The challenge is N≡N triple-bond activation—dissociation energy 945 kJ/mol—under conditions where the hydrogen evolution reaction is thermodynamically favored. DFT screening of N_2 adsorption on stepped surfaces spans coordination environments from c≈4 to c≈9, the full range where the error field operates. uMLIPs underestimate these binding energies due to PES softening, generating false-positive catalyst identifications. Scaling relations between N_2 and NH_x binding energies compound the error: a mistake in one binding energy propagates to all others, potentially inverting turnover-frequency rankings.

N_2 dissociation barriers are corrected on under-coordinated active sites, filtering out catalysts that would be experimentally inactive. More subtly, the field's selective failure identifies scaling-relation-breaking catalysts that conventional screening misses. Single-atom or multi-metal sites with cooperative binding violate the first-shell approximation; impossibility proofs flag them for higher-level treatment, directing ab initio effort to the small subset most likely to exceed the volcano-peak activity limit.

Solar: lead-free perovskite absorbers

Lead-based halide perovskites have climbed from ~3% efficiency in 2009 to >26% single-junction and ~34.6% in perovskite-silicon tandems[14]. But lead toxicity threatens to limit terawatt-scale deployment, and the EU RoHS and REACH frameworks are progressively restricting lead.

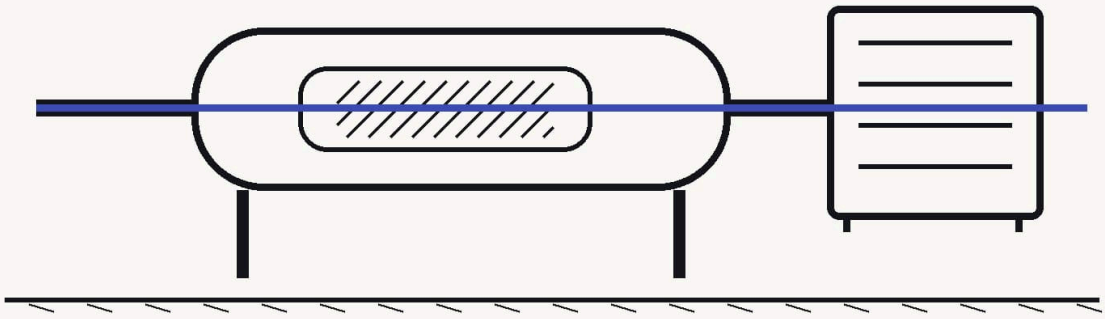
The target is a lead-free perovskite absorber with certified power conversion efficiency above 20% and operational stability above 25 years. Tin perovskites are the most promising alternative, but Sn^{2+} oxidizes to Sn^{4+} in minutes under ambient conditions, and lead-free tin efficiencies have until recently been far below the >26% reached by lead cells[15].

The central degradation mechanism is oxygen insertion through Sn vacancy formation. uMLIPs systematically underestimate Sn vacancy formation energies because vacancies create under-coordinated Sn neighbors that bulk-trained models predict as too stable[16]. Correcting those energies identifies compositions where tin is most strongly bound and most resistant to oxidation, narrowing a vast double-perovskite search space to the subset worth synthesizing.

Metastability adds a second layer of difficulty. Many high-efficiency perovskites are kinetically trapped during solution processing; the equilibrium phase is often a non-perovskite polymorph with poor photovoltaic properties. Standard convex-hull screening discards these metastable phases, eliminating the compositions that achieve the highest efficiencies. Impossibility proofs establish boundaries between candidates where predictions are reliable and candidates requiring synthesis-route engineering to access the metastable phase.

Synthesis: one failure mode, five targets

The five targets span batteries, carbon removal, industrial chemicals, and solar, but they share a common structure. In every case, functional performance is determined by defect-mediated properties in multi-component spaces. In every case, raw uMLIPs systematically soften the potential energy surface at the under-coordinated configurations that govern those properties. And in every case, brute-force DFT is too slow to screen the required compositional space.



The response is not to train a bigger model. It is to measure the systematic error as a physical field over local atomic environments, correct it at runtime with analytic forces, and verify the resulting claims through machine-checked proof. The environment error field achieves Pearson r=0.906 in blind prediction of never-fitted surface energies across 36 (model, material) combinations with zero adjustable parameters.

> Across 36 model–material combinations, the environment error field predicts never-fitted surface energies with Pearson r = 0.906 and zero adjustable parameters.

Runtime correction adds modest overhead in the current Python implementation and will drop below 1% in a compiled LAMMPS overlay. 899 build-locked Lean 4 theorems with zero sorry proofs provide guarantees that

statistical validation cannot match. Where correction fails, the platform proves impossibility rather than reporting a p-value, preventing experimental resources from being spent on computationally unsupported candidates.

> A sub-0.1 eV correction at the atomic scale propagates into orders-of-magnitude device improvements and gigatonne-scale climate impact.

The targets form a portfolio rather than a list. The same correction layer that makes LMR cathode screening reliable also makes halide electrolyte, MOF, ammonia catalyst, and perovskite screening reliable. The moat deepens with each campaign: every screen adds validated field measurements, every impossibility proof sharpens the boundary of applicability, and every experimental validation tightens the feedback loop.

> Every screen, proof, and validation feeds back into the error field, deepening the platform's moat and lowering the cost of each subsequent discovery campaign.

From predictions to partners

Predictions are necessary but not sufficient. The path from a corrected energy landscape to a commercial material runs through named experimental collaborators who can synthesize, characterize, and scale the top candidates.

> Each target is anchored to named labs and companies that can synthesize, characterize, and scale the corrected top candidates.

For LMR cathodes, the immediate partners are the Manthiram Laboratory at UT Austin and TexPower EV Technologies, which operates a pilot facility producing NMA cathodes at >230 mAh/g[17]; the Battery500 Consortium led by Pacific

Northwest National Laboratory, with 350 Wh/kg pouch cells demonstrated at >600 cycles[18]; and Forge Nano, whose atomic-layer-deposition coatings improve cycle life by 30%+ while reducing resistance[19]. For halide electrolytes, the Tier-1 list includes the CEDER Group at UC Berkeley and Lawrence Berkeley National Laboratory, the Janek/Zeier Group at the University of Münster, Argonne National Laboratory, and Solid Power Inc.

For MOFs, the starting points are UC Berkeley's Yaghi and Long groups, Northwestern's Farha Group, and BASF, the first commercial-scale MOF producer at several hundred tons per year[20]. For ammonia catalysts, DTU's Chorkendorff group—which published the landmark Ca-mediated nitrogen reduction result in *Nature Materials* in 2024[21]—and Stanford's SUNCAT Center provide rigorous verification and computational catalysis expertise. For lead-free perovskites, NREL, the University of Queensland's Wang Group—holder of the certified 16.65% tin-halide record[22]—and Tandem PV Inc. anchor the path from absorber to tandem module.

Subsequent work treats partnerships in detail. The methodological point is that a correction layer without experimental partners is an academic exercise, and experimental partners without corrected predictions are flying blind. The 5–12 $GtCO_2$/year figure is reachable only when the two are coupled, and only if each candidate is validated through the chain of synthesis, characterization, cell or module testing, and scale-up that turns a predicted crystal into a deployable technology.

> The path to gigatonne impact runs through a staged chain of corrected predictions, partner synthesis, and device-scale validation.

FOOTNOTES

1. International Energy Agency, *Batteries and Secure Energy Transitions*, IEA, 2024. ↩

2. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025). https://doi.org/10.1038/s41524-024-01500-6 ↩

3. Argonne National Laboratory, “Cobalt-Free and Manganese-Rich Batteries Offer Improved Energy at a Lower Cost,” U.S. Department of Energy / EaCAM consortium, January 2026; J. Chen *et al.*, “Defining Electrode-Level Metrics for Enabling Earth-Abundant Cathodes,” *J. Electrochem. Soc.* (2026). ↩

4. E. Hu *et al.*, “Evolution of redox couples in Li- and Mn-rich cathode materials and mitigation of voltage fade by reducing oxygen release,” *Nature Energy* 3, 690–698 (2018); J.-J. Marie *et al.*, “Trapped O_2 and the origin of voltage fade in layered Li-rich cathodes,” *Nature Materials* 23, 310–317 (2024). https://doi.org/10.1038/s41560-018-0207-z; https://doi.org/10.1038/s41563-024-01833-z ↩

5. International Energy Agency, *The Role of Critical Minerals in Clean Energy Transitions*, IEA, 2022. ↩

6. X. Li *et al.*, “Water-mediated synthesis of a superionic halide solid electrolyte,” *Angewandte Chemie International Edition* 58, 16427–16432 (2019); “Halide-Based Solid Electrolytes for Advanced All-Solid-State Batteries: Design, Interfaces, and Electrochemical Performance,” *Nano-Micro Letters* 18, 225 (2026). ↩

7. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025). ↩

8. IPCC, *Climate Change 2022: Mitigation of Climate Change*, Working Group III Contribution to AR6, Cambridge University Press, 2022. ↩

9. X. Chen *et al.*, “Computational Screening of Amino-Functionalized Molecules for Direct Air Capture of CO_2,” *J. Phys. Chem. A* (2026); P. Chen *et al.*, functionalized MOF-808 DAC study cited therein (1.2 mmol/g at 400 ppm, 50% RH). ↩

10. B. Deng *et al.*, “Systematic softening in universal machine learning interatomic potentials,” *npj Computational Materials* 11, 9 (2025). ↩

11. Royal Society, *Ammonia: Zero-Carbon Fertiliser, Fuel and Energy Store*, policy briefing, 2020; International Energy Agency, *Ammonia Technology Roadmap*, IEA, 2021. ↩

12. W. Chang *et al.*, “Lithium-mediated nitrogen reduction to ammonia via the solvation-enabled N_2 activation,” *Nature Catalysis* 7, 342–352 (2024); X. Cai *et al.*, “Lithium-mediated electrochemical nitrogen reduction: mechanistic insights to enhance performance,” *iScience* 24, 103105 (2021). ↩

13. ARPA-E REFUEL program, “Renewable Energy to Fuels Through Utilization of Energy-Dense Liquids,” U.S. Department of Energy, program targets: >60% energy efficiency, >300 mA/cm^2, >90% Faradaic efficiency. ↩

14. NREL, “Best Research-Cell Efficiency Chart,” National Renewable Energy Laboratory, updated 2025; LONGi 34.85% perovskite-silicon tandem record, NREL-certified, April 2025. ↩

15. P. Chen *et al.*, “2D/3D tin-halide perovskite solar cell with certified 16.65% efficiency,” *Nature Nanotechnology* 20, 742–749 (2025). ↩

16. T. Leijtens *et al.*, “Mechanism of tin oxidation and stabilization by lead substitution in tin halide perovskites,” *ACS Energy Letters* 2, 2159–2165 (2017). ↩

17. TexPower EV Technologies, company website and press releases, 2024–2025; Forge Nano / TexPower SBIR collaboration announcement, January 2024. ↩

18. U.S. Department of Energy / PNNL, “Battery500: Progress Update,” 2020; PNNL, “Powering Up to Address Challenges in Energy Storage,” March 2024. ↩

19. Forge Nano, “US-made ultrahigh-energy cathodes will enable low-cost electric vehicle batteries,” press release, January 2024; Forge Nano website, ALD-enabled battery materials. ↩

20. BASF, “BASF becomes first company to successfully produce metal-organic frameworks on a commercial scale for carbon capture,” press release, October 2023. ↩

21. X. Fu *et al.*, “Calcium-mediated nitrogen reduction for electrochemical ammonia synthesis,” *Nature Materials* 23, 101–107 (2024). https://doi.org/10.1038/s41563-023-01702-1 ↩

22. P. Chen *et al.*, “2D/3D tin-halide perovskite solar cell with certified 16.65% efficiency,” *Nature Nanotechnology* 20, 742–749 (2025); University of Queensland press release, April 2025. ↩

Lupine Science

2026-07-09

From Predicted Crystal to Commercial Cell

https://lupine.science/articles/from-predicted-crystal-to-commercial-cell/

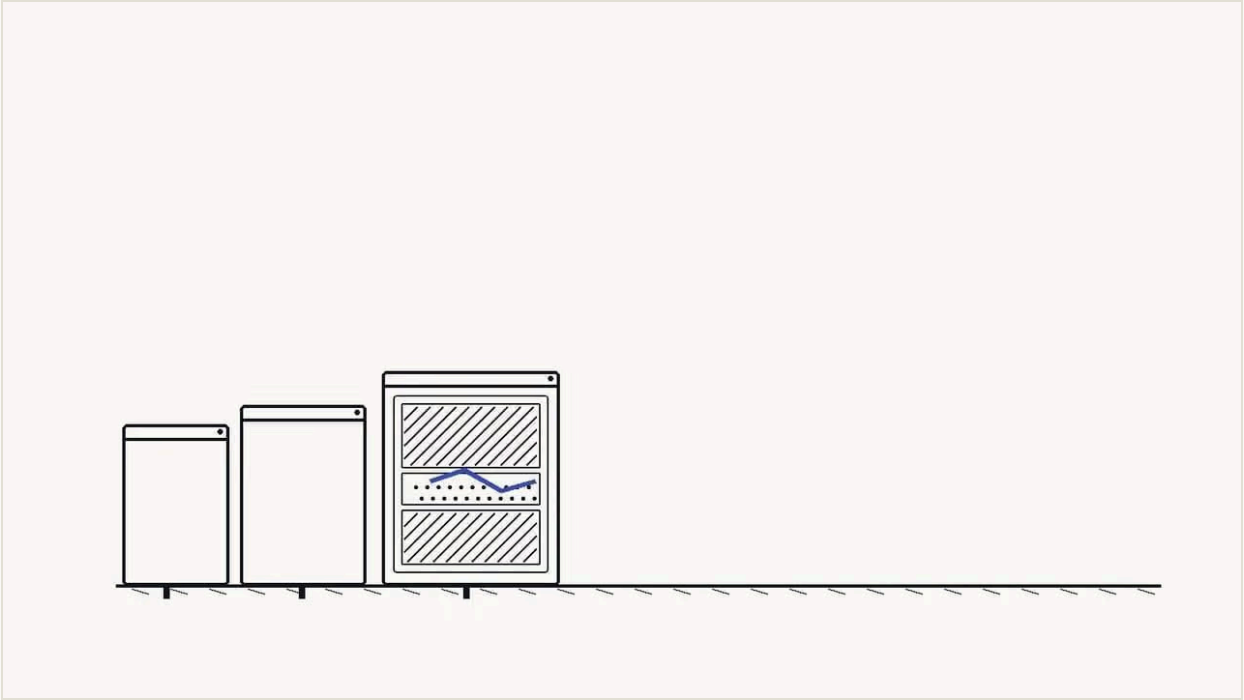
CLIMATE PARTNERSHIPS PROOF PACK

RESEARCH NOTE

From Predicted Crystal to Commercial Cell

How Lupine Science turns computational predictions into commercial climate materials through a sequenced partner chain.

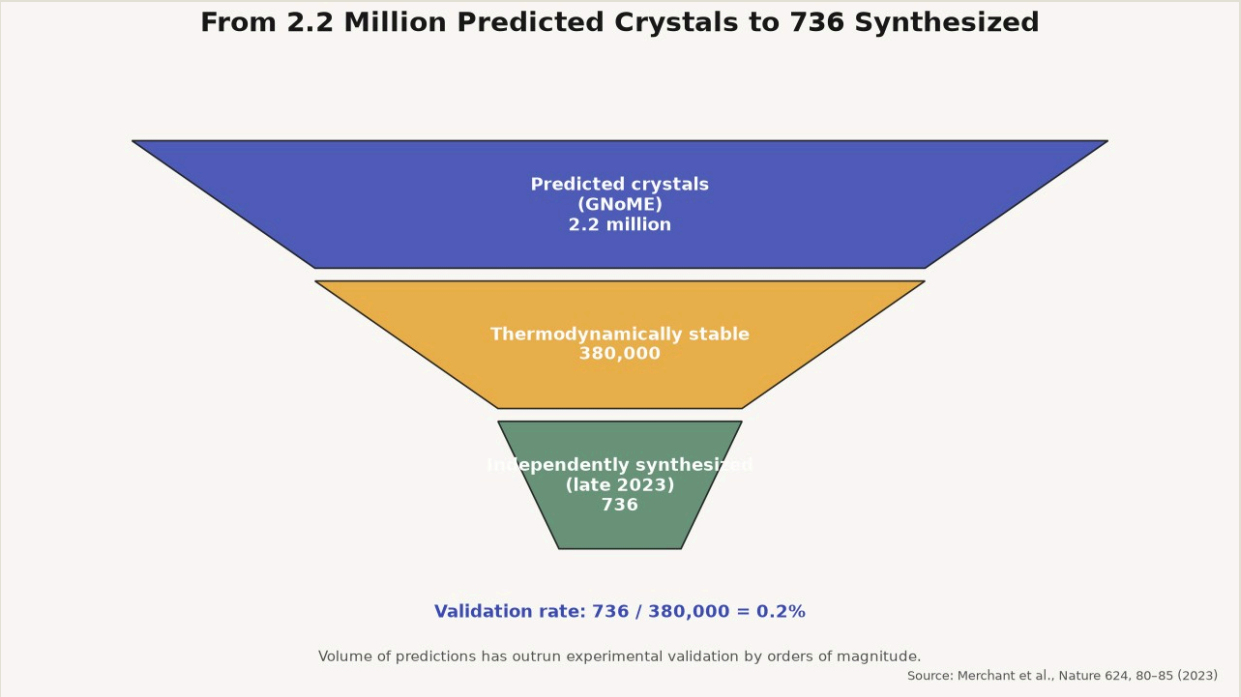
July 9, 2026 · DRAFT



A row of three grid-scale battery cabinets, the nearest a cutaway with one ceramic electrolyte plate seated between two flat electrodes, carrying the single indigo ion path.

A predicted crystal is not a material. A synthesized powder is not a cell. A cell is not a fleet, a filter, a fertilizer plant, or a solar farm. Between a computational screen and climate-relevant deployment stands a chain of specialized capabilities — synthesis, coating, cell fabrication, operando characterization,

module integration, and manufacturing scale-up — that no single institution owns. The question is not simply “which compound is stable?” but “which compound can survive handoff across this chain?”

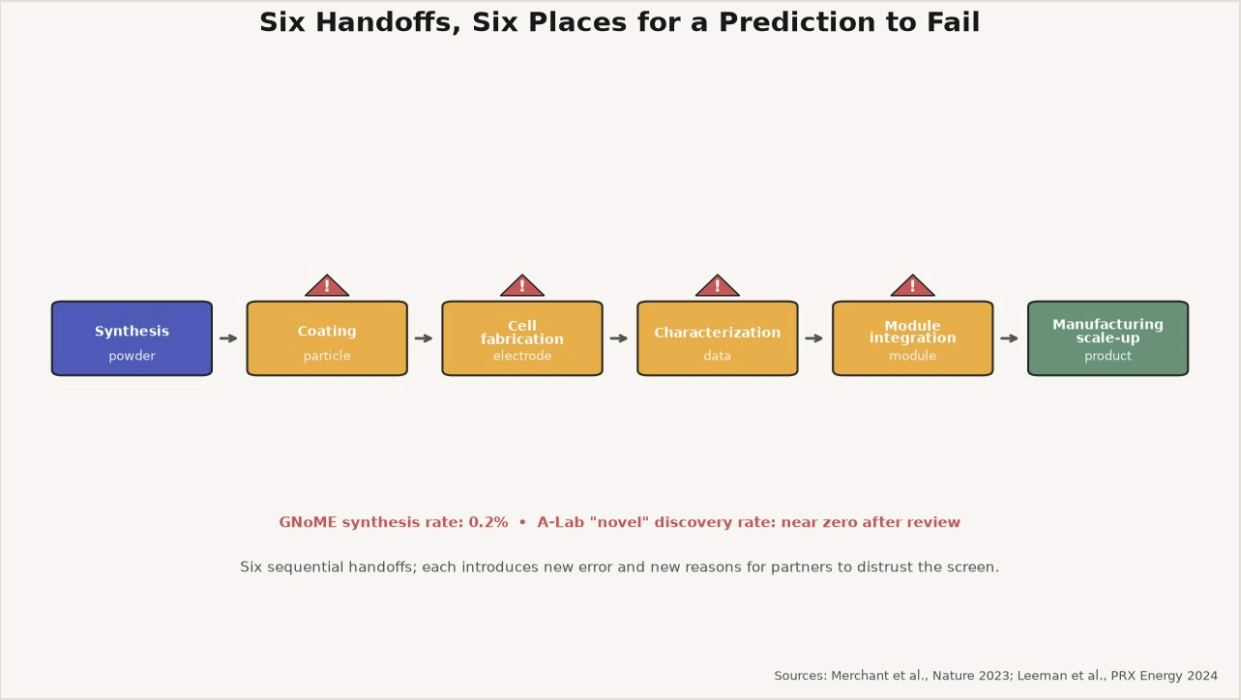


> GNoME’s crystal generator produced 2.2 million predictions, yet only 736 had been independently synthesized by late 2023 — a 0.2% validation rate that exposes the handoff bottleneck.

The conventional answer is to build a bigger generator. That has produced volume without trust. Google DeepMind’s GNoME predicted 2.2 million crystals and identified 380,000 as thermodynamically stable, yet only 736 had been independently synthesized by late 2023 — a 0.2% validation rate[1]. The A-Lab Author Correction (Nature 650:E1, 2026) records 36 confirmed of 57 eligible targets, with 4 inconclusive and one compound removed for training-data leakage; a separate critique disputed the novelty classification of many targets[2]. The bottleneck is not a shortage of predictions; it is the absence of a reliable correction and verification layer between prediction and experiment.

Lupine Science occupies that layer. Rather than competing with structure generators, laboratories, or databases, Lupine corrects the systematic errors of universal machine-learning interatomic potentials (uMLIPs) using a measured environment error field, deploys the correction at runtime with analytic forces, and verifies claims through machine-checked Lean 4 proofs. The result is a partner-facing platform that makes other organizations' predictions trustworthy, accelerating every link from crystal to cell.

The following sections map that chain across the five priority targets — cobalt-free lithium-manganese-rich (LMR) cathodes, earth-abundant halide solid electrolytes, metal-organic frameworks (MOFs) for direct air capture, electrochemical ammonia catalysts, and lead-free perovskite solar absorbers. In each, a prediction is handed to synthesis, then characterization, then integration, then manufacturing. Where each handoff currently fails, a correction-and-verification layer creates value.



> Between a computational screen and a commercial product, a predicted material must survive six sequential handoffs — and current validation rates show most do not.

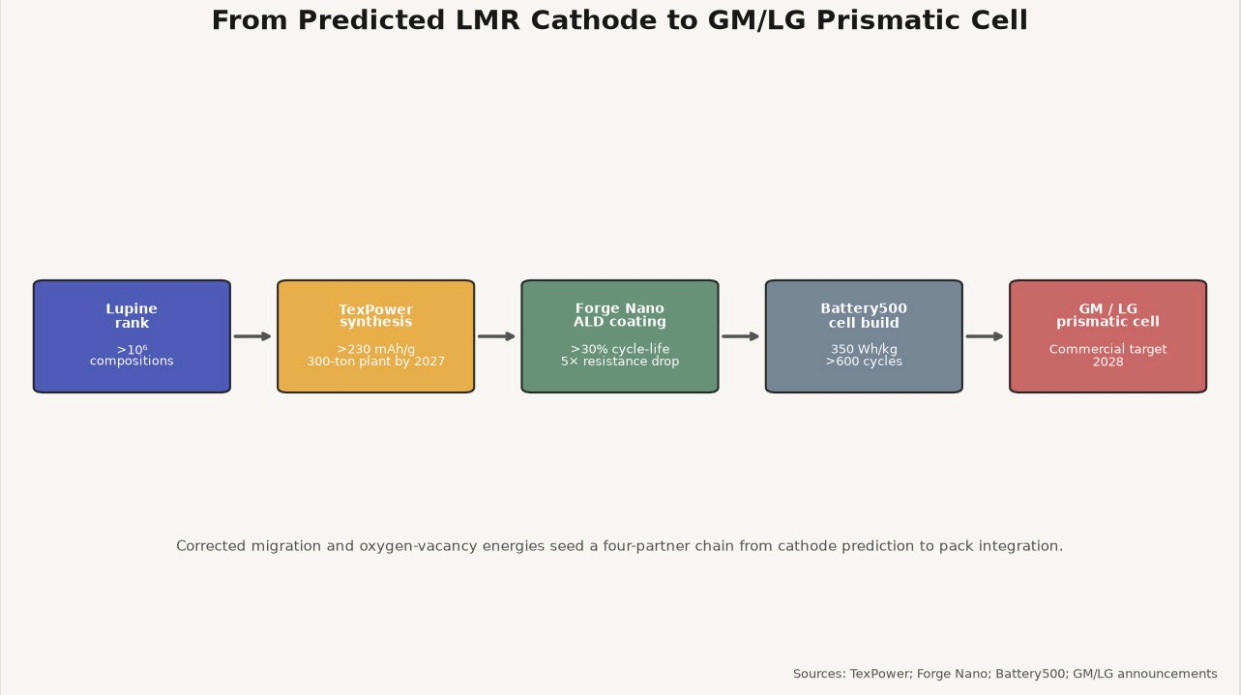
Batteries: From Cathode Prediction to Pack Integration

LMR cathodes are the highest-impact near-term target. By eliminating cobalt and raising energy density above 300 Wh/kg, they could avoid 2–5 $GtCO_2$ cumulatively by 2050. The commercial stakes are visible in the field: POSCO Future M has completed pilot LMR production and is preparing mass manufacturing in 2025[3], while GM and LG Energy Solution aim to begin commercial production of prismatic LMR cells by 2028[4].

The first handoff is from computation to synthesis. The corrected migration barriers and oxygen-vacancy formation energies rank candidates across a compositional space exceeding 10^6 compositions. The Manthiram Laboratory at UT Austin is the natural Tier 1 partner: Arumugam Manthiram invented cobalt-free layered oxide cathodes and has trained more than 300 researchers[5]. TexPower EV Technologies — a Manthiram spinout — extends this to scale-up with a pilot producing NMA cathodes at >230 mAh/g and a 300-ton plant planned by 2027[5:1].

The second handoff is from powder to coated particle. Voltage fade is driven by transition-metal migration and surface reconstruction, both suppressible by atomic-layer-deposition (ALD) coatings. Forge Nano operates powder ALD systems and has demonstrated that ALD coatings improve TexPower cathode cycle life by more than 30% while reducing resistance five-fold[6]. A three-way arrangement — Lupine identifies compositions, Forge Nano applies coatings, TexPower supplies coated material — closes the cathode-to-cell gap.

The third handoff is from material to cell. The Battery500 Consortium, led by Pacific Northwest National Laboratory, operates pouch-cell fabrication lines and has demonstrated 350 Wh/kg pouch cells with more than 600 cycles[7]. Lupine fits as a computational seedling project feeding the consortium's cathode pipeline.

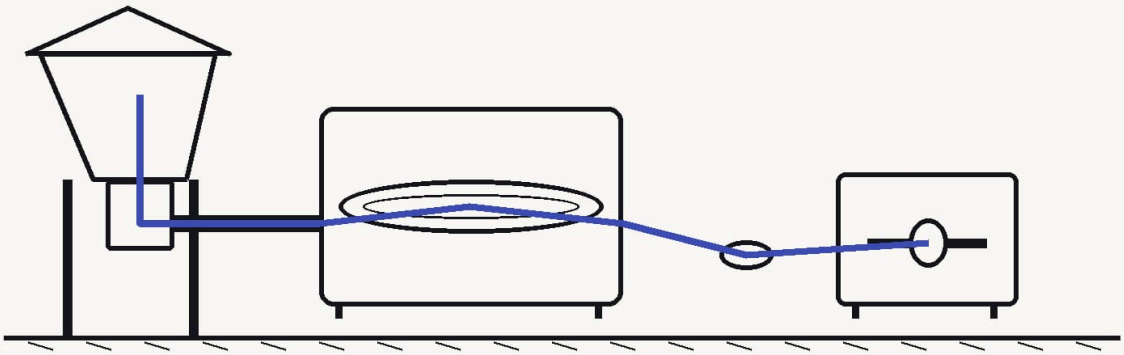


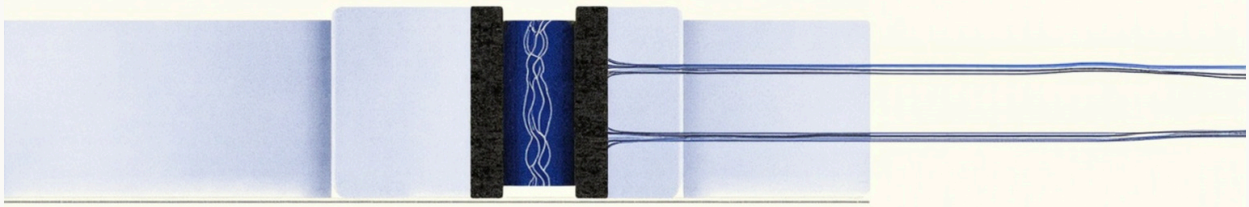
> Lupine ranks candidates across more than a million compositions, then passes corrected cathodes through TexPower, Forge Nano, Battery500, and GM/LG toward 2028 production.

Finally, the OEM handoff. GM's Wallace Battery Cell Innovation Center and LG Energy Solution are developing prismatic LMR cells for GM's electric trucks, with commercial production targeted for 2028[4:1]. These industrial partners do not need another screen; they need a screen they can trust.

Batteries: From Halide Electrolyte to Solid-State Cell

If LMR cathodes are the near-term lever, solid-state batteries with lithium-metal anodes are the long-term leap. Earth-abundant halide solid electrolytes — Li-Zr-Cl and Li-Fe-Cl systems — target >10 mS/cm ionic conductivity, stability versus lithium metal, and compatibility with mechanochemical synthesis. Industry analysts project the solid-state battery market to grow by an order of magnitude or more over the next decade.





> A raw universal machine-learning interatomic potential can underestimate a lithium migration barrier by 60%, shifting conductivity by 5,000×; the runtime correction and Lean 4 proofs bound that error before the next handoff.

The discovery challenge is that raw uMLIPs systematically underestimate Li^+ migration barriers by 60% or more because transition states are under-coordinated relative to bulk training data. A 60% barrier error changes predicted room-temperature conductivity by roughly 5,000×. The environment error field corrects this at uMLIP speed.

The partner chain begins at UC Berkeley with Gerbrand Ceder's group and the Materials Project, which published foundational DFT studies of Li_3YCl_6 and Li_2ZrCl_6[8]. The University of Münster's Janek and Zeier groups add world-leading halide solid-electrolyte synthesis and lithium-metal interface analytics via XPS and cryo-TEM[9]. Argonne National Laboratory contributes operando X-ray absorption spectroscopy and the Materials Engineering Research Facility for kilogram-scale scale-up[10].

Industrial translation runs through Solid Power, which produces sulfide and halide solid electrolytes and has built automotive-scale cells validated by BMW[11]. Factorial Energy, whose quasi-solid-state cells have achieved 391 Wh/kg with Mercedes EQS validation, offers another integration path[12].

> The five target chains — batteries, solid-state electrolytes, carbon removal, ammonia, and solar — map onto gigatonne-scale climate opportunities.

The cross-cutting partner here is again Battery500. The halide-screening results and the consortium's pouch-cell fabrication create a single loop: computational composition → mechanochemical synthesis → impedance spectroscopy → full-cell cycling. With correction, each handoff carries a bounded error budget.

Carbon Removal: From MOF Linker to Gigafactory

Direct air capture is the only carbon-removal technology that can address legacy emissions regardless of source sector. The IPCC estimates cumulative removal needs of 100–1,000 $GtCO_2$ by 2100, with annual rates approaching 10 $GtCO_2$/year by mid-century[13]. MOFs are the most tunable sorbent class, but the gap between a predicted stable framework and a deployed filter is wide: synthesis costs must fall below $50/kg, humidity stability must survive thousands of cycles, and working capacity must exceed 2 mmol/g at 400 ppm CO_2.

The partner chain starts with reticular chemistry. UC Berkeley's Omar Yaghi and Jeffrey Long offer large MOF/COF synthesis libraries and cooperative-binding design. Omar Farha's group at Northwestern brings MOF-808 and Zr-based synthesis expertise, plus DOE programs focused on DAC MOF stability.

The role is to correct hydrolysis-barrier errors and flag unsupported frameworks. Water attack on metal-linker bonds proceeds through under-coordinated transition states — exactly where uMLIPs soften the potential energy surface. Corrected hydrolysis barriers filter frameworks by real-world humidity stability, not idealized gas-phase energy. Amine-functionalized MOF-808 has reached 1.2 mmol/g at 400 ppm and 50% relative humidity[14].

> 899 build-locked Lean 4 theorems let Lupine flag unsupported frameworks or metastable phases before they consume experimental time.

Manufacturing scale runs through BASF, the first commercial MOF producer at several hundred tons per year and exclusive manufacturer of Svante's CALF-20 sorbent[15]. Svante has deployed MOF-based capture at pilot scale and is planning multi-million-tonne-per-year capture facilities. Climeworks, the world's largest DAC deployer with its Mammoth plant in Iceland, offers field validation.

Industrial Decarbonization: From Catalyst Prediction to Ammonia Market

Haber-Bosch ammonia synthesis emits approximately 450 $MtCO_2$/year and consumes 1–2% of global energy[16]. Electrochemical synthesis powered by renewable electricity could eliminate those emissions. Lithium-mediated nitrogen reduction has achieved >90% Faradaic efficiency, but energy efficiency is stuck near 28% because lithium plating imposes a >3 V overpotential penalty[17]. Lupine targets non-lithium mediators — Ca, Mg, Al, Na — with a goal of >60% energy efficiency and >300 mA/cm^2 partial current density[18].

The most rigorous verification partner is the Technical University of Denmark's Villum Center for Sustainable Fuels, led by Ib Chorkendorff. DTU published the landmark Ca-mediated nitrogen reduction result in 2024, achieving 40% Faradaic efficiency, and operates a rigorous ammonia verification protocol[19]. Chorkendorff has described non-lithium mediators as the field's highest priority.

Stanford's SUNCAT Center, co-directed by Jens Nørskov, adds computational catalysis and operando X-ray absorption spectroscopy at SLAC. The loop is tight: Lupine predicts compositions, SUNCAT screens them at the electronic-structure level, and DTU validates the winners.

> Across 36 model-material combinations, the environment error field achieves r = 0.906 in zero-parameter blind prediction of surface energies.

Commercial engineering comes through Jupiter Ionics, which licensed Douglas MacFarlane's lithium-mediated technology and is developing modular MSA Cell technology after a $9 million Series A. ARPA-E's REFUEL program provides funding and metrics: >60% energy efficiency, >300 mA/cm^2, and >90% Faradaic efficiency[18:1].

Downstream, Caltech's Karthish Manthiram group contributes gas-diffusion-electrode engineering, Nitricity offers distributed-fertilizer piloting, and CF Industries — the world's largest ammonia manufacturer — provides the eventual industrial exit.

Solar: From Lead-Free Absorber to Tandem Module

Lead-based halide perovskites have exceeded 26% single-junction efficiency and ~34.6% in perovskite-silicon tandems, but lead toxicity and impending EU restrictions threaten terawatt-scale deployment[20]. Lead-free tin perovskites are the leading alternative, yet Sn^{2+} oxidizes within minutes to hours. The target is >20% certified power conversion efficiency with >25-year operational stability.

The key descriptor is Sn vacancy formation energy, which determines oxidation susceptibility. uMLIPs systematically underestimate this energy because vacancies create under-coordinated Sn neighbors that bulk-trained models predict as too stable. The correction recovers accurate vacancy energetics, while provable boundaries flag metastable phases that convex-hull screening would discard.

The partner chain begins with NREL, which provides certified efficiency measurements, slot-die and roll-to-roll fabrication, and ISOS stability protocols. The University of Queensland's Lianzhou Wang group holds the certified world record for tin halide perovskites at 16.65% and has demonstrated 1,500-hour stability without encapsulation[21]. Tandem PV Inc., which operates a pilot factory in San Jose, integrates perovskite top cells with silicon bottom cells. Northwestern's Mercouri Kanatzidis, who pioneered $CsSnI_3$ and solid-state halide perovskite solar cells, offers synthesis and panoramic methodology for discovering new absorber families.

> A single NREL master CRADA and one UC Berkeley campus partnership give Lupine coverage across four of five target chains, while coordinated ARPA-E submissions provide non-dilutive validation.

The scale-up handoff runs through Oxford Photovoltaics, which began commercial tandem shipments in 2024 and holds a 24%+ module efficiency, and through Swift Solar, whose vapor-deposition manufacturing is especially compatible with oxygen-sensitive tin perovskites.

Cross-Cutting Architecture: NREL, UC Berkeley, and ARPA-E

Each target has its own partner chain, but three institutions create economies of scope.

NREL is the highest-value cross-cutting partner. It spans four of the five targets — cathodes, halide solid electrolytes, ammonia catalysts, and perovskites — and offers certified testing, lifecycle analysis, techno-economic analysis, and a single point of access to DOE funding networks. A master CRADA with NREL replaces multiple one-off agreements.

> The NIST Materials Genome Initiative estimates $123 billion–$270 billion in annual value from better materials innovation infrastructure, while ARPA-E's portfolio has catalyzed billions in follow-on funding.

UC Berkeley offers a complementary two-group engagement: the Ceder group for battery materials and the Yaghi/Long groups for MOF direct air capture. A campus-wide partnership gives Lupine access to synthesis, characterization, and autonomous discovery infrastructure across three targets from one institutional relationship.

ARPA-E should be approached with a portfolio strategy rather than individual proposals. Coordinated submissions across REFUEL (ammonia), IONICS (halide solid electrolytes), and OPEN (MOF DAC) position Lupine as a multi-target computational materials company. The return is non-dilutive capital and validation credibility. ARPA-E's portfolio has catalyzed billions of dollars in private follow-on funding[22], and the NIST Materials Genome Initiative analysis estimates $123 billion–$270 billion in annual value from improved materials innovation infrastructure[23].

The Verification Layer as Partnership Enabler

Partners validate the science; verification makes it trustworthy at scale. Every organization named above has been burned by computational predictions that looked convincing in a paper but failed in the lab. The 0.2% GNoME synthesis rate and the A-Lab disordered-phase episode are warnings that experimental budgets are finite and false positives are expensive.

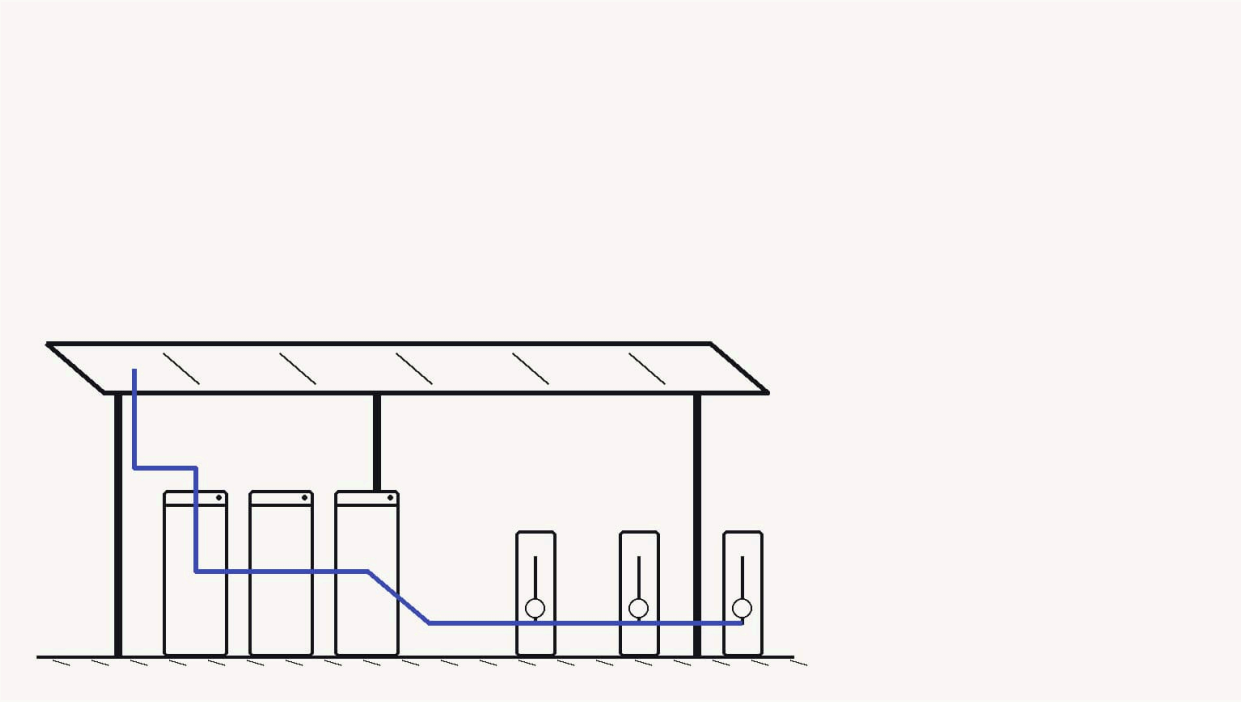
The environment error field, runtime correction, and formal verification address this directly. The field achieves r=0.906 blind prediction of never-fitted surface energies across 36 (model, material) combinations with zero adjustable parameters. Runtime correction adds 15.6% overhead in Python and will drop below 1% in a compiled LAMMPS overlay, keeping corrected uMLIPs many orders of magnitude faster than DFT. 899 build-locked Lean 4 theorems provide machine-checked guarantees. Where correction cannot be applied, impossibility proofs identify the boundary rather than silently extrapolating.

For partners, handoffs carry documentation. A cathode candidate comes with corrected migration barriers and a theorem certificate of the ranking claim. A MOF candidate comes with a hydrolysis-barrier correction and a proof of whether its coordination environment lies inside or outside the field's domain. An ammonia catalyst comes with corrected N_2 dissociation barriers and a flag if

scaling-relation breaking requires higher-level treatment. Experimentalists still synthesize and measure; Lupine removes the guesswork about which candidates deserve the work.

> In a discovery ecosystem that must scale from thousands to millions of validated materials per year, the scarce resource is the trust that turns a predicted crystal into a commercial cell.

In a discovery ecosystem that must scale from thousands to millions of validated materials per year, the scarce resource is not prediction volume. It is the trust that turns a predicted crystal into a commercial cell.



FOOTNOTES

1. A. Merchant *et al.*, “Scaling deep learning for materials discovery,” *Nature* 624, 80–85 (2023). https://doi.org/10.1038/s41586-023-06735-9 ↩

2. N. J. Szymanski *et al.*, “An autonomous laboratory for the accelerated synthesis of novel materials,” *Nature* 624, 86–91 (2023); J. Leeman *et al.*, “Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis,” *PRX Energy* 3, 011002 (2024). https://doi.org/10.1038/s41586-023-06734-7; https://doi.org/10.1103/PRXEnergy.3.011002 ↩

3. POSCO Future M, “POSCO Future M to lead entry-level and standard EV markets with LMR cathode materials,” press release, June 2025. ↩

4. General Motors, “Why LMR batteries will change the outlook for the EV market,” GM Newsroom, May 2025; LG Energy Solution / GM, Battery Innovation of the Year, The Battery Show North America, October 2025. ↩ ↩

5. TexPower EV Technologies, company website and press releases, 2024–2025; A. Manthiram, “A reflection on lithium-ion battery cathode chemistry,” *Nature Communications* 11, 1550 (2020). ↩ ↩

6. Forge Nano, “US-made ultrahigh-energy cathodes will enable low-cost electric vehicle batteries,” press release, January 2024. ↩

7. U.S. Department of Energy / PNNL, “Battery500: Progress Update,” 2020; PNNL, “Powering Up to Address Challenges in Energy Storage,” March 2024. ↩

8. G. Ceder *et al.*, foundational DFT studies of Li_3YCl_6 and Li_2ZrCl_6 halide solid electrolytes; Materials Project. ↩

9. J. Janek and W. G. Zeier groups, University of Münster, halide solid-electrolyte synthesis and interface analytics. ↩

10. Argonne National Laboratory, Materials Engineering Research Facility and Advanced Photon Source capabilities. ↩

11. BMW Group / Solid Power, joint development agreement and solid-state battery cell validation announcements, 2021–2025. ↩

12. Factorial Energy / Mercedes-Benz, solid-state battery cell development and EQS road-test announcements, 2023–2025. ↩

13. IPCC, *Climate Change 2022: Mitigation of Climate Change*, Working Group III Contribution to AR6, Cambridge University Press, 2022. ↩

14. X. Chen *et al.*, "Computational Screening of Amino-Functionalized Molecules for Direct Air Capture of CO_2," *J. Phys. Chem. A* (2026). ↩

15. BASF, "BASF becomes first company to successfully produce metal-organic frameworks on a commercial scale for carbon capture," press release, October 2023. ↩

16. Royal Society, *Ammonia: Zero-Carbon Fertiliser, Fuel and Energy Store*, policy briefing, 2020; International Energy Agency, *Ammonia Technology Roadmap*, IEA, 2021. ↩

17. W. Chang *et al.*, "Lithium-mediated nitrogen reduction to ammonia via the solvation-enabled N_2 activation," *Nature Catalysis* 7, 342–352 (2024). ↩

18. ARPA-E REFUEL program, U.S. Department of Energy, program targets. ↩ ↩

19. X. Fu *et al.*, "Calcium-mediated nitrogen reduction for electrochemical ammonia synthesis," *Nature Materials* 23, 101–107 (2024). https://doi.org/10.1038/s41563-023-01702-1 ↩

20. NREL, "Best Research-Cell Efficiency Chart," National Renewable Energy Laboratory, updated 2025. ↩

21. P. Chen *et al.*, "2D/3D tin-halide perovskite solar cell with certified 16.65% efficiency," *Nature Nanotechnology* 20, 742–749 (2025). ↩

22. ARPA-E, "ARPA-E Impact," U.S. Department of Energy, 2023. ↩

23. NIST / RTI International, *Economic Analysis of the Materials Genome Initiative*, National Institute of Standards and Technology, 2023. ↩

Lupine Science

2026-07-09

Investing in the Trust Layer

https://lupine.science/articles/investing-in-the-trust-layer/

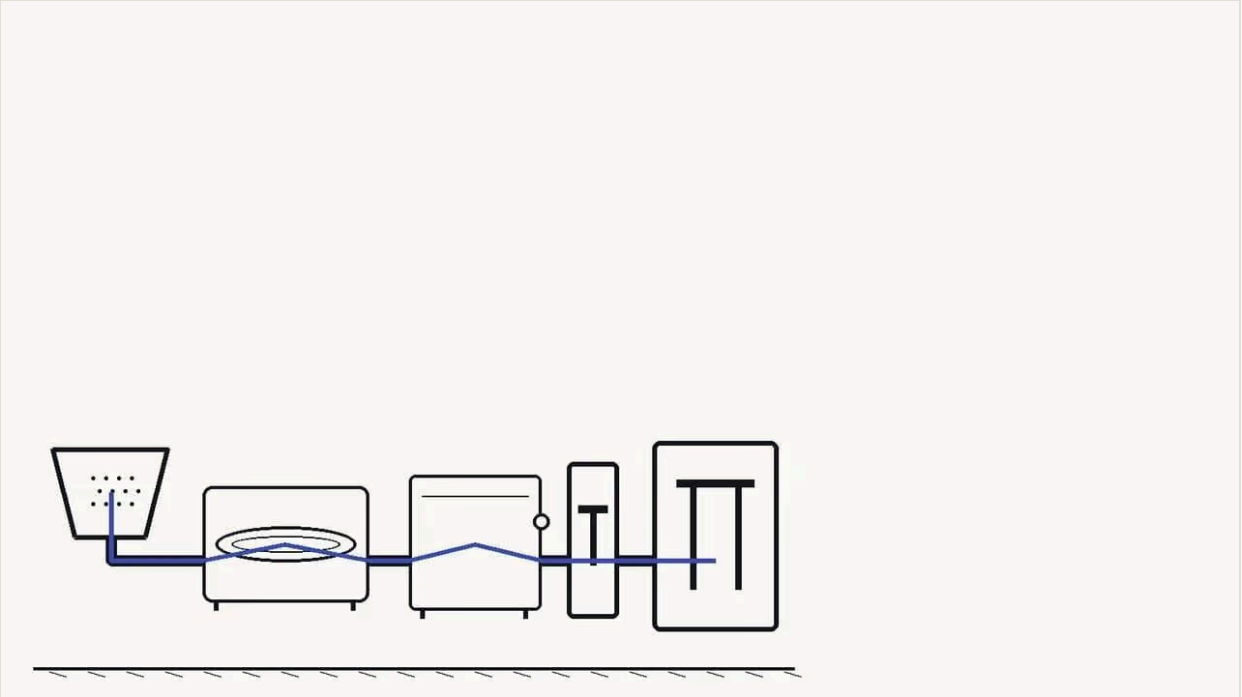
CLIMATE PARTNERSHIPS PROOF PACK

RESEARCH NOTE

Investing in the Trust Layer

The Lupine Science correction-and-verification layer as investable climate infrastructure

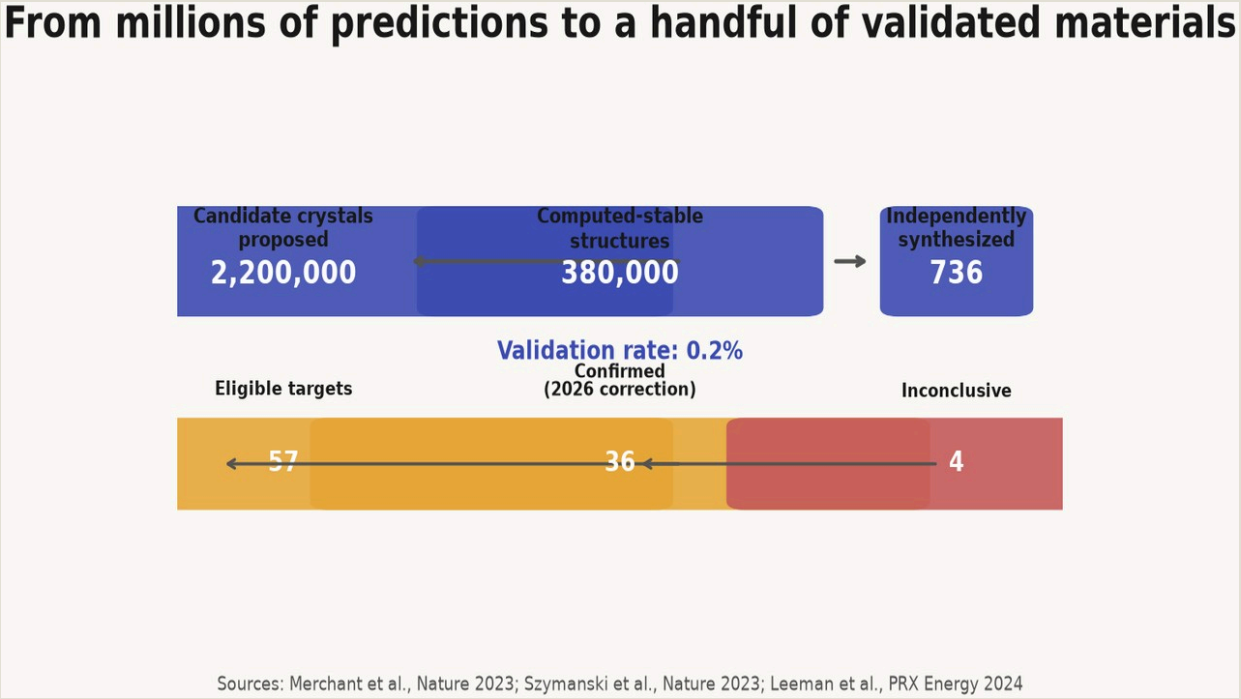
July 9, 2026 · DRAFT



A narrow pilot pipeline from powder hopper through sintering furnace to a pressure-test fixture — one inspection gate admits only measured samples to the pilot press.

Computational materials discovery is no longer short on imagination. Structure generators such as Google DeepMind's GNoME have proposed 2.2 million candidate crystals[1]; universal machine-learned interatomic potentials (uMLIPs) such as CHGNet and MACE evaluate them at roughly 10^{-4} seconds per

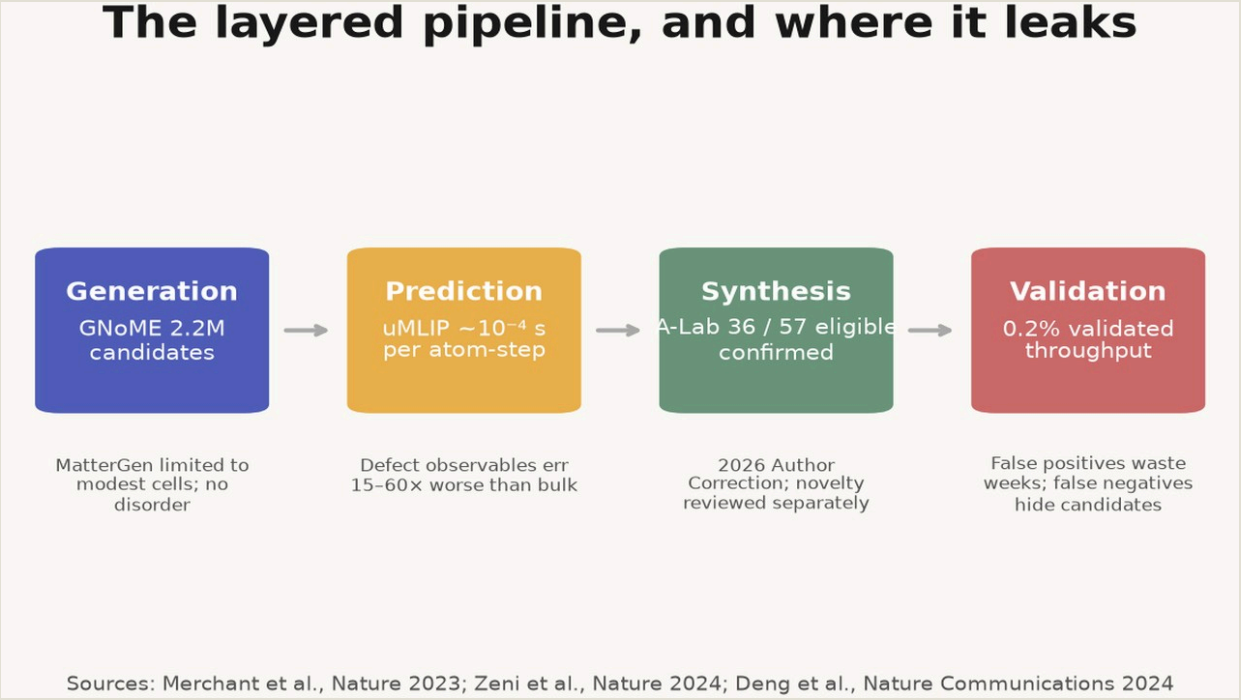
atom-step; and autonomous labs such as the A-Lab at Lawrence Berkeley National Laboratory attempt synthesis around the clock. The scarce resource is no longer the ability to invent or simulate. It is the ability to believe the result.



> GNoME proposed 2.2 million candidate crystals, but only 736 had been independently synthesized by late 2023—a 0.2% validation rate that exposes the trust bottleneck in computational materials discovery.

Belief is scarce because the pipeline is lossy. GNoME reported 380,000 computed-stable structures, yet only 736 had been independently synthesized by late 2023 — a 0.2% validation rate[1:1]. The A-Lab Author Correction (Nature 650:E1, 2026) records 36 confirmed of 57 eligible targets, with 4 inconclusive and one compound removed for training-data leakage; a separate independent review disputed the novelty classification of many targets[2]. These are not failures of ambition. They are structural symptoms of a field that generates faster than it verifies.

For climate materials, the cost of that asymmetry is measured in gigatonnes and years. Batteries alone are directly linked to roughly 20% of the CO_2 reductions required by 2030 and indirectly to another 40%[3]. Each false positive sends a team into weeks of wasted synthesis; each false negative buries a candidate that might have reached a cell test. Speed without accuracy wastes experiments; accuracy without speed misses the deployment window. The winning infrastructure is the layer that de-risks the lab queue.



> Each stage of the discovery pipeline—generation, prediction, synthesis, validation—introduces systematic errors that cascade downstream, turning high activity into low validated throughput.

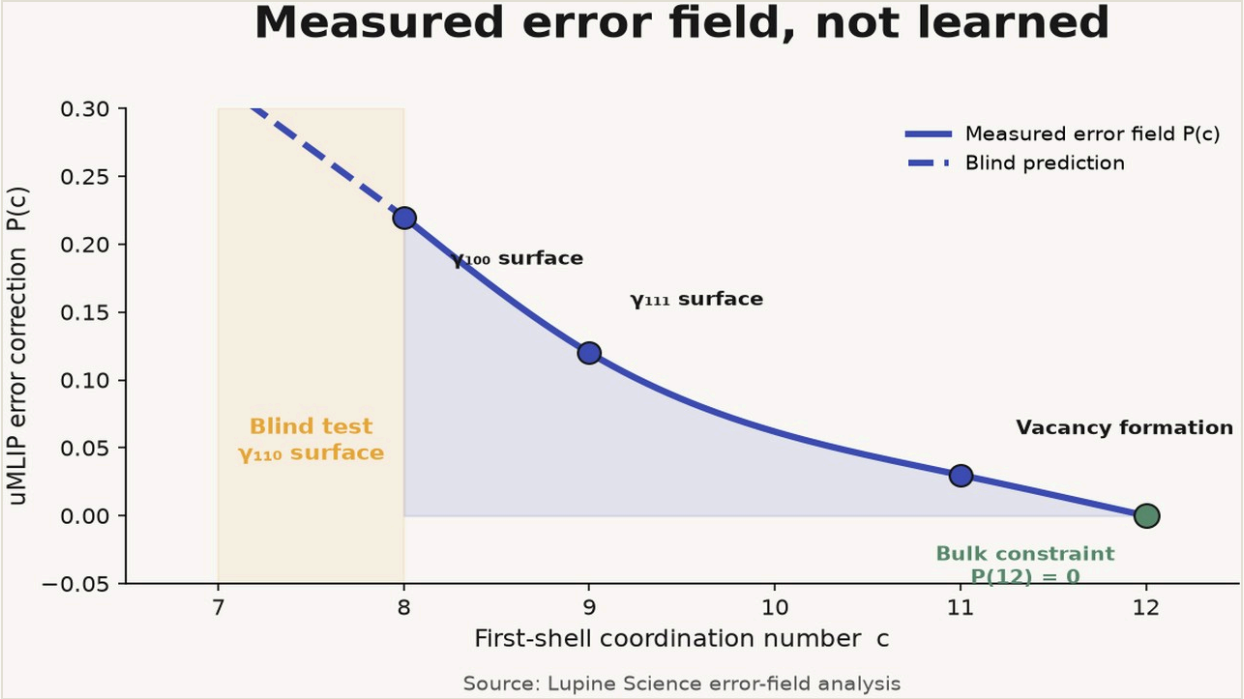
The correction-and-verification layer closes that gap.

The layered pipeline, and where it leaks

Computational discovery can be drawn as four stages: generation, prediction, synthesis, and validation. Each stage has dominant platforms; each produces errors that propagate downstream.

At the generation stage, GNoME and recent generators such as Microsoft's MatterGen produce candidate structures at machine speed. MatterGen reports improved generative-design metrics for inorganic materials, but its outputs are limited to modest unit-cell sizes and do not treat disorder or compositional complexity[4]. GNoME's 2.2 million predictions excluded polymers, glasses, metal-organic frameworks, heterostructures, and composites — the very material classes most relevant to climate technology[1:2].

At the prediction stage, uMLIPs supply near-DFT energies for the generated structures. They are accurate on bulk properties but systematically soften the potential energy surface away from equilibrium. Across multiple uMLIPs and materials, defect-family observables — surfaces, vacancies, stacking faults — err 15–60× worse than bulk observables[5]. The error is not random: it is a smooth function of local atomic environment, concentrated where coordination deviates from the bulk.



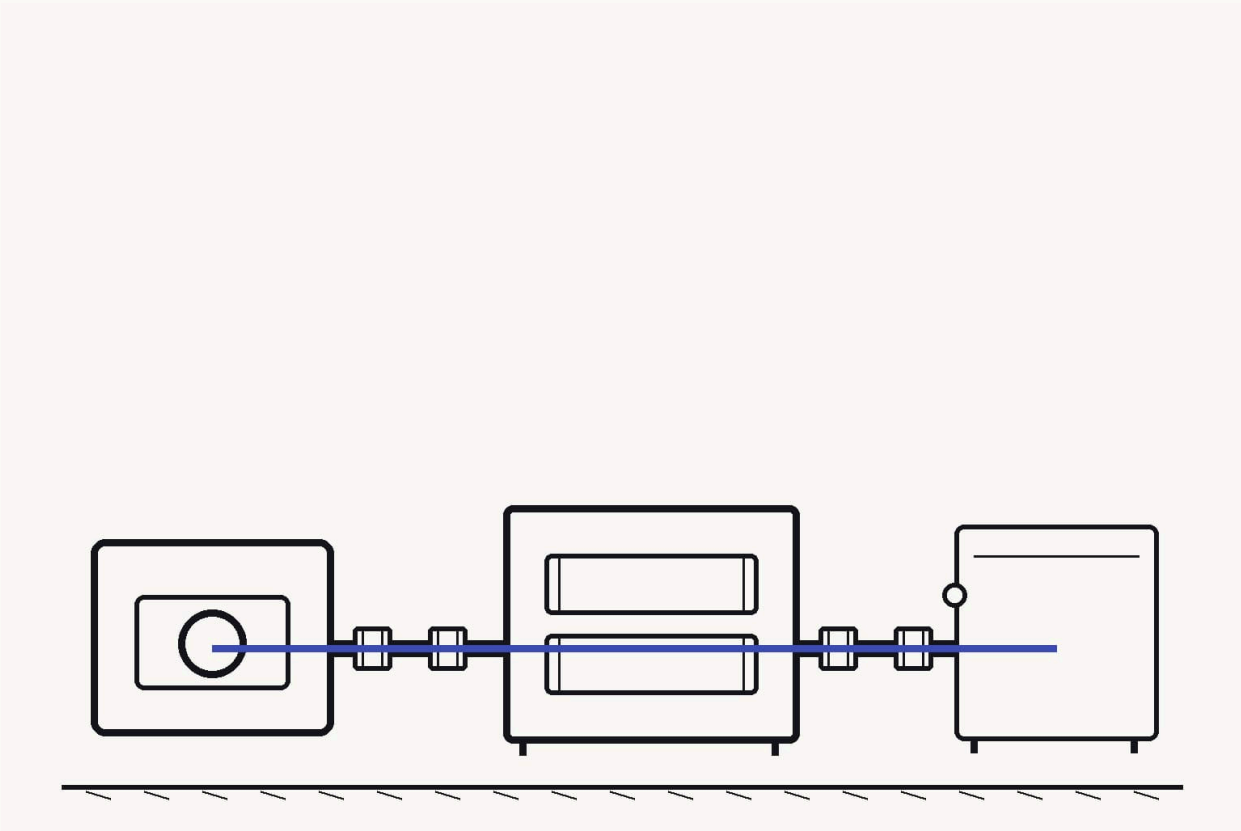
> For fcc metals, uMLIP error is modeled as a cubic-spline field over first-shell coordination, anchored by three standard observables and constrained to zero at bulk coordination 12.

At the synthesis stage, A-Lab and manual labs attempt to make what upstream recommended. Without a characterization of each candidate's failure modes, the queue is filtered by intuition and compute budget rather than by proof. The corrected A-Lab record is 36 confirmed of 57 eligible targets, with 4 inconclusive and one compound removed; independent novelty critiques are a separate question. High activity is not the same as validated throughput.

The correction-and-verification layer does not compete with any of these layers. It sits between prediction and synthesis: it measures the systematic error of the predictor, corrects it at runtime, and proves which claims can be trusted. The value proposition is partnership, not displacement. MatterGen and GNoME generate more candidates as they improve; the market grows because every additional candidate must be checked.

Three pillars of defensibility

The trust layer is defensible only if it is simultaneously accurate, provable, and deployable. The architecture rests on three pillars that are individually difficult and, to current knowledge, collectively unique in materials science.



> Across 36 model-material combinations, the measured error field predicts the never-fitted γ_{110} surface-energy error with Pearson r = 0.906 (95% CI [0.82, 0.96], p = 10^{-4}) and zero adjustable parameters.

Measured error field, not learned. Rather than training a delta-ML model to fit discrepancies per system, the method treats the uMLIP error as a physical field over local atomic environments. For face-centered-cubic metals, the field is a cubic spline over first-shell coordination number, fixed by three anchor observables — γ_{100}, γ_{111}, and the vacancy formation energy — with the bulk value constrained to zero at coordination 12. A fourth observable, γ_{110}, is held

back as a blind test. Across 36 independent (model, material) combinations, the field predicts the never-fitted γ_{110} error with Pearson r = 0.906 (material-clustered 95% CI [0.82, 0.96], p = 10^{-4}) and zero adjustable parameters.

The distinction from delta-ML and fine-tuning matters commercially. Delta-ML requires retraining for every new material, negating the speed advantage of universal potentials[6]. Fine-tuning requires curated target-system data that does not exist for unexplored composition spaces. The correction is measured from three standard observables and transfers within a structure family; no per-system training is required. The improvement is concrete: on the Ni(110) blind facet, relative error falls from 9.7% to 1.5%; on Cu(110), from 28.0% to 13.7%.

Machine-checked proof. The quantitative claims are sealed as Lean 4 theorems over integer-scaled data carrying SHA-256 provenance. The formalization currently comprises 899 build-locked theorems and zero `sorry` proofs — Lean's escape hatch for unproven claims. The count is generated from the Lean source rather than transcribed into this article. The theorems cover ordering inequalities, isotonic correction bounds, and impossibility results where no monotone correction can recover the reference ranking.

This is a categorical difference from statistical validation. Google DeepMind and Microsoft can scale compute indefinitely, but they cannot produce machine-checked proofs of why a prediction should be believed. The Lean kernel once rejected a claim that had survived statistical filtering, reducing a reported success count from 27/36 to 26/36 at integer precision. That episode is the approach in miniature: a trust layer must be able to say no to itself.

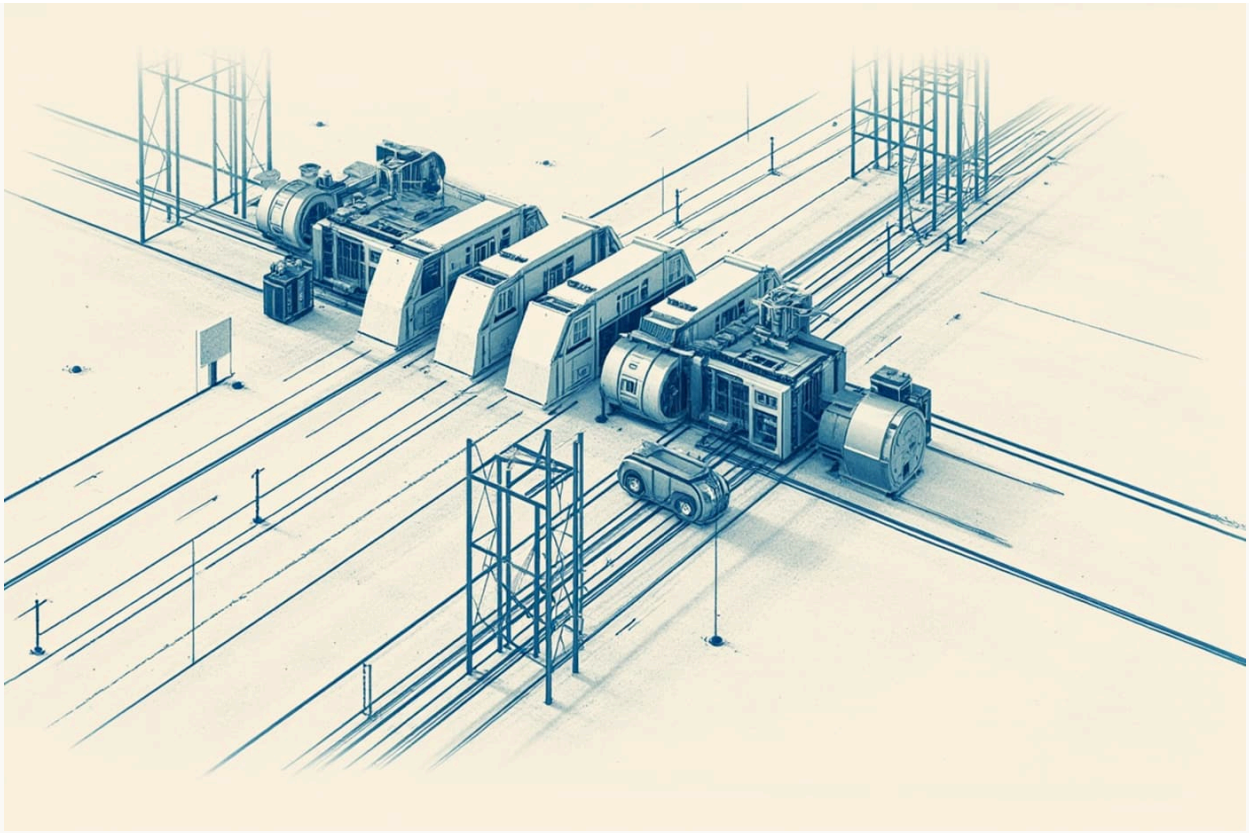
> The correction layer deploys beside existing CHGNet or MACE calculators as a LAMMPS overlay, adding 15.6% overhead in Python and cutting relative error on blind facets by up to an order of magnitude.

Runtime compatibility. The correction term is shaped like an EAM embedding function — a sum over atoms of a function of local coordination — so it deploys as a LAMMPS overlay pair style beside a live CHGNet or MACE calculator. Measured wall-time overhead in Python is 15.6%; a compiled C/C++ or CUDA implementation is expected to drop this below 1%. Even at 15.6%, corrected uMLIPs remain many orders of magnitude faster than DFT for structurally complex cells. Users keep their existing simulators; the trust layer corrects them without retraining.

Why the moat deepens with use

Unlike data-scale moats that erode as competitors scrape the same training corpora, the trust layer's position strengthens with each screening campaign. Every measured error field adds to a reference database of systematic biases organized by (model, material, structure family). Each impossibility proof sharpens the boundary of applicability. Each experimental validation tightens the feedback loop between computation and synthesis.

Transfer is the mechanism. The Li–Zr–Cl field informs Li–Fe–Cl halide solid-electrolyte screening; the fcc field informs body-centered-cubic and hexagonal-close-packed extensions planned for Phase 0; the cobalt-free cathode field transfers to layered oxide and spinel dopant spaces. Theorem families expand in parallel: extending the formalization from fcc to bcc, hcp, and layered structures adds machine-checked protections for the material classes that dominate battery and catalyst targets.



> A master CRADA with NREL and tier-1 experimental collaborators close the loop between prediction and synthesis, turning every validated measurement into a more trustworthy next prediction.

The partnership architecture reinforces the flywheel. A master CRADA with NREL, which spans four of the five priority targets, would provide certified testing infrastructure and a single point of access to DOE funding networks. Tier-1 experimental collaborators — the Manthiram Laboratory and TexPower for cobalt-free cathodes, the Ceder group at UC Berkeley and LBNL for halide electrolytes, the University of Münster for Li-metal interface analytics, DTU's Chorkendorff group for ammonia verification, and the Yaghi and Long groups at UC Berkeley for MOF direct air capture — supply the synthesis and characterization data that close the loop. Each validated measurement makes the next prediction more trustworthy.

The economic case

The macroeconomic argument is not speculative. The NIST Materials Genome Initiative economic analysis estimates that improved materials innovation infrastructure would deliver $123 billion–$270 billion in annual value to U.S. industry[7]. ARPA-E's portfolio has catalyzed billions of dollars in private follow-on funding and substantial IPO or acquisition value[8]. Self-driving laboratories funded by ARPA-E and CHIPS Act investments could compress materials-development timelines from 10–20 years to 2–5 years.

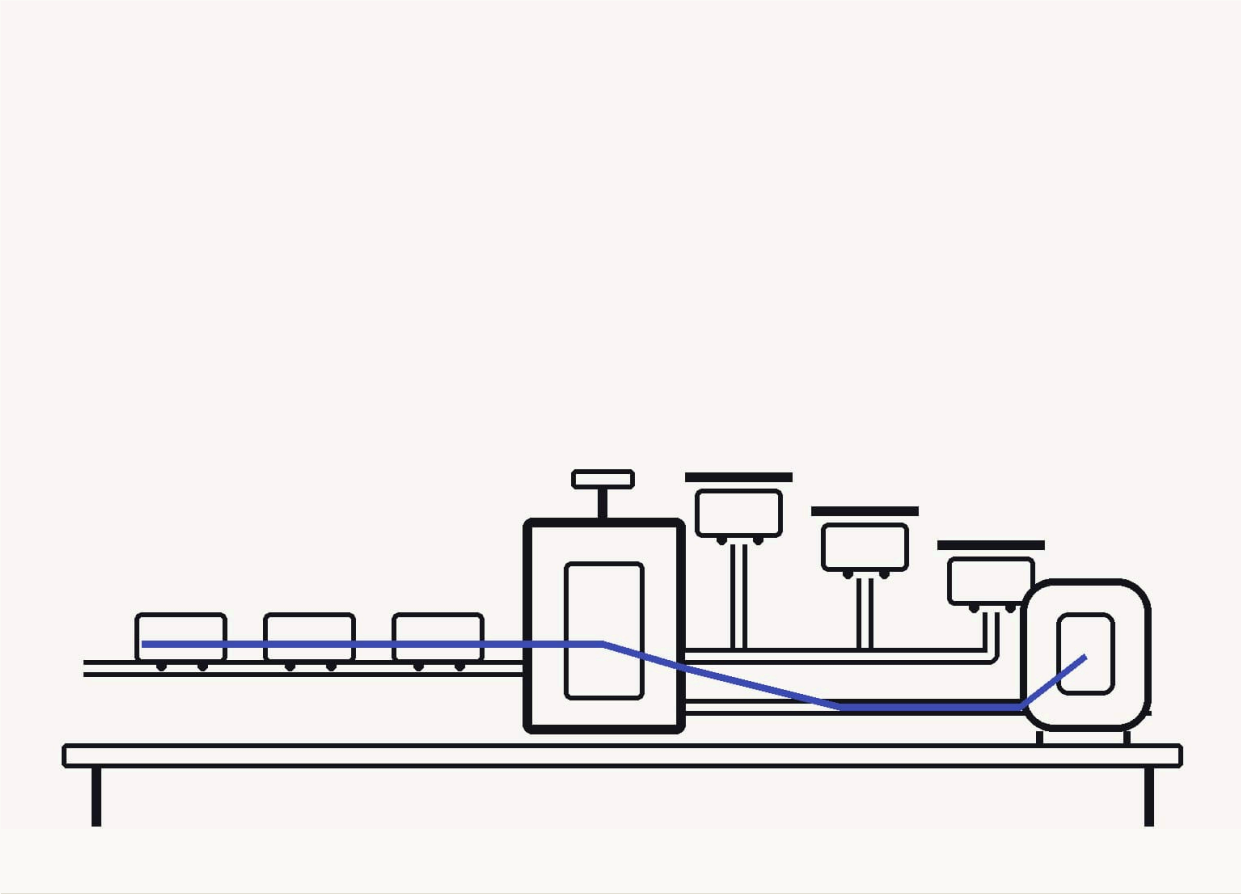
Against that backdrop, the cumulative 36-month budget is approximately $3.2 million. The leverage ratio justifies the expenditure even if only one of the five priority targets — cobalt-free cathodes, halide solid electrolytes, MOF sorbents, electrochemical ammonia catalysts, or lead-free perovskites — yields a commercially viable material. Each target addresses a large and growing clean-energy market, and the combined climate impact potential is substantial.

> Batteries are directly tied to roughly 20% of the CO_2 reductions needed by 2030 and indirectly to another 40%, so false positives and false negatives in materials screening carry climate-scale consequences.

The investment thesis is therefore not a bet on a single material. It is a bet on the infrastructure that makes every material-discovery program more capital-efficient. Corrected screening campaigns become economically possible where brute-force DFT is not. Verified predictions reduce the false-positive tax that currently consumes lab time. Partners across the stack — generators, databases, autonomous labs, OEMs, national labs — capture more value from their own data because the trust layer exists.

Risk and honesty

A proof-first voice requires naming what is still open. The fcc error field is established; extension to bcc, hcp, and layered structures is the critical path item in Phase 0, with a go/no-go gate at Month 3. A second-shell correction may be needed for systems where first-shell coordination alone does not resolve the error geometry. Industrial validation partners are targeted but not yet contracted; until synthesis data arrive, the transfer claims remain computational.



> The fcc error field is established; the Phase 0 critical path extends it to bcc, hcp, and layered structures, with a Month 3 go/no-go gate and honest caveats about second-shell corrections and pending validation partners.

These are engineering risks, not conceptual ones. The conceptual question — whether uMLIP errors are structured and measurable — has been answered affirmatively for fcc metals. The remaining work is to extend the field's jurisdiction and to prove, with each new structure family, where correction applies and where it does not.

The trust layer for a real-world Replicator

The long-term vision is a discovery pipeline in which AI-designed matter can be made, measured, and believed. Structure generators propose; the trust layer corrects and verifies; experimental partners synthesize; the result feeds back into the field database and the theorem library. The loop accelerates not because errors are hidden but because they are characterized.

> The ~$3.2 million, 36-month budget is tiny against the NIST-estimated $123 billion–$270 billion annual value of improved materials innovation infrastructure, so even one successful target among five justifies the spend.

The climate window is unforgiving. A material discovered in 2035 may miss the 2040 deployment window, and the cumulative emissions difference between deployment in 2035 versus 2045 is measured in tens of gigatonnes. The world does not need more predicted crystals. It needs a smaller, defensible set of predictions that laboratories can act on.

> The trust layer closes the loop between AI-generated candidates, corrected simulations, and experimental validation—turning an abundance of predicted materials into a scarce, defensible set that labs can act on before the climate window closes.

That is the trust layer. It is the infrastructure that converts the abundance of generated materials into the scarcity of validated ones — and it is the investment that makes the rest of the climate-materials pipeline economically rational.

> For climate investors: the full slide presentation, *The Trust Layer for Climate-Critical Materials*, walks through the capital ramp, the five targets, the moat, and the ask.
>
> **Open the presentation →**

FOOTNOTES

---

1. A. Merchant *et al.*, "Scaling deep learning for materials discovery," *Nature* 624, 80–85 (2023). https://doi.org/10.1038/s41586-023-06735-9 ↩ ↩ ↩

2. N. J. Szymanski *et al.*, "An autonomous laboratory for the accelerated synthesis of novel materials," *Nature* 624, 86–91 (2023); J. Leeman *et al.*, "Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis," *PRX Energy* 3, 011002 (2024). https://doi.org/10.1038/s41586-023-06734-7; https://doi.org/10.1103/PRXEnergy.3.011002 ↩

3. International Energy Agency, *Global EV Outlook 2024*, IEA, 2024; IEA, *Energy Technology Perspectives 2023*, IEA, 2023. ↩

4. C. Zeni *et al.*, "A generative model for inorganic materials design," *Nature* 633, 812–818 (2024). https://doi.org/10.1038/s41586-024-07816-6 ↩

5. Y. Deng *et al.*, "Benchmarking machine learning potentials for the discovery of new materials," *Nature Communications* 15, 1673 (2024). https://doi.org/10.1038/s41467-024-45870-1 ↩

6. L. Hu *et al.*, “Machine learning Δ-ML potentials for materials: Progress, challenges, and outlook,” *Current Opinion in Solid State and Materials Science* 26, 100996 (2022). https://doi.org/10.1016/j.cossms.2022.100996 ↩

7. NIST / RTI International, *Economic Analysis of the Materials Genome Initiative*, National Institute of Standards and Technology, 2023. ↩

8. ARPA-E, “ARPA-E Impact,” U.S. Department of Energy, 2023. ↩