Matter - AI A Multimodal AI Framework for Accelerated Materials Discovery in Physical Sciences

Prudvi Saisaran Ponduru

Doctor of Computer Science Researcher
Department of Computer Science, University of the Potomac.

Abstract:
Materials discovery in physical sciences is being transformed by the convergence of large language models, graph neural networks, foundation atomistic models, high-throughput databases, physics-based simulation, autonomous laboratories, and sustainability-aware decision making. However, many current pipelines remain fragmented: literature mining is separated from crystal-structure learning, machine learning prediction is separated from physics validation, and performance optimization is often separated from cost, toxicity, energy, recyclability, and environmental impact. This paper proposes Matter AI, a multimodal artificial intelligence framework for accelerated and responsible materials discovery. The proposed framework integrates literature mining using large language models, federated collection of materials databases, chemical and crystal-structure feature extraction, machine learning and graph neural network property prediction, physics-informed validation, multi-objective candidate ranking, human or robotic experimental validation, and sustainability scoring. The paper is grounded in recent 2023 to 2026 developments, including structure-aware multimodal language models, pretrained atomistic potentials, generative crystal models, benchmark-driven stability prediction, large open density functional theory datasets, source-tracked literature extraction, and closed-loop autonomous synthesis. No fabricated experiment is reported. Instead, a rigorous framework, mathematical formulation, implementation protocol, evaluation metrics, and ablation plan are presented for future empirical deployment. The main contribution is a unified discovery architecture that treats scientific evidence, atomistic structure, physical laws, uncertainty, experimental validation, and sustainability as coupled components rather than isolated stages.

Keywords: Artificial Intelligence, Materials Discovery, Graph Neural Networks, Large Language Models, Physics Informed Machine Learning, Sustainable Materials

1. Introduction

Materials discovery is a central challenge in physical sciences because the search space formed by elements, stoichiometries, crystal symmetries, synthesis pathways, processing conditions, defects, and operating environments is far larger than what can be explored by conventional trial-and-error experimentation. A single practical material must satisfy several constraints at once, including thermodynamic stability, kinetic accessibility, functional performance, manufacturability, cost, safety, and environmental acceptability. The need for faster discovery is especially urgent in energy storage, photovoltaics, catalysis, semiconductors, thermoelectrics, structural alloys, carbon capture, and sustainable manufacturing.

Artificial intelligence has changed this landscape by enabling models to learn from high-throughput calculations, published literature, crystal structures, chemical descriptors, and experimental records. Graph neural networks can treat atoms as nodes and interatomic relations as edges, while pretrained atomistic models can approximate energy, force, stress, and finite-temperature behavior at a fraction of the cost of repeated first-principles calculations. Large language models can retrieve and structure knowledge from unstructured scientific papers. Generative models can propose new crystals under property constraints. Autonomous laboratories can close the loop between prediction and synthesis. These advances have been demonstrated by recent systems such as CHGNet, MatterSim, MatterGen, GNoME, A-Lab, Matbench Discovery, and MatterChat [1, 2, 3, 5, 6, 7, 9].

Despite this progress, the field still lacks a disciplined framework that connects all discovery stages without losing provenance, uncertainty, physical consistency, and sustainability information. A text-only model may extract a plausible synthesis route but fail to understand three-dimensional periodic geometry. A graph model may predict a promising property but ignore synthesis feasibility and lifecycle risk. A generative model may propose a stable crystal that relies on toxic, scarce, or expensive elements. A robotic laboratory may produce a sample whose structural assignment remains ambiguous. These failures are not isolated model weaknesses; they are interface failures between literature, databases, models, physics, experiment, and environmental decision making.

This paper proposes Matter AI as a multimodal framework for accelerated materials discovery in physical sciences. The framework is designed to connect eight components: literature mining using large language models, materials database collection, feature extraction from chemical and crystal structures, machine learning and graph neural network property prediction, physics-informed validation, candidate ranking, human or robotic experimental validation, and sustainability scoring. The objective is not to replace materials scientists but to create a reliable discovery system in which human expertise, physical laws, computational models, and experimental feedback are integrated in a traceable loop.

The paper makes four contributions. First, it defines a complete architecture for multimodal materials discovery that joins text, structure, simulation, experiment, and sustainability. Second, it presents a mathematical ranking formulation that combines desirability, uncertainty, stability, synthesis feasibility, novelty, and sustainability. Third, it specifies an implementation and evaluation protocol using public materials databases, modern foundation models, graph neural networks, and source-tracked literature extraction. Fourth, it discusses limitations and safeguards needed to avoid hallucinated records, unstable candidates, misleading novelty claims, and unsustainable recommendations.

2. Recent Literature and Research Gap

Recent progress from 2023 to 2026 indicates that materials artificial intelligence is shifting from isolated predictors toward foundation-scale, multimodal, and closed-loop discovery systems. The CHGNet model demonstrated that a pretrained graph neural network potential can learn from energies, forces, stresses, and magnetic moments in Materials Project trajectories, enabling charge-informed atomistic modeling for inorganic materials [2]. MatterSim expanded this direction by presenting a deep learning atomistic model across elements, temperatures from 0 K to 5000 K, and pressures up to 1000 GPa, with reported improvements in precision and data efficiency for downstream fine-tuning [3]. MACE-class foundation models further indicate that large pretrained atomistic representations are becoming reusable components for materials chemistry and simulation [4].

Generative discovery has also advanced. GNoME used graph networks and active learning to scale the search for stable inorganic crystals, reporting millions of candidate structures and hundreds of thousands of stable predictions [6]. MatterGen introduced a diffusion-based generative model for inorganic crystals and showed that generated structures could be steered toward multiple target properties, including low supply-chain-risk magnets [5]. These systems demonstrate that the discovery problem is no longer limited to screening known structures; it now includes goal-directed generation, validation, and ranking under constraints.

The literature layer is being transformed by large language models. ChatExtract and related approaches have shown that large language models can extract structured materials data from papers when prompts, retrieval, and verification are carefully designed [11, 12]. More recent source-tracked multi-stage extraction pipelines report strong performance in narrow experimental domains by requiring every extracted attribute to carry provenance and by decomposing complex composition-processing-microstructure-property relationships [13]. At the same time, unbiased extraction benchmarks show that recall can remain low for general literature mining tasks, proving that large language models should not be treated as unattended database builders [14].
Multimodal models have begun to bridge language and atomistic structure. MatterChat is an important milestone because it aligns a pretrained materials encoder with a pretrained language model through a trainable bridge, allowing material structural data and textual prompts to be processed jointly [1]. This

direction is especially relevant because materials discovery requires reasoning over both human scientific language and periodic atomic structures. However, benchmark studies such as MaScQA and MatTools show that large language models still have measurable limitations in materials knowledge, tool usage, and physics-grounded reasoning [15, 16].

Benchmarking has also matured. Matbench Discovery emphasizes prospective stability discovery instead of ordinary regression accuracy alone, showing that materials models must be evaluated by their ability to accelerate the discovery of stable candidates [9]. OMat24 contributes an open large-scale dataset of more than 110 million density functional theory calculations and pretrained models, addressing the lack of open data for foundation-scale materials learning [10]. These developments make it possible to design a framework that is measurable, reproducible, and aligned with practical discovery outcomes.

The research gap is therefore clear. Existing systems often excel in one stage: extraction, prediction, generation, simulation, or robotic synthesis. Matter AI addresses the missing integration layer. It is proposed as a governed architecture in which literature evidence, database records, crystal representations, predictive models, physical constraints, experimental outcomes, and sustainability metrics remain connected through common identifiers, provenance metadata, uncertainty estimates, and active-learning feedback.

Table 1: Core Components of Matter AI Framework

Component	**Primary Function**	**Main Output**	**Key Risk Controlled**
Literature mining using LLMs	Extract synthesis routes, properties, processing conditions, and claims from papers	Source-tracked structured records	Hallucination and missing provenance
Materials database collection	Federate Materials Project, OQMD, AFLOW, NOMAD, JARVIS, MPtrj, and OMat24 data	Canonical materials knowledge graph	Data duplication and label inconsistency
Feature extraction	Represent composition, crystal symmetry, local environment, and graph structure	Descriptors, graphs, tensors, embeddings	Loss of structural information
ML and GNN prediction	Predict target properties and uncertainty	Property distributions and calibrated intervals	Overconfident false positives
Physics-informed validation	Apply stability, charge, symmetry, phonon, elastic, and simulation checks	Physics-consistent shortlist	Violation of physical constraints
Candidate ranking	Optimize property, uncertainty, stability, novelty, synthesis feasibility, and sustainability	Pareto-ranked candidates	Single-objective bias
Human or robotic validation	Synthesize, characterize, and adjudicate outputs	Experimental evidence and feedback	Unverified novelty or ambiguous phase identification

Component	Primary Function	Main Output	Key Risk Controlled
Sustainability scoring	Assess cost, energy, toxicity, recyclability, supply risk, and environmental impact	Sustainability-aware score and report	Environmentally harmful optimization

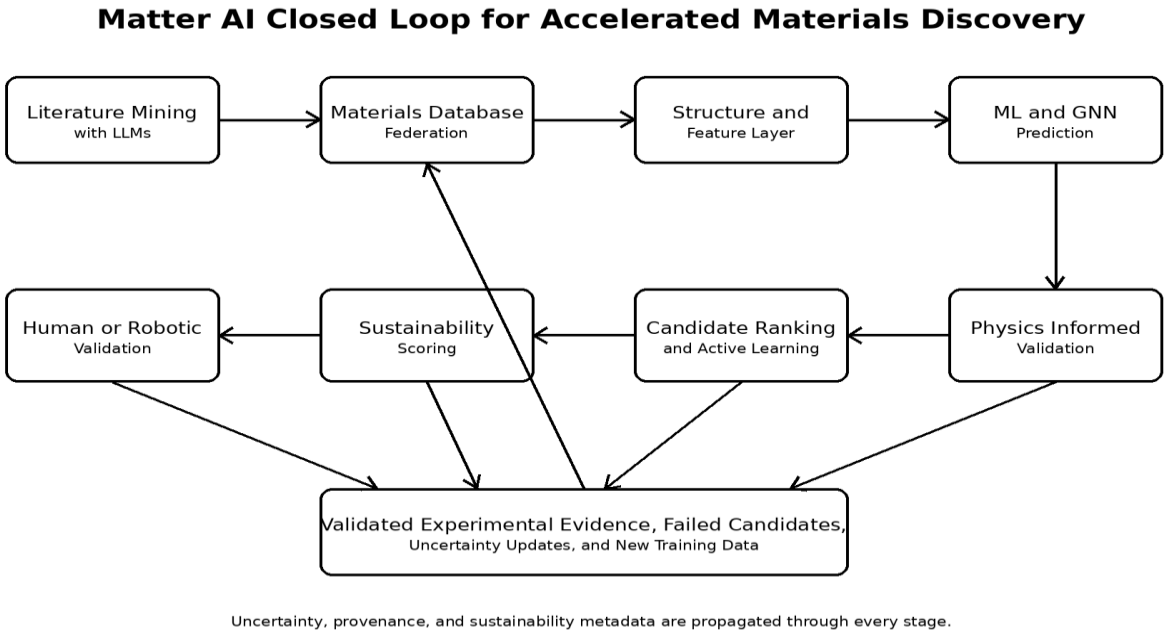
3. Proposed Matter AI Framework

Matter AI is designed as a closed-loop scientific system rather than a single model. The architecture begins with literature and database ingestion, converts heterogeneous evidence into a canonical provenance-preserving representation, builds multi-view features from chemical and crystal structures, predicts properties using machine learning and graph neural networks, validates candidates using physics-based checks, ranks candidates under multiple objectives, sends high-confidence candidates for human or robotic experimental validation, and returns all outcomes to the knowledge base. Figure 1 presents the proposed architecture.

A central design principle is that every model output must carry metadata. A property prediction should include the input structure, source database, training split status, model family, uncertainty estimate, calibration status, and physical validation status. A literature-extracted record should include the paper source, sentence or table evidence, unit, experimental condition, extraction prompt version, verifier decision, and confidence score. A sustainability score should identify whether it was estimated from composition proxies, lifecycle inventory data, toxicity databases, or full process assumptions. This metadata discipline is essential because materials discovery errors often emerge after several apparently reasonable stages are chained together.

The framework is multimodal because it combines text, tables, chemical compositions, crystal structures, graphs, atomistic trajectories, simulation outputs, spectroscopy or microscopy records, and experimental outcomes. It is physics-informed because its predictions are constrained by charge neutrality, symmetry, energy above hull, dynamical stability, elastic stability, and domain-specific physical laws. It is sustainability-aware because candidate utility is not defined only by performance but also by energy use, cost, toxicity, criticality, recyclability, and environmental impact.

Figure 1: Matter AI Closed Loop Framework



3.1 Literature Mining Using Large Language Models

The literature mining module converts unstructured papers, supplementary files, and tables into source-tracked structured records. Its role is not merely to summarize papers but to extract material names, compositions, synthesis routes, processing parameters, temperatures, durations, atmospheres, measured properties, characterization methods, and experimental outcomes. A retrieval-augmented design is preferred because the large language model should answer from relevant evidence chunks rather than from general memory. Every extracted field should be paired with a source span, unit, condition, and confidence score.
Recent work supports a schema-first extraction strategy. ChatExtract showed that conversational prompting and deliberate follow-up can improve extraction reliability [11]. Structured scientific extraction using large language models has shown that entity and relation extraction can be made more flexible when prompts are constrained by schema definitions [12]. Source-tracked multi-stage extraction has shown that complex materials records can be extracted with higher reliability when the task is decomposed and every field is verified against evidence [13]. Matter AI adopts these lessons by requiring a three-stage pipeline: retrieval, extraction, and verification.

The proposed literature mining workflow has four safeguards. First, the model is allowed to abstain when evidence is incomplete. Second, extracted numerical values are normalized into consistent units. Third, cross-field consistency is checked, such as whether a reported processing temperature is compatible with the stated synthesis method. Fourth, extracted records are not promoted into the training database until a verifier model or human reviewer has accepted the evidence. This conservative design is necessary because recent benchmarks show that large language models can achieve useful precision while still missing many valid records in unbiased scientific corpora [14].

3.2 Materials Database Collection and Provenance

The data layer is federated across public and institutional sources. Materials Project provides open computed information on known and predicted materials, including structures and properties that are widely used in machine learning benchmarks [17]. OQMD contributes density functional theory thermodynamic and structural properties for large inorganic spaces [18]. AFLOW provides high-throughput standardized calculations and a large set of derived properties [19]. NOMAD contributes parser-normalized archives and metadata-rich provenance. JARVIS contributes additional property diversity from a NIST-backed ecosystem [22]. OPTIMADE provides an interoperability standard that reduces database lock-in and enables common querying across providers [20, 21].

Matter AI stores records in a canonical knowledge graph. Each material node links to database identifiers, composition, structure, symmetry, oxidation-state estimates, computed and experimental properties, synthesis evidence, model predictions, sustainability indicators, and validation history. Each edge records the relation type, evidence source, timestamp, license status, and confidence. This knowledge graph makes it possible to ask not only which material has a high predicted property, but also why it was recommended, where the evidence originated, whether it has been synthesized, and what environmental risks are attached to it.

Data cleaning is performed at three levels. Syntactic cleaning normalizes chemical formulae, units, material identifiers, and structure files. Structural cleaning removes duplicates using reduced cells and structure matching. Semantic cleaning harmonizes labels such as band gap, formation energy, energy above hull, elastic constants, ionic conductivity, and synthesis yield while preserving the original method and conditions. The original labels are not overwritten because differences in exchange-correlation functional, pseudopotential, temperature, pressure, or measurement method can create meaningful shifts.

3.3 Feature Extraction from Chemical and Crystal Structures

Feature extraction is treated as a multi-view representation problem. Composition-only descriptors are used for rapid screening and small-data baselines. These include stoichiometric statistics, elemental properties, electronegativity, valence electron counts, orbital features, and composition-based physical heuristics. Tools such as matminer remain valuable because they provide interpretable descriptors and reproducible baseline features [24].

Structure-aware descriptors add the missing geometry. Crystal features include lattice constants, volume, density, space group, crystal system, Wyckoff positions, coordination environment, packing fraction, bond-length distributions, local order parameters, and symmetry-derived attributes. Local environment descriptors such as smooth overlap of atomic positions can encode three-dimensional neighborhoods and remain useful for kernel methods, clustering, and hybrid models [25, 26].

Graph representations form the central predictive view. In a crystal graph, atoms are nodes and neighbor relationships are edges. Edge attributes encode interatomic distances, bond directions, and local environments. Crystal graph convolutional neural networks established the usefulness of this representation for periodic materials [23]. More advanced graph neural networks and graph transformers can add angular information, line graphs, attention, or equivariance. For tensorial properties, symmetry-aware architectures are preferred because they enforce the transformation behavior required by physical tensors [27].

Pretrained atomistic embeddings are the newest feature source. CHGNet, MatterSim, MACE-class models, and related foundation potentials can serve as structural encoders whose internal representations are transferred into downstream property predictors [2, 3, 4]. Matter AI therefore uses three feature families in parallel: interpretable descriptors, explicit crystal graphs, and pretrained atomistic embeddings. The final model can select or combine these views depending on data size, property type, required interpretability, and computational budget.

3.4 Machine Learning and Graph Neural Network Property Prediction

The prediction layer is organized in tiers. Tier one uses classical machine learning models such as random forests, gradient boosting, support vector machines, and calibrated linear baselines on composition and structure descriptors. These models are fast and interpretable, making them useful for error analysis and early screening. Tier two uses supervised graph neural networks and graph transformers for structure-property prediction. Tier three uses pretrained atomistic models and adapters for high-value candidates, molecular dynamics, relaxation, fine-tuning, and transfer learning.

Target properties depend on the physical-science use case. For energy materials, target labels can include voltage, ionic conductivity, diffusion barriers, phase stability, dielectric constant, and electrochemical window. For semiconductors, targets can include band gap, effective mass, dielectric response, defect tolerance, and thermal conductivity. For structural materials, targets can include elastic moduli, yield-related descriptors, density, thermal stability, oxidation resistance, and fatigue-related proxies. Matter AI is designed as a general framework in which property heads can be attached without changing the data and validation architecture.

Uncertainty quantification is mandatory. A prediction without uncertainty is unsafe for active discovery because it can rank out-of-distribution materials too highly. Matter AI therefore stores predictive mean, epistemic uncertainty, aleatoric uncertainty when available, calibration error, and conformal prediction intervals. Deep ensembles, representation-distance measures, and conformal calibration can be used together. Candidates with high predicted performance but high uncertainty are not rejected automatically; instead, they may be selected by the active-learning acquisition function if the expected information gain is large.

Multimodal reasoning is handled through a bridge design. Inspired by structure-aware multimodal language models, a pretrained atomistic encoder can represent the crystal while a language model processes textual goals, synthesis evidence, and user constraints [1]. This bridge allows queries such as: identify lithium conductors with high predicted stability, low cobalt and nickel content, low toxicity, and synthesis routes below a chosen temperature. The language model provides flexible interaction, while the structural encoder and predictive models provide grounded materials reasoning.

3.5 Physics Informed Validation

Physics-informed validation reduces false positives before expensive computation or experiment. The first filter checks basic chemical plausibility, including charge balance, oxidation-state reasonableness, atomic

overlap, coordination environment, density, and symmetry consistency. The second filter checks thermodynamic plausibility through formation energy and energy above hull. The third filter applies advanced tests such as phonon stability, elastic tensor stability, magnetic ordering consistency, defect plausibility, or finite-temperature simulation when the application requires it.

The validation layer does not require every candidate to undergo density functional theory immediately. Instead, a graduated pipeline is used. Low-cost heuristics and pretrained potentials triage large candidate sets. Medium-cost surrogate simulations relax structures and estimate local stability. High-cost density functional theory, molecular dynamics, phonon calculations, or defect calculations are reserved for candidates that survive early filtering. This staged design aligns computational cost with scientific value.

Physics validation also improves interpretability. A candidate can be rejected because it violates charge neutrality, because it is far above the convex hull, because relaxation collapses its proposed symmetry, because predicted uncertainty is too large, or because sustainability constraints dominate its utility. Such explicit rejection reasons are essential for human trust and for model improvement.

3.6 Candidate Ranking and Active Learning

Candidate ranking is performed as a multi-objective decision problem. The output is not a single best material but a Pareto-ranked set of candidates across performance, uncertainty, stability, synthesis feasibility, novelty, cost, toxicity, supply risk, recyclability, and environmental impact. This is important because materials discovery rarely has one objective. A candidate with the highest predicted property may be unacceptable if it depends on critical elements, toxic precursors, extreme synthesis energy, or poor recyclability.

The proposed utility function for a candidate material m is expressed as follows.

$$U(m) = aP(m) - bV(m) - cH(m) + dF(m) + eT(m) + fS(m) + gN(m) \quad (1)$$

In Equation 1, P(m) is property desirability, V(m) is predictive uncertainty, H(m) is energy above hull or instability penalty, F(m) is physics-consistency score, T(m) is synthesis feasibility, S(m) is sustainability score, N(m) is novelty, and a to g are non-negative weights selected for the discovery objective. The equation is not intended to replace Pareto analysis. It is used only after hard constraints and Pareto filtering have removed candidates that are clearly dominated.

Active learning controls which candidates should be simulated or synthesized next. The acquisition function can prioritize high predicted utility, high uncertainty, novelty, expected improvement, expected hypervolume improvement, or expected sustainability improvement. A balanced loop is preferred: exploitation identifies likely winners, while exploration identifies uncertain regions where the model can learn. GNoME and foundation-model active learning show that iterative model-guided search is a realistic path for large-scale discovery [6, 10].

3.7 Human or Robotic Experimental Validation

Experimental validation is the final authority in Matter AI. Human researchers or robotic platforms synthesize selected candidates, characterize the resulting materials, and return successes, failures, partial outcomes, and ambiguous outcomes to the knowledge graph. A-Lab demonstrated that autonomous synthesis planning, robotics, computation, and historical knowledge can be integrated for inorganic materials synthesis [7]. However, subsequent correction and reanalysis also showed that novelty claims and structural identification require careful auditing [8]. Matter AI therefore uses graduated autonomy rather than full automation without review.

Before any robotic execution, the framework generates a synthesis plan with precursor identities, masses or molar ratios, heating profile, atmosphere, dwell time, cooling schedule, safety constraints, expected phase, and characterization plan. Human approval is required when hazardous precursors, high temperatures, uncertain phase behavior, or ambiguous literature evidence are involved. After synthesis, characterization can

include X-ray diffraction, electron microscopy, energy-dispersive spectroscopy, Raman or infrared spectroscopy, thermal analysis, electrochemical testing, or other domain-specific measurements.
The validation outcome is not binary. Matter AI records whether the target phase was obtained, whether secondary phases were present, whether characterization was sufficient, whether the synthesis was repeatable, and whether the measured property matched the prediction. Failed synthesis is treated as useful data because it improves synthesis-feasibility models and prevents repeated recommendations of inaccessible candidates.

3.8 Sustainability Scoring

Sustainability scoring is treated as a first-class module rather than an afterthought. For each candidate, Matter AI estimates cost, embodied energy, toxicity, recyclability, supply risk, and environmental impact. Early screening can use composition-level proxies, but final candidates should be assessed using life-cycle assessment principles, toxicity databases, criticality data, and process-specific assumptions. ISO life-cycle assessment standards, ecoinvent, openLCA, USEtox, EPA CompTox, USGS mineral commodity data, European critical raw materials information, and circularity tools provide relevant external infrastructure [28, 29, 30, 31, 32, 33, 34].

The sustainability score is defined as follows.

$$S(m) = w1C(m) + w2E(m) + w3Q(m) + w4R(m) + w5L(m) \quad (2)$$

In Equation 2, C(m) is normalized cost score, E(m) is embodied energy and greenhouse-gas score, Q(m) is toxicity and hazard score, R(m) is recyclability and circularity score, L(m) is supply-risk and criticality score, and w1 to w5 are user-defined weights. Each score is normalized so that a larger value indicates a more desirable sustainability profile. The score should include uncertainty because early-stage materials often lack complete lifecycle inventory data.
Sustainability-aware ranking prevents technically impressive but irresponsible outcomes. For example, a candidate that improves performance by a small amount while increasing toxicity, cost, and supply risk should not dominate a safer candidate with slightly lower performance. MatterGen's low-supply-chain-risk magnet example indicates that supply-risk constraints can be built directly into generative materials design [5]. Matter AI generalizes this idea by applying sustainability scoring to both generated and database-derived candidates.

Table 2: Public Data Sources and Benchmarks Relevant to Matter AI

Resource	**Role in Matter AI**	**Main Data Type**	**Example Use**
Materials Project	Core computed inorganic materials source	Structures, energies, phase data, properties	Training, screening, hull analysis
OQMD	Thermodynamic and structural database	DFT calculations and phase stability	Convex-hull screening and transfer validation
AFLOW	High-throughput standardized repository	Large computed property collections	Broad feature and label diversity
NOMAD	Metadata-rich archive and parser ecosystem	Raw and normalized computational records	Provenance and reproducibility
JARVIS	NIST-supported materials data	Electronic, elastic, vibrational, and low-dimensional data	Independent benchmark and property diversity
OPTIMADE	Interoperability layer	Common API schema	Federated database access

Resource	Role in Matter AI	Main Data Type	Example Use
MPtrj and OMat24	Foundation-model training data	Relaxation trajectories and DFT calculations	Pretraining and model comparison
Matbench Discovery	Prospective stability benchmark	Stability labels and leaderboard metrics	Discovery-oriented evaluation

Table 3: Evaluation Metrics for the Proposed Framework

Layer	Metrics	Purpose
Literature mining	Precision, recall, F1, null precision, tuple completeness, source-span accuracy	Measures extraction quality and auditability
Property prediction	Mean absolute error, root mean squared error, R squared, precision at k, F1, AUROC	Measures predictive accuracy and screening utility
Uncertainty	Expected calibration error, conformal coverage, negative log likelihood, sparsification error	Measures reliability of confidence estimates
Physics validation	Energy above hull, relaxation success, phonon stability, elastic stability, defect plausibility	Measures physical consistency
Candidate ranking	Pareto hypervolume, discovery acceleration, novelty rate, constraint satisfaction	Measures usefulness of shortlist
Experiment	Synthesis success rate, repeatability, characterization confidence, property agreement	Measures real-world validation
Sustainability	Cost score, embodied energy, toxicity score, recyclability, supply-risk score	Measures responsible material selection

Table 4: Sustainability Dimensions in Matter AI

Dimension	Operational Meaning	Representative Data Source
Cost	Raw material and precursor burden	USGS commodity summaries, supplier data
Energy	Embodied energy and process energy	Life-cycle inventory and process models
Toxicity	Human and ecological hazard potential	USEtox, EPA CompTox, PubChem
Recyclability	End-of-life recovery and circularity potential	Material Circularity Indicator and sector recycling data
Environmental impact	Greenhouse gas, water, land, and pollution indicators	ISO life-cycle assessment, ecoinvent, openLCA

Dimension	Operational Meaning	Representative Data Source
Supply risk	Criticality, import reliance, geographic concentration	European critical raw materials data and USGS data

4. Mathematical Formulation of Discovery Loop

The discovery loop can be formulated as an iterative decision process. Let M be the current material search space, D be the current evidence database, f be the predictive model, p be the physics-validation operator, s be the sustainability-scoring function, and e be the experimental validation operator. At each iteration, a candidate subset is selected from M using predicted property desirability, uncertainty, physics feasibility, synthesis feasibility, novelty, and sustainability.

The selected candidate set C is defined as follows.

C = ParetoTopK(P(m), -V(m), -H(m), F(m), T(m), S(m), N(m)), where m belongs to M (3)

In Equation 3, ParetoTopK returns non-dominated candidates and then selects the top K candidates under the final utility policy. The negative signs indicate that uncertainty and instability are minimized while the other terms are maximized. The output C is sent to additional simulation or experiment depending on cost and confidence.
After validation, the evidence database is updated as follows.

D(t + 1) = D(t) union e(C) union p(C) union s(C) (4)

Equation 4 emphasizes that the database is expanded not only with successes but also with failures, inconclusive characterization, corrected labels, updated sustainability estimates, and model uncertainty. This creates a learning system in which each cycle improves the next cycle.

5. Implementation and Evaluation Protocol

A practical implementation of Matter AI should begin with a well-defined use case, such as lithium solid electrolytes, lead-free perovskites, low-cobalt cathodes, high-temperature alloys, thermoelectrics, or low-supply-risk magnets. The target property, acceptable chemistry, safety constraints, operating conditions, and sustainability priorities must be specified before model training. Without this problem definition, the framework may optimize generic metrics that do not correspond to useful materials.

The data pipeline should be implemented in five steps. First, materials records should be retrieved from public databases through official APIs and OPTIMADE where possible. Second, literature records should be extracted using retrieval-augmented large language models with source tracking. Third, all records should be normalized into a canonical schema containing structures, properties, methods, units, conditions, provenance, and uncertainty. Fourth, training and test splits should be constructed to avoid information leakage through near-duplicate structures or polymorphs. Fifth, dataset manifests and license information should be stored with the model artifacts.

The model pipeline should compare multiple baselines. Classical descriptor models should be used to establish interpretable baselines. Crystal graph neural networks and graph transformers should be used for structure-property learning. Pretrained atomistic models should be evaluated as feature extractors, relaxers, or fine-tuned predictors. Multimodal bridge models should be evaluated on structure-conditioned question answering, synthesis suggestion, and retrieval-grounded explanation. No single model should be declared superior without property-specific, split-specific, and uncertainty-aware evaluation.

The evaluation protocol should include ablations. The minimum ablations are: no retrieval in literature mining, no source tracking, no structure encoder, no symmetry features, no pretrained atomistic embeddings,

no uncertainty calibration, no physics filter, no sustainability term, and no human review gate. Each ablation should be compared against the full Matter AI system using predictive metrics, discovery metrics, and practical review effort. This prevents inflated claims from a single favorable benchmark.

Reproducibility should be built into the workflow. Code should include environment files, random seeds, versioned datasets, train-test split files, prompt templates, parsing scripts, model cards, and evaluation reports. Simulation inputs should include pseudopotentials, convergence settings, exchange-correlation functional, relaxation criteria, and software versions. Experimental validation should include protocols, raw characterization files, phase-identification assumptions, and repeatability records.

6. Results and Discussion

Because this paper proposes a framework and does not report newly executed experiments, the results are presented as an evidence-based framework analysis rather than fabricated numerical outcomes. The recent literature supports the feasibility of each Matter AI component. Source-tracked literature extraction has shown high performance in focused materials domains, but broader unbiased extraction remains recall-limited [13, 14]. Therefore, Matter AI uses large language models for evidence extraction and triage, not as unattended authorities.

For structure-aware prediction, the evidence supports pretrained atomistic models and graph neural networks as core modules. CHGNet, MatterSim, MACE-class models, and OMat24 demonstrate that large-scale pretraining can improve transfer and simulation capability [2, 3, 4, 10]. However, high benchmark accuracy is not enough. Matbench Discovery shows that discovery-oriented stability evaluation is required because the practical goal is to identify stable candidates efficiently rather than to minimize a generic regression error [9]. Matter AI therefore combines prediction accuracy with uncertainty calibration and stability-oriented screening.

For generation and active learning, MatterGen and GNoME show that machine learning can propose large numbers of candidate materials and can guide discovery beyond known databases [5, 6]. The challenge is that generated candidates must be filtered for physics, synthesis feasibility, novelty, and sustainability. Matter AI addresses this by treating generation as only one part of the loop. A generated candidate must still pass database deduplication, stability checks, uncertainty analysis, sustainability scoring, and experimental planning.

For experimental validation, A-Lab is a major example of autonomous materials synthesis [7]. The later correction is equally important because it highlights the need for careful novelty definitions and characterization standards [8]. Matter AI therefore avoids a fully automatic success label. It stores evidence strength, characterization confidence, repeatability, and human adjudication status. This makes the framework safer for publication-quality claims.

The strongest advantage of Matter AI is the coupling of layers. Literature mining informs synthesis feasibility. Databases provide structures and labels. Features represent chemistry and geometry. Models predict properties and uncertainty. Physics filters remove impossible candidates. Ranking balances performance and responsibility. Experiments validate or reject candidates. Sustainability scoring prevents harmful optimization. The framework becomes valuable because it turns hidden handoffs into monitored interfaces.

7. Limitations and Ethical Considerations

Several limitations must be acknowledged. First, large language models can hallucinate, omit evidence, or misread tables and supplementary files. Source tracking and abstention reduce but do not eliminate this risk. Second, materials databases are biased toward computable, stable, publishable, and historically popular chemistries. Third, graph neural networks and foundation atomistic models can fail under out-of-distribution chemistry, unusual oxidation states, defects, disorder, surfaces, or high-temperature conditions. Fourth, density functional theory labels contain method-dependent errors and are not identical to experimental reality.

Sustainability scoring also has uncertainty. Early-stage candidates often lack process-specific life-cycle inventory data. Toxicity may depend on chemical form, exposure route, and manufacturing context. Recyclability depends on product design and collection infrastructure, not only composition. Criticality changes over time with geopolitics, mining capacity, substitution, and recycling. Matter AI should therefore represent sustainability outputs as uncertainty-aware decision aids, not absolute judgments.

Ethical deployment requires safety controls. The framework should not recommend unsafe syntheses without human approval. Hazardous precursors, high-pressure or high-temperature processes, toxic emissions, and restricted materials require institutional review. Open science must be balanced with responsible release when candidate materials or processes could create safety or environmental risks. The role of the scientist remains essential: to define meaningful goals, evaluate evidence, interpret anomalies, and decide what should be made.

8. Conclusion

Matter AI is proposed as a multimodal, physics-informed, sustainability-aware framework for accelerated materials discovery in physical sciences. The framework integrates literature mining using large language models, federated materials databases, chemical and crystal-structure feature extraction, machine learning and graph neural network prediction, physics-based validation, candidate ranking, human or robotic experimental validation, and sustainability scoring. It is designed for the current 2023 to 2026 research landscape in which foundation atomistic models, generative materials models, multimodal language systems, open large-scale datasets, and autonomous laboratories are all advancing rapidly.

The core argument is that the future of materials discovery will not be driven by a single model. It will be driven by governed closed-loop systems that preserve provenance, quantify uncertainty, respect physical constraints, validate experimentally, and account for environmental consequences. Matter AI provides such an architecture. It can be implemented with existing databases and model families, but its main contribution is the integration logic: every prediction should remain connected to evidence, physics, validation, and sustainability. This design can help reduce discovery time, improve reliability, and guide physical sciences toward materials that are not only high performing but also responsible and deployable.

REFERENCES:

1. Tang Y., et al., "A Multimodal Large Language Model for Materials Science", Nature Machine Intelligence, 2026. https://www.nature.com/articles/s42256-026-01214-y
2. Deng B., et al., "CHGNet as a Pretrained Universal Neural Network Potential for Charge-Informed Atomistic Modelling", Nature Machine Intelligence, 2023. https://arxiv.org/abs/2302.14231
3. Yang H., et al., "MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures", arXiv, 2024. https://arxiv.org/abs/2405.04967
4. Batatia I., et al., "A Foundation Model for Atomistic Materials Chemistry", The Journal of Chemical Physics, 2025, 163 (18), 184110. https://pubs.aip.org/aip/jcp/article/163/18/184110/3372267/A-foundation-model-for-atomistic-materials
5. Zeni C., et al., "MatterGen: A Generative Model for Inorganic Materials Design", Nature, 2025. https://www.nature.com/articles/s41586-025-08628-5
6. Merchant A., et al., "Scaling Deep Learning for Materials Discovery", Nature, 2023. https://www.nature.com/articles/s41586-023-06735-9
7. Szymanski N.J., et al., "An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials", Nature, 2023. https://www.nature.com/articles/s41586-023-06734-w
8. Szymanski N.J., et al., "Author Correction: An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials", Nature, 2026. https://www.nature.com/articles/s41586-025-09992-y
9. Riebesell J., et al., "Matbench Discovery: A Framework to Evaluate Machine Learning Crystal Stability Predictions", Nature Machine Intelligence, 2025. https://www.nature.com/articles/s42256-025-01055-1
10. Barroso-Luque L., et al., "Open Materials 2024 OMat24 Inorganic Materials Dataset and Models", arXiv, 2024. https://arxiv.org/abs/2410.12771

11. Polak M.P., Morgan D., "Extracting Accurate Materials Data from Research Papers with Conversational Language Models and Prompt Engineering", Nature Communications, 2024. https://www.nature.com/articles/s41467-024-45914-8
12. Dagdelen J., et al., "Structured Information Extraction from Scientific Text with Large Language Models", Nature Communications, 2024. https://www.nature.com/articles/s41467-024-45563-x
13. Wang X., Raj A., Luebbe M., Wen H., Xu S., Lu K., "Reliable End-to-End Material Information Extraction from the Literature with Source-Tracked Multi-Stage Large Language Models", arXiv, 2025. https://arxiv.org/abs/2510.05142
14. Ning W., Li M., Reimers J.R., Kobayashi R., "Optimizing Data Extraction from Materials Science Literature: A Study of Tools Using Large Language Models", arXiv, 2025. https://arxiv.org/abs/2512.09370
15. Zaki M., Jayadeva, Mausam, Krishnan N.M.A., "MaScQA: A Question Answering Dataset for Investigating Materials Science Knowledge of Large Language Models", arXiv, 2023. https://arxiv.org/abs/2308.09115
16. Liu S., Xu J., Ye B., Hu B., Srolovitz D.J., Wen T., "MatTools: Benchmarking Large Language Models for Materials Science Tools", arXiv, 2025. https://arxiv.org/abs/2505.10852
17. Jain A., et al., "The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation", APL Materials, 2013, 1 (1), 011002. https://next-gen.materialsproject.org/
18. Saal J.E., Kirklin S., Aykol M., Meredig B., Wolverton C., "Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database", JOM, 2013, 65, 1501-1509. https://oqmd.org/
19. Curtarolo S., et al., "AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput Ab Initio Calculations", Computational Materials Science, 2012, 58, 227-235. https://aflow.org/
20. Andersen C.W., et al., "OPTIMADE, an API for Exchanging Materials Data", Scientific Data, 2021, 8, 217. https://www.nature.com/articles/s41597-021-00974-z
21. Evans M.L., et al., "Developments and Applications of the OPTIMADE API for Materials Discovery, Design, and Data Exchange", arXiv, 2024. https://arxiv.org/abs/2402.00572
22. Choudhary K., et al., "The Joint Automated Repository for Various Integrated Simulations JARVIS for Data-Driven Materials Design", npj Computational Materials, 2020. https://www.nist.gov/programs-projects/jarvis-dft
23. Xie T., Grossman J.C., "Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties", Physical Review Letters, 2018, 120, 145301. https://link.aps.org/doi/10.1103/PhysRevLett.120.145301
24. Ward L., et al., "Matminer: An Open Source Toolkit for Materials Data Mining", Computational Materials Science, 2018, 152, 60-69. https://hackingmaterials.lbl.gov/matminer/
25. Himanen L., et al., "DScribe: Library of Descriptors for Machine Learning in Materials Science", Computer Physics Communications, 2020, 247, 106949. https://singroup.github.io/dscribe/latest/tutorials/descriptors/soap.html
26. Laakso J., Himanen L., Homm H., Morooka E.V., Jager M.O.J., "Updates to the DScribe Library: New Descriptors and Derivatives", The Journal of Chemical Physics, 2023, 158 (23), 234802. https://pubs.aip.org/aip/jcp/article/158/23/234802/2897350/Updates-to-the-DScribe-library-New-descriptors-and
27. Wen T., et al., "MatTen: A Materials Tensor Representation and Learning Framework", Digital Discovery, 2024. https://pubs.rsc.org/en/content/articlehtml/2024/dd/d3dd00233k
28. International Organization for Standardization, "ISO 14040: Environmental Management - Life Cycle Assessment - Principles and Framework", ISO, 2006. https://www.iso.org/standard/37456.html
29. USEtox, "USEtox Model: Scientific Consensus Model for Characterizing Human and Ecotoxicological Impacts", USEtox, 2026. https://usetox.org/model
30. United States Geological Survey, "Mineral Commodity Summaries 2025", United States Geological Survey, 2025. https://pubs.usgs.gov/publication/mcs2025
31. European Commission, "Critical Raw Materials", European Commission, 2026. https://single-market-economy.ec.europa.eu/sectors/raw-materials/areas-specific-interest/critical-raw-materials_en

32. GreenDelta, "openLCA: Open Source Life Cycle and Sustainability Assessment Software", GreenDelta, 2026. https://www.openlca.org/
33. United States Environmental Protection Agency, "CompTox Chemicals Dashboard and DSSTox Database", United States Environmental Protection Agency, 2026. https://www.epa.gov/comptox-tools/comptox-chemicals-dashboard
34. Ellen MacArthur Foundation, "Material Circularity Indicator", Ellen MacArthur Foundation, 2026. https://www.ellenmacarthurfoundation.org/material-circularity-indicator