

ORIGINAL RESEARCH

Open access

# Knowledge Graphs vs Property Predictors: Competing Infrastructures of Materials Intelligence

Fatima Zahra Amrani<sup>1\*</sup>, Youssef Benali<sup>1</sup>

## Abstract

In the evolving landscape of computational and data-driven materials engineering, the integration of machine learning techniques has transformed traditional discovery paradigms into intelligent, autonomous systems. Materials informatics leverages vast datasets from high-throughput computations and multimodal sources to accelerate the design of novel materials with tailored properties. However, a conceptual gap persists in understanding the infrastructural roles of knowledge graphs and property predictors as competing yet complementary architectures for materials intelligence. Knowledge graphs offer relational representations that capture complex interdependencies among materials entities, enabling semantic querying and inference across disparate data modalities. In contrast, property predictors, often based on graph neural networks or deep learning models, focus on direct regression or classification of material attributes, prioritizing predictive accuracy over holistic system integration. This manuscript introduces a novel conceptual framework, termed the Dual-Infrastructure Materials Cognition (DIMC) model, which interprets the dynamic interplay between these infrastructures through layered computational workflows and feedback mechanisms. By examining representation learning, uncertainty quantification, and closed-loop discovery logics, the framework elucidates trade-offs in scalability, interpretability, and epistemic robustness. Implications for the field include enhanced steering of autonomous discovery systems, improved coupling of simulation and experimentation, and refined strategies for inverse materials design. Ultimately, this interpretive lens fosters a more cohesive ecosystem for materials intelligence, bridging isolated predictive tools with knowledge-centric infrastructures to advance data-driven innovation in materials science.

**Keywords** Autonomous discovery, Materials informatics, Graph neural networks, Representation learning, Knowledge graphs, Property predictors

\*Correspondence:

Fatima Zahra Amrani

fatima.amrani@outlook.com

<sup>1</sup> Department of Computational Materials Science, Faculty of Engineering, Mohammed V University, Rabat, Morocco

## Introduction

### The emergence of computational materials ecosystems

The advent of computational materials science has ushered in an era where data-driven approaches dominate the quest for new materials with superior performance. High-throughput computations, powered by density functional theory and molecular dynamics simulations, generate expansive datasets that fuel machine learning models for property prediction and materials design [1, 2]. This shift from empirical trial-and-error to informatics-based strategies has been instrumental in addressing grand challenges in energy storage, catalysis, and structural materials. For instance, machine learning frameworks have enabled the rapid screening of vast chemical spaces, identifying candidates for batteries and perovskites with enhanced stability and efficiency [3, 4]. Yet, as datasets grow in complexity and volume, the need for

robust infrastructural elements becomes paramount. Knowledge graphs and property predictors emerge as pivotal components, each offering distinct mechanisms for organizing and extracting intelligence from materials data.

Knowledge graphs, rooted in semantic web technologies, structure materials information as interconnected nodes and edges, representing entities such as atoms, compounds, and processes along with their relationships [5]. This relational architecture facilitates querying across multimodal datasets, integrating experimental observations with computational simulations. Property predictors, conversely, employ supervised learning paradigms to map input representations—often featurized as graphs or vectors—to output properties like band gaps or mechanical strengths [6, 7]. These predictors thrive on pattern recognition within large datasets, but their black-box nature can obscure underlying physical insights. The competition between these infrastructures lies in their differing emphases: graphs prioritize connectivity and context, while predictors emphasize efficiency and specificity.

## Challenges in representation and inference

A core challenge in materials intelligence is the faithful representation of materials' structural and functional attributes. Traditional featurization methods, such as atomic fingerprints or descriptors, have evolved into sophisticated representation learning techniques, including graph neural networks that encode crystal structures for property forecasting [8]. However, these representations often struggle with multimodality, where data from spectroscopy, microscopy, and simulations must be harmonized [9]. Knowledge graphs address this by embedding ontologies that link disparate modalities, enabling inference beyond mere prediction. For example, graphs can infer novel synthesis routes by traversing relational paths, a capability less inherent in standalone predictors [10].

Inference mechanisms further highlight the infrastructural divide. Property predictors rely on statistical correlations, often incorporating uncertainty quantification to assess prediction reliability [11, 12]. Techniques like Gaussian processes or ensemble methods provide error estimates, crucial for guiding experimental validation [13]. In contrast, knowledge graphs support logical inference, drawing on rule-based systems or embedding-based similarities to uncover hidden associations [14]. This duality raises questions about epistemic completeness: predictors may excel in quantitative accuracy but falter in qualitative reasoning, while graphs offer breadth at the potential cost of depth in specific predictions.

The integration of these elements into autonomous discovery systems amplifies their competing roles. Closed-loop experimentation, where predictions inform robotic synthesis and characterization, demands seamless data flow [15]. Here, property predictors drive hypothesis generation, but knowledge graphs ensure contextual coherence, preventing siloed discoveries [16]. Uncertainty quantification plays a bridging role, steering workflows toward regions of high informational value [17].

## Paradigmatic shifts in data-driven discovery

The paradigm of inverse materials design exemplifies the need for infrastructural synergy. Rather than forward prediction, inverse approaches start from desired properties and navigate backward to candidate structures [18]. Property predictors facilitate this through generative models or optimization algorithms, but they require robust representations to avoid extrapolation errors [19]. Knowledge graphs enhance this by providing prior knowledge constraints, such as thermodynamic stability rules derived from literature syntheses [20].

Foundation models for science, inspired by large language models, represent a convergence point, where pre-trained architectures handle both predictive and relational tasks [21]. These models ingest multimodal data, learning emergent patterns that span prediction and graph-based reasoning [22]. However, their infrastructural underpinnings remain underexplored, often blending predictor efficiency with graph-like connectivity without explicit demarcation.

Simulation-experiment coupling further underscores infrastructural competition. High-fidelity simulations generate data for predictors, but experimental feedback loops necessitate graph structures to track provenance and iterations [23]. This

coupling demands adaptive workflows, where predictors update in real-time while graphs maintain long-term knowledge repositories [24].

## Positioning the conceptual inquiry

Amid these developments, a conceptual void exists in interpreting knowledge graphs and property predictors not as tools, but as competing infrastructures shaping materials intelligence. This manuscript synthesizes recent advancements to propose a framework that interprets their interactions through computational dynamics and epistemic structures. By delineating layers of representation, inference, and discovery steering, the framework offers insights into optimizing hybrid ecosystems for sustainable materials innovation.

# Theoretical Background & Literature Synthesis

## Foundations of materials informatics and machine learning integration

Materials informatics has matured as a discipline that harnesses data science to expedite materials discovery, drawing on principles from computational chemistry and machine learning [1]. Early efforts focused on curating databases from high-throughput computations, enabling the application of regression models for property estimation [2]. Machine learning's role expanded with the adoption of deep learning architectures, particularly graph neural networks, which model materials as graphs with atoms as nodes and bonds as edges [3]. These networks have proven adept at capturing local and global structural features, leading to improved predictions for electronic and mechanical properties [6, 8].

The synthesis of literature reveals a trajectory toward more sophisticated representation learning. Initial descriptors were hand-crafted, but automated feature extraction via neural networks has supplanted them, allowing for end-to-end learning from raw structural data [7, 12]. This shift has implications for handling sparse or noisy datasets, common in materials research, where techniques like transfer learning mitigate data scarcity [14, 18].

## Representation learning paradigms in materials systems

Representation learning constitutes the epistemic and computational substrate upon which contemporary materials intelligence systems are constructed, functioning as the foundational layer for both predictive modeling architectures and relational knowledge infrastructures. Within property prediction pipelines, representations are engineered to maximize performance on downstream inferential tasks, including the classification of thermodynamic stability, regression of formation energies, bandgap estimation, and prediction of mechanical or electrochemical properties [4, 13]. The success of these predictors is deeply contingent upon the structural fidelity of the underlying representations. Graph-based encodings have emerged as particularly effective in this regard because they intrinsically capture atomistic connectivity, coordination environments, periodic boundary conditions, and crystallographic symmetries that define solid-state materials systems [15]. By embedding these structural priors directly into model architectures, graph neural frameworks reduce the need for handcrafted descriptors while preserving physically meaningful relational signals.

Recent literature advances this paradigm further through the development of multi-fidelity representation learning strategies, wherein embeddings are trained across heterogeneous accuracy regimes [16, 22]. In such systems, coarse-grained simulation outputs or lower-cost computational approximations are fused with sparse, high-precision quantum calculations or experimental measurements. This fusion enables models to internalize broad chemical trends while remaining anchored to high-confidence reference points, thereby improving robustness, calibration, and generalization in data-scarce regimes. The representational space thus becomes hierarchically structured, encoding both approximate global patterns and localized high-fidelity corrections.

In contrast to predictors, knowledge graphs operationalize representation through semantic and relational logics rather than task-optimized embeddings. Here, the emphasis shifts from numerical compression to ontological expressivity.

Materials entities are situated within structured schemas that encode relationships among composition, processing pathways, synthesis conditions, characterization outputs, and performance metrics [5, 9]. These semantic representations enable machine-interpretable reasoning across experimental and theoretical domains, effectively transforming fragmented datasets into navigable knowledge infrastructures. Such relational encoding is particularly conducive to multimodal integration. Textual evidence extracted from journal articles, patents, or technical reports can be linked to simulation-derived numerical attributes, thereby constructing unified knowledge layers that transcend modality boundaries [10, 21].

Synthesis of recent work demonstrates that knowledge graphs are increasingly capable of extracting latent materials knowledge from unstructured corpora through natural language processing and information extraction pipelines [19, 20]. This capability allows relational systems to expand dynamically as new literature emerges, enriching predictive ecosystems with contextual insights that would otherwise remain inaccessible to purely numerical models. Consequently, knowledge graphs function not only as storage architectures but also as epistemic amplifiers, augmenting discovery pipelines with historically accumulated scientific intelligence.

A central conceptual tension emerges at the intersection of these two paradigms—namely, the trade-off between featurization depth and featurization breadth. Property predictors are optimized for representational compression, distilling materials structures into dense, task-specific embeddings that maximize predictive efficiency [11, 17]. Knowledge graphs, conversely, prioritize representational expansiveness, preserving relational granularity to enable complex querying and inferential traversal. This divergence has direct implications for scalability and functional scope. Predictors excel in high-throughput virtual screening contexts, where millions of candidate materials must be evaluated computationally. Knowledge graphs, however, enable cross-domain reasoning, such as linking alloy compositions to lifecycle environmental impacts, regulatory constraints, or supply-chain considerations—forms of inference that extend beyond the predictive reach of isolated models [23, 24].

## Architectures for prediction and relational intelligence

The architectural foundations of property prediction systems are anchored in deep learning paradigms specifically adapted to atomistic and crystalline data structures. Convolutional neural networks, originally developed for Euclidean image domains, have been reconfigured to process voxelized or grid-based materials representations, while message-passing graph neural networks operate directly on atomic graphs to propagate chemical information across bonding topologies [3, 6]. These architectures transform structural inputs into scalar or tensorial property outputs, often integrating physics-informed constraints—such as energy conservation laws or symmetry invariances—to stabilize extrapolation and improve generalization beyond training distributions [7, 12].

Uncertainty quantification mechanisms further enhance the epistemic reliability of these predictors. Bayesian neural formulations, ensemble modeling strategies, and evidential deep learning approaches generate probabilistic outputs that estimate predictive confidence alongside property values [13, 17, 25]. Such calibrated uncertainty estimates are indispensable in materials discovery contexts, where experimental validation costs necessitate risk-aware candidate prioritization.

Knowledge graph architectures, by contrast, derive intelligence from relational embedding techniques that map entities and edges into continuous vector spaces while preserving logical structure [5, 14]. These embeddings enable hybrid reasoning modalities in which symbolic queries—such as identifying synthesis pathways—can be combined with similarity-based vector operations. Within autonomous laboratory ecosystems, such graph infrastructures track experimental lineage, encode procedural dependencies, and infer optimization trajectories based on accumulated empirical histories [15, 26].

Comparative synthesis across the literature reveals complementary functional strengths. Predictors specialize in direct structure-to-property mappings under conditions of data completeness, whereas graphs demonstrate resilience in

incomplete knowledge environments by propagating inference across relational pathways [19, 27]. This distinction positions predictors as engines of quantitative precision and graphs as infrastructures of contextual reasoning.

These infrastructural distinctions and complementarities are synthesized in Table 1.

**Table 1.** Comparative Infrastructural Functions of Knowledge Graphs and Property Predictors in Materials Intelligence

| Infrastructure Dimension | Knowledge Graphs                     | Property Predictors                   | Hybrid DIMC Interpretation         |
|--------------------------|--------------------------------------|---------------------------------------|------------------------------------|
| Representational Logic   | Ontological, relational schemas      | Latent vector embeddings              | Dual encoding substrate            |
| Data Modalities          | Textual, experimental, computational | Primarily structural + numerical      | Multimodal fusion pipelines        |
| Inference Mechanism      | Logical traversal, rule reasoning    | Statistical regression/classification | Coupled reasoning–prediction       |
| Uncertainty Handling     | Propagated through relational edges  | Output-level probabilistic estimates  | Harmonized uncertainty fields      |
| Scalability              | Limited by graph complexity          | Highly scalable screening             | Tiered infrastructure scaling      |
| Interpretability         | High semantic transparency           | Often black-box                       | Context-augmented interpretability |
| Discovery Role           | Contextual navigation                | Candidate prioritization              | Closed-loop orchestration          |
| Inverse Design Support   | Feasibility constraints              | Generative design engines             | Constrained generative discovery   |
| Epistemic Risks          | Ontological gaps                     | Extrapolation errors                  | Risk counterbalancing              |
| Autonomous Labs Role     | Workflow memory + lineage            | Hypothesis generation                 | Integrated steering cognition      |

Hybrid architectures increasingly dissolve the boundary between these systems. Predictive models are now embedded within graph frameworks, populating nodes with inferred attributes that expand relational knowledge layers [21, 22]. Conversely, graph-derived contextual priors are used to regularize predictor training, constraining learning trajectories within chemically and experimentally plausible regimes [10, 28]. This bidirectional integration mitigates representational bias, enhances interpretability, and supports more epistemically grounded discovery processes [4, 8].

High-throughput and autonomous discovery workflows

High-throughput computational infrastructures form the data-generative backbone of contemporary materials intelligence ecosystems [1, 2]. Automated density functional theory pipelines and combinatorial simulation platforms produce vast property datasets that feed predictive screening architectures. Machine learning predictors operate atop this substrate, filtering expansive chemical search spaces to identify high-potential candidates prior to costly experimental validation [4, 29].

Autonomous discovery frameworks extend this paradigm into fully closed-loop systems that iteratively integrate prediction, synthesis, and characterization [9, 15]. Within such environments, experimental outcomes are reintegrated into training datasets, enabling continuous model recalibration and adaptive hypothesis refinement.

Knowledge graphs play a critical infrastructural role in maintaining workflow coherence across these iterative cycles. By storing dynamic experimental records and relational dependencies, graphs ensure that discovery trajectories remain contextually grounded [5, 10]. Steering algorithms operating on these graphs direct exploration toward underrepresented compositional or structural domains, thereby counteracting search redundancy and epistemic stagnation [16, 19].

Trade-off analyses synthesized in the literature underscore functional complementarities. Predictors accelerate discrete computational steps, enabling rapid candidate triaging, whereas graphs provide longitudinal epistemic continuity, mitigating risks such as confirmation bias in data acquisition strategies [17, 20]. Uncertainty quantification operates across both layers: predictors generate variance-informed prioritization signals, while graphs encode confidence propagation through weighted relational edges [11, 13, 14, 18].

## Inverse design and multimodal data handling

Inverse design frameworks invert the traditional forward-prediction paradigm by generating candidate materials conditioned on desired functional properties [15, 18]. Generative architectures—including variational autoencoders and generative adversarial networks—construct latent design spaces from which novel structural configurations can be sampled [19, 21]. These systems enable goal-directed exploration but face persistent challenges in ensuring thermodynamic stability and synthetic feasibility.

Knowledge graphs address this constraint by embedding feasibility priors derived from historical synthesis knowledge and processing constraints [5, 10]. When integrated into generative loops, these relational filters constrain design outputs to experimentally plausible regimes, thereby improving translational viability.

The integration of multimodal datasets further complicates representational alignment. Predictors frequently encounter difficulties in reconciling heterogeneous modalities—such as spectroscopy, microscopy, and textual annotations—within unified embedding spaces [3, 7]. Knowledge graphs, however, provide ontological scaffolding that links modalities through shared entity anchors, enabling coherent cross-modal reasoning [9, 14]. Foundation-scale pretraining paradigms increasingly bridge these systems, producing universal representations capable of supporting both predictive inference and relational querying [16, 22].

## Epistemic and infrastructural considerations

Epistemic risk constitutes an inherent structural feature of materials AI ecosystems, emerging from dataset biases, representation incompleteness, and inductive modeling assumptions [11, 17]. Predictive architectures remain vulnerable to distributional overfitting, particularly when extrapolating into sparsely sampled compositional regimes. Knowledge graphs, while interpretively rich, may suffer from ontological gaps or incomplete relational encoding [20, 23].

Literature synthesis increasingly advocates hybrid infrastructural paradigms as epistemic counterbalances. Within such systems, predictors provide quantitative precision and screening scalability, while graphs deliver interpretability, contextual reasoning, and historical grounding [4, 8, 12].

From an infrastructural scaling perspective, these paradigms diverge. Predictive systems scale primarily through computational acceleration and parallelization, whereas graph systems depend on advances in distributed querying, ontology harmonization, and relational database optimization [2, 6, 13]. Closed-loop discovery architectures derive strategic advantage from integrating both modalities—leveraging predictors for rapid iterative cycles and graphs for cumulative knowledge consolidation across temporal horizons [9, 15, 18].

## Proposed conceptual framework

The Dual-Infrastructure Materials Cognition (DIMC) framework interprets knowledge graphs and property predictors as intertwined yet distinct layers within a unified materials intelligence ecosystem. This framework conceptualizes materials

discovery as a multi-layered computational pipeline, where data ingestion flows into representation formation, inference generation, and discovery steering. At its core, DIMC delineates three structural layers: the representational substrate, the inferential engine, and the steering nexus. The representational substrate harmonizes raw multimodal data—spanning atomic configurations, spectral profiles, and synthesis parameters—into compatible formats for both graph-based relational encoding and predictor-oriented vector embeddings.

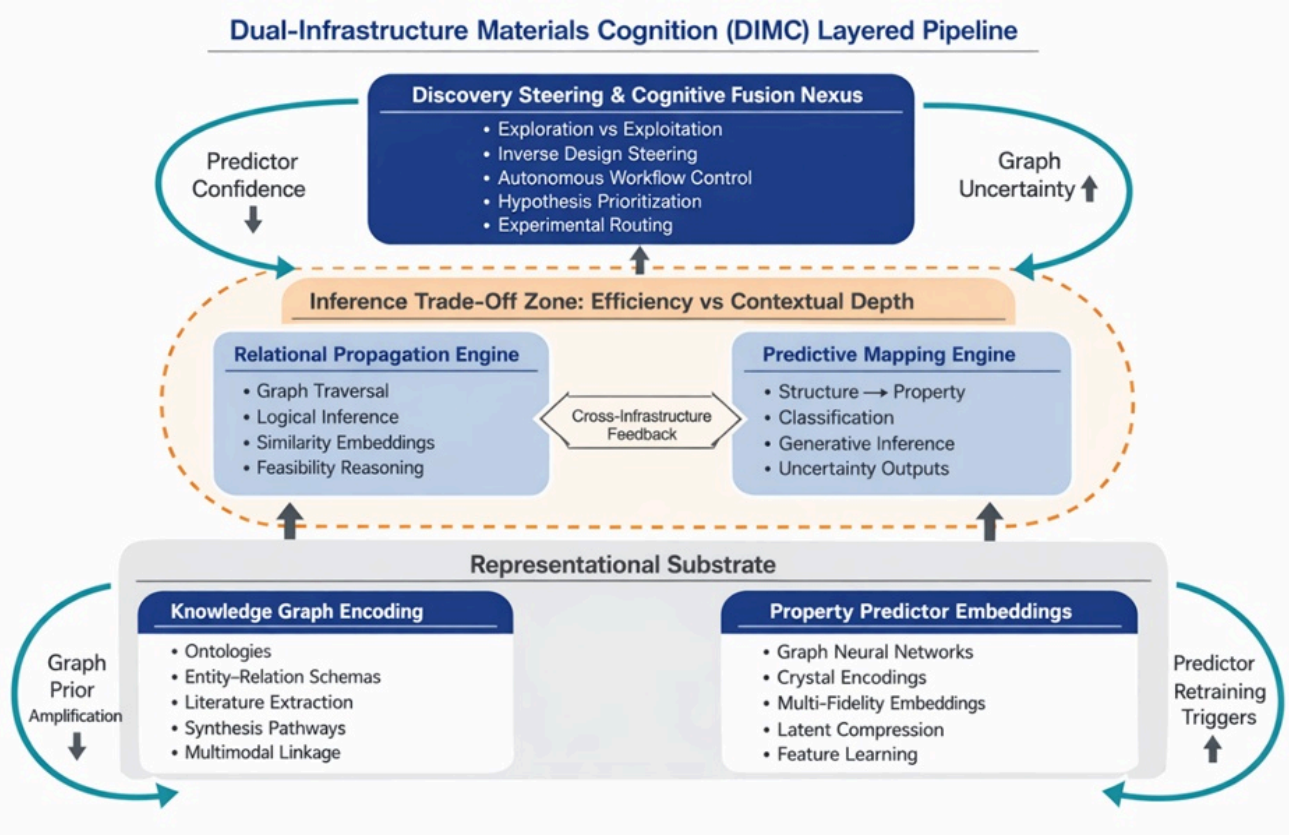
In the inferential engine layer, knowledge graphs handle relational propagation, while property predictors execute direct mappings. Feedback loops connect these, allowing predictor outputs to enrich graph nodes and graph-derived contexts to refine predictor inputs. The steering nexus oversees workflow dynamics, modulating exploration versus exploitation based on epistemic signals.

A key dynamic within DIMC captures the interaction between representational fidelity and inferential efficiency, which may be expressed as  $\alpha = \frac{I}{f(R_g, R_p)} \cdot \alpha$ , where  $R_g$  denotes graph relational depth,  $R_p$  property predictor specificity,  $f$  a fusion function, and  $\alpha$  an adaptability coefficient reflecting feedback strength. This expression interprets how fused representations enhance overall intelligence without implying empirical measurement.

Another aspect formalizes trade-offs in uncertainty handling: 
$$U = \beta \frac{U_g + U_p}{\gamma | U_g - U_p |},$$
 with  $U_g$  and  $U_p$  as graph and predictor uncertainties,  $\beta$  a propagation factor, and  $\gamma$  a dissonance penalty. This captures how aligned uncertainties bolster system robustness, while discrepancies introduce epistemic risks.

Discovery steering logics in DIMC emphasize adaptive pipelines, where closed-loop mechanisms adjust based on infrastructural signals. For instance, if predictor confidence dips, graph inference amplifies to provide contextual priors. This interplay fosters resilient materials intelligence, as conceptualized in **Figure 1**.





**Figure 1.** Dual-Infrastructure Materials Cognition (DIMC) architecture for integrated materials intelligence.

The framework conceptualizes materials discovery as a layered cognitive system composed of a representational substrate, an inferential engine, and a steering nexus. Knowledge graphs and property predictors operate as parallel infrastructures that exchange feedback signals across relational and predictive domains. Their integration supports uncertainty-aware discovery steering, adaptive inverse design, and closed-loop experimental orchestration.

A final conceptualization addresses feedback loop dynamics: 
$$\frac{F}{\delta} = \int (G \circ P) dt,$$
 where  $G$  and  $P$  represent graph and predictor operations,  $\circ$  their composition,  $\delta$  a damping term, and the integral over time denoting cumulative learning. This interprets sustained infrastructure evolution without predictive claims.

Through these elements, DIMC offers interpretive insights into optimizing computational workflows for materials discovery.

Analytical implications

The DIMC framework yields interpretive insights into the computational dynamics of materials intelligence, particularly in how knowledge graphs and property predictors interplay to shape discovery pipelines. One implication concerns representation-inference interactions, where the fusion of graph relational structures with predictor embeddings can mitigate epistemic gaps in multimodal data handling. For instance, in inverse design workflows, graphs provide ontological constraints that temper predictor-generated candidates, ensuring alignment with physical plausibility derived from high-throughput data [1, 15, 18]. This interaction interprets a balanced approach to scalability, where predictor efficiency scales computations while graph connectivity preserves systemic coherence [2, 3].

Systems-level insights emerge in uncertainty quantification, as DIMC highlights propagation mechanisms across infrastructures. Predictors often localize uncertainty to specific outputs, but graphs distribute it through relational



networks, offering a holistic view of confidence landscapes [11–13]. An interpretive formalization of this dynamic may be expressed as  $\overset{Q}{=} \eta (Q_p \otimes Q_g)$ , where  $Q_p$  and  $Q_g$  represent predictor and graph quantifications,  $\otimes$  a tensor product symbolizing multi-dimensional integration, and  $\eta$  a harmonization scalar. This captures how intertwined uncertainties enhance decision-making in autonomous systems, without empirical validation [17, 20].

Feedback loops within DIMC imply refined discovery steering logics. In closed-loop experimentation, predictor outputs can trigger graph updates, iteratively refining representations [9, 16]. This loop interprets adaptive behaviors, where initial predictor biases are corrected via graph-inferred contexts, fostering robust exploration of materials spaces [4, 19]. Epistemic risk structures are thus minimized, as the dual infrastructure counters isolated predictor overconfidence with graph-mediated validations [5, 14].

Infrastructure trade-offs become evident in simulation-experiment coupling. Predictors accelerate property forecasting from simulations, but graphs integrate experimental feedback to evolve models [23, 24]. DIMC interprets this as a workflow where data  $\rightarrow$  model  $\rightarrow$  discovery pipelines are modulated by infrastructural synergies, optimizing for both speed and depth [6, 7]. In foundation model contexts, the framework suggests interpretive layers where pre-trained predictors embed within dynamic graphs, enhancing multimodal adaptability [21, 22].

Another implication addresses computational workflow dynamics in high-throughput settings. DIMC elucidates how predictors handle volume-intensive tasks, while graphs manage complexity through semantic layering [8, 10]. A conceptual expression for this trade-off is  $\overset{T}{=} \kappa \left( \frac{V_p}{C_g} \right)$ , with  $V_p$  as predictor volume capacity,  $C_g$  graph complexity overhead, and  $\kappa$  trade-off modulator. This interprets equilibrium points where hybrid deployments maximize intelligence output [2, 18].

Overall, these implications interpret DIMC as a lens for engineering resilient materials ecosystems, where competing infrastructures converge to steer innovation.

## Results and Discussion

The interpretive scope of DIMC extends to broader field dynamics, illuminating how knowledge graphs and property predictors redefine computational paradigms in materials engineering. In representation learning, the framework underscores the shift from static descriptors to dynamic, infrastructure-dependent encodings [3, 7, 12]. Graphs enrich representations with relational depth, enabling inference across datasets that predictors might treat independently [5, 14]. This discussion interprets such interactions as foundational to addressing data sparsity, common in emerging materials like perovskites or alloys [4, 19].

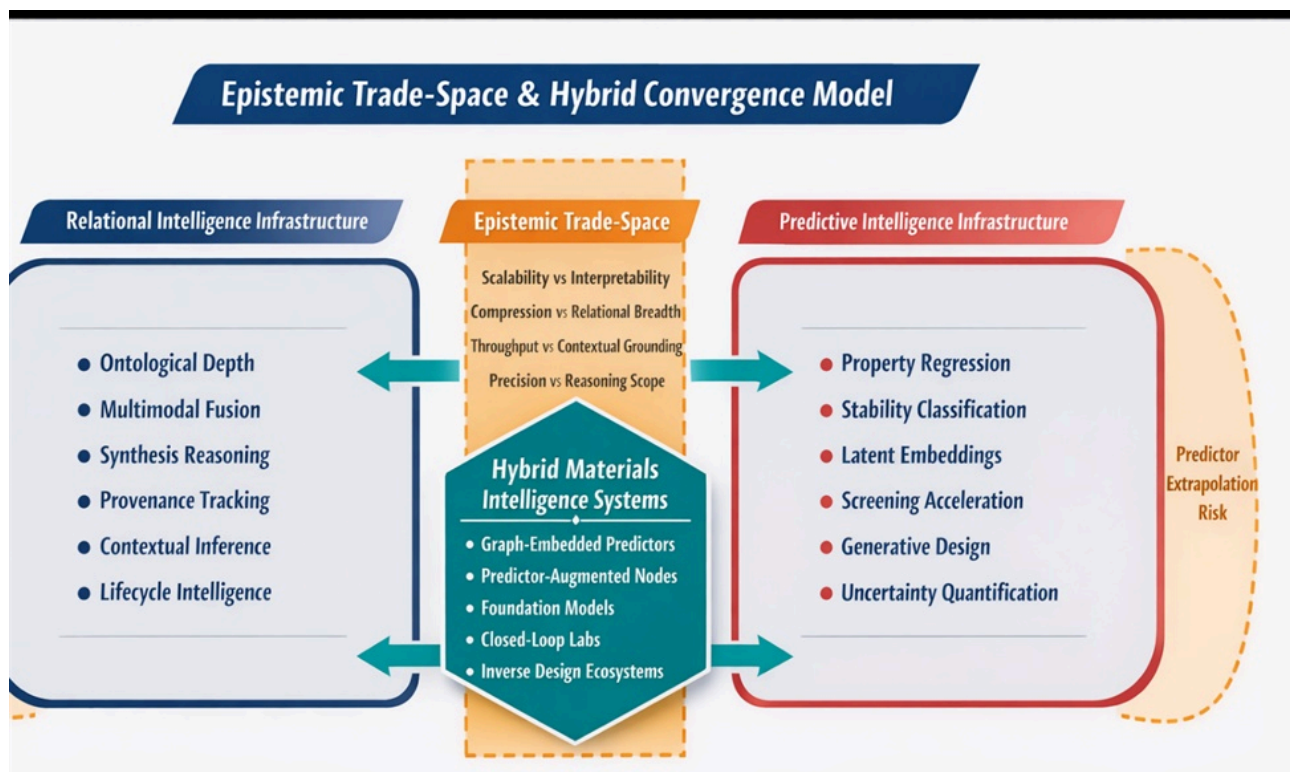
In terms of discovery steering, DIMC highlights logics that balance exploration and exploitation. Predictors favor exploitation through precise mappings, but graphs enable exploratory traversals via linked entities [15, 16]. This duality interprets enhanced autonomy in discovery systems, where steering adapts to uncertainty signals, prioritizing high-value experiments [11, 13, 17]. For inverse design, the framework discusses how predictor-driven generation benefits from graph constraints, interpreting a pathway to more feasible candidates without exhaustive searches [18, 21].

Epistemic risk structures warrant discussion, as DIMC reveals vulnerabilities in isolated infrastructures. Standalone predictors risk extrapolation errors in unseen regimes, while graphs may propagate outdated relations [20, 23]. Hybrid interpretations via DIMC suggest risk mitigation through feedback, where predictor updates refresh graph knowledge [9, 10]. This discusses a move toward epistemic robustness, crucial for trustworthy AI in materials [2, 6].

Computational workflow dynamics are reinterpreted under DIMC, particularly in multimodal integration. Graphs serve as hubs for fusing simulation and experimental data, augmenting predictor inputs [22, 24]. Discussion points to implications

for foundation models, where DIMC interprets layered cognition: predictors as core engines, graphs as contextual scaffolds [1, 8]. This fosters cohesive ecosystems, bridging silos in materials informatics.

Trade-offs in infrastructure scalability are central. Predictors scale with parallel computing, but graphs demand efficient storage and querying [3, 7]. DIMC discusses optimizations, such as embedding predictors within graphs for hybrid efficiency [14, 19]. In closed-loop contexts, this interprets iterative refinements that accelerate discovery cycles (Figure 2).



**Figure 2.** Epistemic trade-space and hybrid convergence between knowledge graph and property predictor infrastructures.

The diagram maps competing and complementary intelligence functions across relational and predictive systems. While predictors optimize throughput and quantitative precision, knowledge graphs enable contextual reasoning and multimodal integration. Hybrid architectures emerge within the convergence zone, where bidirectional feedback mitigates epistemic risk and enhances discovery resilience.

Uncertainty quantification dynamics offer further discussion. DIMC interprets integrated approaches where graph-propagated uncertainties inform predictor ensembles, enhancing calibration [11-13]. This has bearings on simulation-experiment coupling, discussing streamlined validations [17, 20].

Ultimately, DIMC's interpretations discuss a paradigm where competing infrastructures evolve into symbiotic systems, driving sustainable advancements in data-driven materials engineering [4, 5, 18].

## Conclusion

The DIMC framework provides an interpretive structure for understanding knowledge graphs and property predictors as competing infrastructures in materials intelligence. By delineating representational, inferential, and steering layers, it

elucidates dynamics that enhance computational workflows and mitigate epistemic risks. Insights into feedback loops and trade-offs underscore the potential for hybrid ecosystems, fostering integrated discovery pipelines. This conceptual lens advances the field toward more adaptive, robust systems for materials innovation, bridging predictive precision with relational depth.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 12 Jul 2023

Revised: 27 Nov 2023

Accepted: 15 Dec 2023

Published online: 18 March 2024

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ramprasad R, Batra R, Pilania G, Mannodi-Kanakithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. *npj Comput Mater.* 2017;3(1):54.
- Pilania G. Machine learning in materials science: From explainable predictions to autonomous design. *Comput Mater Sci.* 2021;193:110360.
- Jablonka KM, Ongari D, Moosavi SM, Smit B. Big-data science in porous materials: Materials genomics and machine learning. *Chem Rev.* 2020;120(16):8066-129.
- Liu Y, Zhao T, Ju W, Shi S. Materials discovery and design using machine learning. *J Mater.* 2017;3(3):159-77.

Chen L, Tran H, Batra R, Kim C, Ramprasad R. Machine learning models for the prediction of energy, forces, and stresses for platinum. *npj Comput Mater.* 2019;5(1):111.

Dunn A, Wang Q, Ganose A, Dopp D, Jain A. Benchmarking materials property prediction methods: The Matbench test set and Automatminer reference algorithm. *npj Comput Mater.* 2020;6(1):138.

Lv C, Chen C, Chen T, Li Q, Liu H, Zhao L, et al. Machine learning: An advanced platform for materials development and state prediction in lithium-ion batteries. *Adv Mater.* 2022;34(25):2101474.

Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G. Gaussian process regression for materials and molecules. *Chem Rev.* 2021;121(16):10073-141.

Noh J, Gu GH, Kim S, Jung Y. Machine-enabled inverse design of inorganic solid materials: Promises and challenges. *Chem Sci.* 2020;11(19):4871-81.

Yan Q, Yu S, Gong W, Zhao Z, Zhang Y, Gong S, et al. Accelerating materials discovery using integrated deep machine learning approaches. *J Mater Chem A.* 2023;11(10):5084-94.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. *Nature.* 2023;624(7990):80-5.

Vincent-Cuaz C, Flamaray R, Corneli M, Vayer T, Courty N. Semi-relaxed Gromov-Wasserstein divergence with applications on graphs. *arXiv preprint arXiv:2110.02753.* 2021 Oct 6.

Fung V, Hu J, Ganesh P, Sumpter BG. Machine learned features from density of states for accurate adsorption energy prediction. *Nat Commun.* 2021;12(1):88.

Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions. *npj Comput Mater.* 2021;7(1):173.

Chen C, Ye W, Zuo Y, Zheng C, Ong SP. Graph networks as a universal machine learning framework for molecules and crystals. *Chem Mater.* 2019;31(9):3564-72.

Xie T, Grossman JC. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys Rev Lett.* 2018;120(14):145301.

Park CW, Wolverton C. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys Rev Mater.* 2020;4(6):063801.

Chen C, Zuo Y, Ye W, Li X, Ong SP. Learning properties of ordered and disordered materials from multi-fidelity data. *Nat Comput Sci.* 2021;1(1):46-53.

Lu S, Zhou Q, Ouyang Y, Guo Y, Li Q, Wang J. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat Commun.* 2018;9(1):3405.

Goodall REA, Lee AA. Predicting materials properties without crystal structure: Deep representation learning from stoichiometry. *Nat Commun.* 2020;11(1):6280.

Schmidt J, Shi J, Borlido P, Chen L, Botti S, Marques MAL. Predicting the thermodynamic stability of solids with machine learning. *Chem Mater.* 2017;29(12):5090-103.

Jelenic S, Bindics J, Czermak P, Pillai BR, Ruer M, Hoege C, et al. Composition can buffer protein dynamics within liquid-like condensates. *bioRxiv.* 2022:2022-05.

DeCost B, Hattrick-Simpers JR, Trautt Z, Kusne AG, Campo E, Green ML. Scientific AI in materials science: Apath to a sustainable and scalable paradigm. *Mach Learn Sci Technol.* 2020;1(3):033001.

Lin HY, Lu YJ, Chou HH, Ou KL, Huang BH, Lan WC, et al. Biomimetic ceramic composite: characterization, cell response, and in vivo biocompatibility. *Materials.* 2021;14(23):7374.

Wei J, Chu X, Sun XY, Xu K, Deng HX, Chen J, et al. Machine learning in materials science. *InfoMat.* 2019;1(3):338-58.

Li Y, Li C, Zhou K, Zhang C, Yang J, Zhang X, et al. Methods, progresses, and opportunities of materials informatics. *InfoMat.* 2023;5(3):e12425.

Chan MH, Chan CH, Liu M, Li J, Li X, Hu H, et al. Application of machine learning for advanced material prediction and design. *EcoMat.* 2022;4(4):e12194.

Shi S, Zuo Y, Li X, Li H, Guo X, Zhu X, et al. A review on the applications of graph neural networks in materials science at the atomic scale. *Mater Genome Eng Adv.* 2024;2(1):e50.

Shen ZH, Liu HX, Shen Y, Hu JM, Chen LQ, Nan CW. Machine learning in energy storage materials. *Interdiscip Mater.* 2022;1(2):175-95.