ORIGINAL RESEARCH **Open access**

The Coordination Problem in Multi-Model Materials AI Pipelines

Li Zhang[1*], Wei Chen[1]

Abstract

Materials science increasingly relies on artificial intelligence (AI) pipelines that integrate multiple models of varying fidelities, architectures, and objectives to accelerate discovery and design. These multi-model workflows—encompassing low-fidelity approximations, high-fidelity simulations, machine learning surrogates, and experimental feedback—promise efficiency but introduce a fundamental coordination problem: reconciling disparate predictions, managing conflicts, ensuring interoperability, and mitigating emergent behaviors or bottlenecks. This conceptual manuscript examines the coordination challenges in such pipelines, drawing on recent advances in multi-fidelity learning, active learning, hybrid modeling, and workflow orchestration. It analyzes how integration conflicts arise from differences in scale, accuracy, and data provenance, potentially leading to consensus failures, validation cascades, and optimization bottlenecks. The discussion highlights conceptual strategies for robust coordination, including uncertainty-aware fusion, adaptive sampling, and iterative refinement, while underscoring the need for principled frameworks to harness the full potential of multi-model systems in materials AI.

Keywords Coordination problem, Multi-fidelity systems, Epistemic uncertainty, Model interoperability, Emergent behavior, Validation cascades

*Correspondence:
Li Zhang
li.zhang@outlook.com
[1] Department of Computational Materials Science, School of Materials Engineering, Tsinghua University, Beijing, China

Introduction

The advent of machine learning (ML) and artificial intelligence (AI) has profoundly reshaped materials discovery, displacing traditionally sequential, intuition-driven, and resource-intensive workflows with high-throughput, data-centric paradigms [1, 2]. Rather than relying on isolated computational or experimental steps, contemporary materials research increasingly operates through integrated pipelines in which multiple computational models interact to generate, evaluate, and refine candidate materials at scale. These pipelines promise accelerated discovery, broader exploration of chemical and structural spaces, and improved utilization of computational and experimental resources.

A defining feature of modern materials AI pipelines is their reliance on heterogeneous model ensembles spanning multiple fidelity levels. Low-cost approximations—such as semi-empirical methods, heuristic descriptors, or simplified physical proxies—enable rapid screening across vast candidate spaces. Mid-fidelity models, including graph neural networks and other representation-learning architectures, provide scalable property prediction and ranking capabilities. High-fidelity approaches, most notably first-principles ab initio calculations or carefully controlled experiments, serve as validation anchors and sources of epistemic authority [3, 4]. In principle, this stratified architecture exploits complementary strengths: computational efficiency from surrogates, expressive power from learned representations, and physical rigor from physics-based methods.

Yet the promise of such multi-model architectures introduces a fundamental conceptual challenge: how to coordinate disparate models into a coherent scientific workflow. Coordination extends beyond simple data exchange or pipeline automation. It encompasses alignment across model assumptions, representational choices, uncertainty semantics, optimization objectives, and interpretive roles within the discovery process. When models differ in fidelity, scope, or inductive bias, their outputs are not trivially commensurable. Instead, coordination requires explicit or implicit rules for how predictions are compared, reconciled, prioritized, or overridden—rules that are often underspecified or tacitly embedded in workflow design.

The coordination problem manifests most clearly when heterogeneous models generate conflicting or inconsistent signals. Differences in training domains, feature representations, or physical constraints can yield divergent rankings, incompatible uncertainty estimates, or contradictory recommendations for downstream validation. Errors introduced at low-fidelity stages may propagate silently, shaping candidate selection in ways that are difficult to detect or reverse once high-fidelity resources are engaged [5, 6]. Conversely, high-fidelity models may invalidate large regions of the search space favored by surrogates, revealing latent misalignments rather than incremental refinements. Such dynamics complicate not only accuracy but also interpretability, trust, and decision-making within the pipeline.

Beyond prediction inconsistencies, coordination failures give rise to a range of secondary system-level issues. These include orchestration inefficiencies arising from mismatched computational costs or scheduling constraints; interoperability barriers between models built on incompatible data schemas or software ecosystems; consensus failures when ensemble members disagree without a principled resolution mechanism; validation cascades in which early approximation errors amplify downstream; and emergent behaviors driven by nonlinear feedback between models, data acquisition, and optimization loops [7, 8]. In extreme cases, pipelines may converge prematurely on narrow regions of material space, not because of genuine scientific promise but because of structural biases embedded in coordination logic.

Importantly, these challenges cannot be reduced to implementation details or engineering optimization alone. They reflect deeper epistemic questions about how knowledge claims are constructed, compared, and legitimized when no single model provides ground truth. As materials AI systems increasingly operate under conditions of data scarcity, extrapolation, and open-ended exploration, coordination becomes a central determinant of scientific reliability rather than a peripheral concern.

This manuscript addresses the coordination problem from a purely conceptual perspective, deliberately avoiding empirical methods or performance evaluation. The analysis focuses on how coordination failures arise, how they shape interpretive outcomes, and why they pose persistent obstacles to robust materials discovery. Building on recent literature on multi-fidelity integration [9, 10], active and transfer learning within discovery workflows [11, 12], and the emergence of autonomous or semi-autonomous materials pipelines [13, 14], the paper synthesizes existing insights into a coherent conceptual account. By reframing coordination as an epistemic and systems-level challenge, this work aims to clarify the conditions under which multi-model materials AI pipelines succeed—or fail—to function as reliable engines of scientific discovery. **Figure 1** schematically illustrates the coordination problem in multi-model materials AI pipelines, highlighting how heterogeneous fidelity levels are integrated through an orchestration layer and how coordination failures give rise to consensus breakdowns, validation cascades, and emergent system-level behaviors.

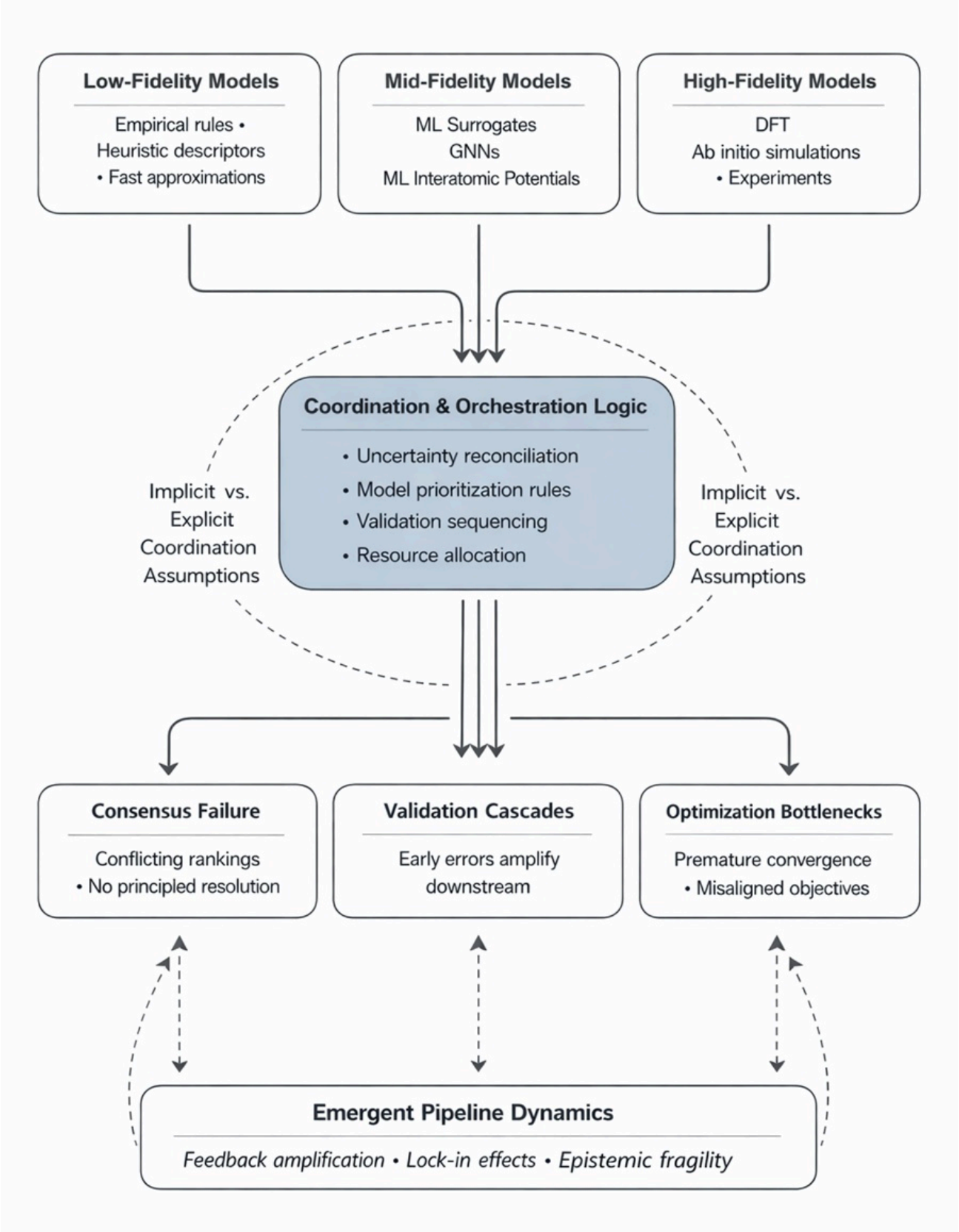

Figure 1. Conceptual diagram of the coordination problem in multi-model materials AI pipelines

Theoretical Background

The coordination problem in multi-model materials AI pipelines originates from a fundamental tension between model heterogeneity and workflow integration. Contemporary materials discovery systems routinely combine multiple modeling paradigms—ranging from low-fidelity empirical or semi-empirical approximations, through mid-fidelity graph neural networks (GNNs) and machine-learning interatomic potentials (MLIPs), to high-fidelity first-principles methods such as density functional theory (DFT)—within a single iterative discovery loop [1, 2]. Each modeling class embodies distinct assumptions about physical representation, data dependence, uncertainty structure, and computational cost. While this diversity enables broad exploration and targeted refinement, it also complicates aligning model outputs into a unified inferential process.

Theoretical foundations for such multi-model systems draw primarily from multi-fidelity modeling, ensemble learning, and hierarchical probabilistic frameworks. In classical multi-fidelity learning, models are arranged in a stratified hierarchy in which lower-fidelity surrogates enable rapid, large-scale screening, and higher-fidelity models are selectively deployed to refine predictions in promising regions [3, 4]. This architecture formalizes an explicit trade-off between computational efficiency and predictive accuracy. However, it also presumes that fidelity levels are meaningfully comparable and that information can be transferred upward or downward in the hierarchy without distortion—assumptions that are often violated in practice due to mismatched representations or domain shifts.

Ensemble learning provides an alternative theoretical lens, treating multiple models as parallel hypothesis generators whose collective behavior can outperform any single constituent. Ensembles mitigate variance, hedge against model misspecification, and offer robustness under uncertainty. Yet in materials AI pipelines, ensembles rarely operate as symmetric peers. Instead, models differ substantially in epistemic status: a DFT calculation is typically treated as authoritative, while ML predictions are provisional. This asymmetry complicates standard ensemble fusion strategies and raises questions about how confidence, credibility, and evidentiary weight should be assigned across models operating at different fidelities.

A closely related theoretical pillar concerns the representation, propagation, and reconciliation of uncertainty. Disagreement among models may arise from divergent training datasets, incompatible feature spaces, architectural inductive biases, or differing physical approximations [5, 6]. From a statistical decision-theoretic perspective, such disagreement constitutes a coordination dilemma: the system must determine how to aggregate, prioritize, or defer conflicting outputs. Bayesian model averaging, weighted ensemble fusion, and active learning strategies provide formal mechanisms for addressing disagreement, often reframing it as a signal of epistemic uncertainty rather than mere noise [7]. However, these approaches typically assume well-calibrated uncertainty estimates and shared probabilistic semantics—conditions that are difficult to satisfy when combining physics-based and data-driven models.

Beyond probabilistic considerations, coordination must also address alignment with optimization. Individual models are often trained or tuned to optimize local objectives—such as prediction accuracy, energy minimization, or uncertainty reduction—without explicit consideration of system-level goals. As a result, locally optimal behavior can yield globally suboptimal outcomes, such as premature convergence to narrow regions of material space or inefficient allocation of high-fidelity resources. This misalignment underscores the need for coordination mechanisms that operate not only at the level of predictions but also at the level of objectives and decision rules.

From a systems-theoretic standpoint, multi-model materials AI pipelines can be understood as complex adaptive systems. Their behavior arises from nonlinear interactions among models, data-acquisition processes, and decision policies rather than from any single component in isolation [8, 9]. Feedback loops—where predictions guide experiments that, in turn, reshape the training data—can amplify initial biases or errors. Cascading failures may occur when inaccuracies at low-fidelity stages propagate downstream, influencing candidate selection well before validation. Optimization bottlenecks may arise when computational or data constraints interact with rigid coordination logic, limiting adaptability.

Taken together, these theoretical perspectives suggest that coordination in multi-model materials AI pipelines is not merely a technical integration problem but a systems-level epistemic challenge. Effective coordination requires explicit mechanisms for aligning representations, uncertainties, objectives, and interpretive authority across heterogeneous models. Absent such mechanisms, pipelines risk producing internally consistent yet scientifically fragile outcomes—results that reflect the workflow's structure more than the structure of the materials space itself.

The nature of multi-model pipelines in materials AI

Contemporary AI pipelines for materials are explicitly designed to manage the vast combinatorial space of chemical, structural, and compositional properties by integrating diverse modeling paradigms within a single discovery architecture. Rather than relying on a monolithic predictive model, these pipelines decompose the discovery task into stages that differ in scope, fidelity, and computational cost. High-throughput screening typically begins with low-fidelity models—such as empirical rules, heuristic descriptors, or simplified surrogates—that enable rapid elimination of implausible candidates across large search spaces. Promising subsets are then subjected to progressively higher-fidelity evaluation using machine-learned surrogates or physics-based simulations [15, 16].

Hybrid pipelines further complicate this structure by combining distinct representational and generative components. Graph-based machine learning models are commonly used to encode atomic connectivity and local environments. At the same time, density functional theory (DFT) and related electronic-structure methods provide physically grounded estimates of energetics, stability, and electronic properties. In parallel, generative models—such as variational autoencoders, diffusion models, or reinforcement learning agents—may propose novel compositions or structures that lie outside existing datasets [17, 18]. These components do not merely operate sequentially; they interact iteratively, with outputs from one stage reshaping the inputs, objectives, or constraints of another.

As a result, multi-model pipelines exhibit an inherently hierarchical, bidirectional, multi-fidelity structure. Information flows upward as low-fidelity predictions guide the allocation of high-cost computational or experimental resources, and downward as high-fidelity validations recalibrate surrogate models through transfer learning, active learning, or dataset augmentation [19, 20]. This bidirectional flow is central to pipeline efficiency and adaptability, allowing discovery strategies to evolve dynamically rather than following a fixed evaluation order.

However, this orchestration introduces substantial coordination demands. Models must exchange information across incompatible data formats, representational spaces, and uncertainty semantics. Decisions made at each stage—such as which candidates to advance, which regions of space to explore, or when to terminate search—implicitly encode coordination logic that shapes the overall epistemic trajectory of the pipeline. When such logic is underspecified or misaligned, the pipeline may prioritize efficiency at the expense of reliability, or convergence at the expense of exploration. Thus, the structure that enables scalability also creates fragility, making coordination a central determinant of scientific robustness.

Core coordination challenges

Integration and conflict between disparate models

A primary challenge in multi-model materials AI pipelines is integrating models that differ fundamentally in their physical assumptions, representational choices, length and time scales, and data modalities. Low-fidelity models may rely on coarse descriptors or simplified physics that systematically distort certain regions of materials space. At the same time, high-fidelity methods encode detailed interactions at substantially higher computational cost. Without explicit correction or calibration mechanisms, biases introduced early in the pipeline can steer downstream exploration in subtle but persistent ways. **Table 1** summarizes the primary coordination challenges in multi-model materials AI pipelines, their sources of tension, and their downstream epistemic consequences at the system level.

Table 1. Coordination challenges and epistemic consequences in multi-model materials AI pipelines

Coordination dimension	Source of tension	Manifestation in pipelines	Epistemic consequence	Illustrative mitigation logic (conceptual)
Model heterogeneity	Divergent fidelities, representations, and assumptions	Conflicting predictions and rankings across models	Ambiguous knowledge claims; unclear inferential authority	Explicit fidelity hierarchies; discrepancy-aware fusion
Uncertainty semantics	Incompatible or poorly calibrated uncertainty estimates	Inconsistent confidence signals across stages	False consensus or misplaced trust	Uncertainty normalization and propagation rules
Objective misalignment	Local optimization targets differ across models	Surrogates optimize proxies misaligned with validation goals	Premature convergence; distorted discovery priorities	System-level objective alignment and adaptive weighting
Workflow orchestration	Incompatible data formats and software ecosystems	Fragile integration; hidden preprocessing assumptions	Reduced reproducibility and interpretability	Modular orchestration with provenance tracking
Consensus formation	Lack of principled disagreement resolution	Heuristic averaging or model privileging	Masked uncertainty; epistemic opacity	Disagreement-as-signal decision frameworks
Validation sequencing	Early low-fidelity errors propagate downstream	Late detection of invalid candidates	Validation cascades and wasted resources	Active learning with uncertainty-triggered escalation
Feedback coupling	Strong nonlinear interaction between models and data	Self-reinforcing biases and lock-in	Emergent epistemic fragility	Controlled feedback strength and damping mechanisms

Model disagreement is an especially acute manifestation of this challenge. Ensemble members or fidelity tiers may produce conflicting rankings, incompatible uncertainty estimates, or divergent recommendations for candidate selection [21, 22]. Such conflicts are not merely technical inconveniences; they represent epistemic fractures within the pipeline. When disagreement lacks a principled resolution strategy, pipelines may stall, default to heuristic decision rules, or implicitly privilege certain models based on convention rather than justification.

Conceptual frameworks such as multi-fidelity Gaussian processes, hierarchical surrogate modeling, or denoising-based fusion approaches attempt to address this issue by explicitly modeling inter-model discrepancies as structured signals rather than treating them as noise [9, 23]. In these formulations, disagreement becomes informative, revealing regions where model assumptions break down or where additional information is most valuable. However, these approaches presuppose that discrepancies are statistically learnable and that models can be aligned within a shared probabilistic or representational framework—assumptions that are often strained when combining data-driven and physics-based methods.

The stakes of unresolved integration conflict are particularly high in inverse design and goal-directed discovery. In such settings, candidate generation, property prediction, and validation must remain tightly aligned with target specifications. Misalignment across fidelities can lead to candidates that satisfy surrogate objectives but fail under high-fidelity evaluation, resulting in wasted computational effort and distorted assessments of design feasibility. Over time, these failures can induce feedback loops in which pipelines overfit to regions where coordination is easier rather than to those where scientific opportunity is greatest.

More broadly, integration conflicts expose a deeper issue: coordination mechanisms frequently operate implicitly, embedded in workflow design choices rather than articulated as explicit epistemic rules. As a consequence, pipelines may appear internally consistent while systematically privileging certain representations, fidelities, or objectives. Understanding and addressing integration conflict, therefore, requires not only improved fusion algorithms but also conceptual clarity about how heterogeneous models are intended to jointly support scientific inference.

Workflow orchestration and interoperability

Workflow orchestration in multi-model materials AI pipelines extends beyond the sequential execution of computational steps. It involves the deliberate coordination of models with heterogeneous inputs and outputs, the management of data provenance across iterative cycles, and the maintenance of semantic consistency as information traverses the pipeline. Orchestration decisions determine not only which models are executed and when, but also how their outputs are interpreted, transformed, and reintegrated into subsequent stages.

A persistent challenge lies in interoperability. Models across fidelity levels often rely on incompatible data formats, feature representations, software ecosystems, and metadata standards. In the absence of widely adopted interfaces or ontologies, integration frequently depends on ad hoc data transformations and custom glue code. While such mappings enable short-term functionality, they introduce hidden assumptions about equivalence between representations, obscure provenance, and increase the risk of silent errors that are difficult to trace or correct [24, 25]. Over time, these fragilities accumulate, reducing reproducibility and undermining confidence in pipeline outputs.

Recent platforms and workflow frameworks—such as integrated hubs that centralize preprocessing, feature engineering, model execution, and result aggregation—represent important steps toward unified orchestration [15]. These systems aim to modularize discovery workflows, allowing components to be swapped or updated without redesigning the entire pipeline. However, most current approaches emphasize computational convenience and scalability rather than epistemic coherence. In particular, uncertainty information is often flattened, truncated, or discarded as predictions move between modules, breaking the chain of inferential accountability.

Conceptually, scalable orchestration requires treating workflows as composable decision systems rather than linear data pipelines. This entails explicit coordination protocols governing how uncertainty is propagated, how confidence thresholds trigger transitions between fidelity levels, and how conflicting signals are escalated for resolution. Without such protocols, orchestration logic becomes implicit, embedded in implementation details rather than articulated as part of the scientific methodology. As a result, pipelines may appear operationally robust while remaining epistemically opaque.

Consensus failure and validation cascades

Consensus failure arises when multiple models within a pipeline produce conflicting recommendations that cannot be reconciled through simple aggregation. For example, one surrogate model may prioritize thermodynamic stability, while another emphasizes functional performance or synthesizability, leading to incompatible rankings of candidate materials [15]. Such conflicts reflect not only predictive disagreement but also bigger differences in objective functions, inductive biases, and implicit value judgments encoded within models.

In many pipelines, consensus is resolved heuristically—by privileging a particular model class, averaging scores, or applying manually defined rules. While expedient, these approaches often mask uncertainty rather than resolving it, creating an illusion of agreement. More critically, unresolved disagreement can stall discovery by preventing clear advancement decisions or, conversely, force premature commitment to candidates that satisfy coordination logic rather than scientific merit.

Validation cascades represent a related but distinct failure mode. When low-fidelity models introduce systematic errors or miscalibrated confidence early in the workflow, these inaccuracies can cascade downstream, shaping candidate selection long before high-fidelity validation occurs [17]. By the time discrepancies are detected, substantial computational or

experimental resources may already have been expended. In this sense, validation cascades amplify the epistemic consequences of early-stage coordination failures, transforming minor approximation errors into high-stakes decision risks.

Active learning and related adaptive sampling strategies offer a conceptual mitigation pathway by reframing discovery as an iterative uncertainty management process rather than a one-directional filtering operation [11, 18]. In this view, disagreement and uncertainty are treated as signals that guide the selective acquisition of high-fidelity information, dynamically reallocating resources to regions where coordination is weakest. However, the effectiveness of such strategies depends on how uncertainty is defined, compared, and prioritized across heterogeneous models—once again returning coordination to the center of the problem.

Taken together, consensus failure and validation cascades highlight a key insight: coordination mechanisms determine not only computational efficiency but also the epistemic trajectory of materials discovery. Pipelines lacking principled strategies for resolving disagreement and containing error-propagation risk converge on results that are operationally optimized yet scientifically fragile. Addressing these issues, therefore, requires explicit conceptual frameworks for consensus formation, validation sequencing, and uncertainty-aware decision-making across multi-model systems.

Emergent system behavior

As multi-model materials AI pipelines evolve from linear workflows into tightly coupled, iterative systems, their behavior increasingly reflects the dynamics of complex adaptive systems rather than simple computational chains. Nonlinear interactions among heterogeneous models—each with distinct update rules, learning dynamics, and uncertainty structures—can give rise to emergent phenomena that are not predictable from any single component in isolation. These include unexpected synergies between models, instability in candidate rankings, or abrupt phase-like transitions in predicted material properties as pipeline parameters or feedback intensities change [3].

A primary driver of such emergence is the presence of feedback loops. Predictions from surrogate models guide sampling strategies, candidate selection, or generative proposals, which in turn reshape training datasets and model updates. While feedback is essential for efficiency and adaptivity, it can also amplify latent biases. For example, early surrogate errors may skew sampling toward narrow regions of materials space, reinforcing artifacts rather than genuine structure–property relationships. Over successive iterations, pipelines may become increasingly confident in systematically distorted predictions, exhibiting a form of self-reinforcing overfitting that is difficult to detect through conventional validation [2].

From a conceptual standpoint, emergent behavior depends critically on the coupling strength between pipeline components. Strong coupling—where small changes in one model rapidly propagate throughout the system—can accelerate convergence but also increase the system's susceptibility to instability and lock-in. Weak coupling promotes robustness and diversity but may sacrifice efficiency or delay discovery. Achieving emergent robustness, therefore, requires explicit control over how tightly models interact, including thresholds for feedback activation, damping mechanisms to prevent uncertainty amplification, and constraints on how rapidly models adapt to newly acquired data.

Viewing materials AI pipelines through the lens of complex adaptive systems reframes coordination as a problem of system-level stability and resilience. Robust discovery does not emerge automatically from model accuracy alone; it depends on how interactions are structured, how feedback is regulated, and how diversity of hypotheses is maintained over time. Without such considerations, pipelines risk converging toward internally consistent yet scientifically brittle equilibria.

Optimization bottlenecks

Despite their promise, multi-model materials AI pipelines are constrained by persistent optimization bottlenecks arising from computational expense, limited data availability, and sensitivity to model and hyperparameter choices. High-fidelity

methods such as first-principles simulations impose strict resource constraints, while low-fidelity models may provide misleading guidance if their limitations are not adequately accounted for. These tensions create bottlenecks that shape not only efficiency but also the epistemic trajectory of discovery.

Multi-fidelity strategies are conceptually motivated as a solution to these constraints, leveraging inexpensive approximations to guide selective investment in high-cost evaluations [5, 10]. However, optimization across fidelity levels introduces its own challenges. Poorly calibrated transitions between fidelities can trap pipelines in locally optimal regions of materials space, privileging candidates that perform well under surrogate objectives but fail to generalize under more rigorous evaluation. Similarly, rigid allocation strategies may overexploit familiar domains while underexploring uncertain but potentially high-impact regions.

Optimization bottlenecks are further compounded by objective misalignment. Different models may implicitly optimize different criteria—accuracy, stability, novelty, or uncertainty reduction—without a unifying system-level objective. As a result, optimization dynamics may become fragmented, with individual components improving locally while overall discovery efficiency stagnates.

Bayesian optimization, reinforcement learning, and related sequential decision-making frameworks offer conceptual pathways for addressing these challenges by reframing discovery as an adaptive resource allocation problem [8]. Rather than maximizing a single performance metric, these approaches prioritize informative queries, explicitly balancing exploitation of known promising regions against exploration of uncertain or poorly characterized domains. Crucially, their effectiveness depends on coherent uncertainty modeling and principled coordination across fidelity levels.

Taken together, optimization bottlenecks underscore that efficiency gains in materials AI are inseparable from the quality of coordination. Without explicit mechanisms to align objectives, regulate feedback, and adaptively allocate resources, pipelines risk expending computational effort without commensurate epistemic return. Addressing these bottlenecks, therefore, requires not only more sophisticated optimization algorithms but also conceptual clarity about how optimization goals are defined, compared, and integrated across heterogeneous models.

Conceptual strategies for robust coordination

Effective coordination demands uncertainty quantification, adaptive fusion, and iterative refinement. Multi-fidelity co-kriging or transfer learning bridges fidelities without full high-cost data [7, 19]. Active learning with integrated variance sampling enhances generalizability by targeting informative regions [1]. Hybrid workflows—combining physics-informed constraints with data-driven surrogates—reduce conflicts by grounding ML in domain knowledge [11].

Future conceptual directions include agentic frameworks where models negotiate via communication protocols, or foundation models fine-tuned for multi-fidelity reasoning [13]. Emphasis on provenance tracking and explainable fusion could prevent cascades and emergent pitfalls.

Results and Discussion

The coordination problem constitutes one of the most significant conceptual barriers to achieving scalable, autonomous materials discovery. While multi-model pipelines offer substantial theoretical advantages in balancing accuracy and efficiency, their practical effectiveness is severely constrained by unresolved integration conflicts, consensus failures, and orchestration inefficiencies [1, 3, 8].

Current approaches — such as multi-fidelity Gaussian processes, active learning, and hybrid physics-informed ML — provide partial mitigation but remain fundamentally limited. Most existing strategies address coordination reactively rather than proactively. They often rely on heuristic fusion rules or post-hoc reconciliation rather than on designing coordination mechanisms into the pipeline's core architecture [6, 12].

A critical analytical insight is that coordination is not merely a technical issue of data fusion but a deeper architectural and epistemological challenge. Different models embody different assumptions about reality (e.g., continuum vs. atomistic, empirical vs. physics-based), making perfect alignment theoretically difficult. This raises important questions about the limits of model pluralism in materials science [9, 13].

Looking forward, more sophisticated coordination strategies are needed, including agent-based negotiation frameworks, uncertainty-aware orchestration platforms, and foundation models specifically trained for multi-fidelity reasoning. The field must move from loosely coupled ensembles toward tightly coordinated, self-correcting systems that can dynamically adapt model weighting, sampling strategies, and validation protocols [14].

Conclusion

The coordination problem in multi-model materials AI pipelines represents a central conceptual and practical challenge in contemporary materials science. As pipelines grow increasingly complex and heterogeneous, the ability to effectively integrate, reconcile, and orchestrate disparate AI models will determine the success or failure of autonomous materials discovery.

This manuscript has examined the theoretical foundations, core coordination challenges, analytical implications, and conceptual strategies related to multi-model integration. Key issues — integration conflicts, consensus failures, validation cascades, emergent behaviors, and optimization bottlenecks — highlight the urgent need for principled coordination frameworks.

Ultimately, overcoming the coordination problem will require shifting from model-centric to workflow-centric design philosophies. Only through deliberate architectural focus on interoperability, uncertainty propagation, adaptive consensus, and robust orchestration can materials AI realize its full transformative potential.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 15 Apr 2021 Revised: 05 Jun 2021 Accepted: 12 Jul 2021
Published online: 18 January 2022

References

DeCost BL, Hattrick-Simpers JR, Trautt Z, Kusne AG, Campo E, Green ML. Scientific AI in materials science: a path to a sustainable and scalable paradigm. Mach Learn Sci Technol. 2020;1(3):032001.

Bai X, Zhang X. Artificial intelligence-powered materials science. Nano-Micro Lett. 2025;17(1):135. https://doi.org/10.1007/s40820-024-01634-8.

Zivic F, Kaplarevic Malisic A, Grujovic N, Stojanovic B, Ivanovic M. Materials informatics: a review of AI and machine learning tools, platforms, data repositories, and applications to architectured porous materials. Mater Today Commun. 2025;48:113525. https://doi.org/10.1016/j.mtcomm.2025.113525.

Back S, Aspuru-Guzik A, Ceriotti M, Gryn'ova G, Grzybowski B, Gu GH, et al. Accelerated chemical science with AI. Digit Discov. 2024;3(1):23-33.

Wang F, Jiang S, Li J. The AI-driven transformation in new materials manufacturing and the development of intelligent sports. Appl Sci. 2025;15(10):5667. https://doi.org/10.3390/app15105667.

Hegi H, Heitz J, Kredel R. Sensor-based augmented visual feedback for coordination training in healthy adults: a scoping review. Front Sports Act Living. 2023;5:1145247. https://doi.org/10.3389/fspor.2023.1145247.

Mohammadi M, Tajik E, Martinez-Maldonado R, Sadiq SS, Tomaszewski W, Khosravi H. Artificial intelligence in multimodal learning analytics: a systematic literature review. Comput Educ Artif Intell. 2025;8:100426. https://doi.org/10.1016/j.caeai.2025.100426.

Zhou H, Xu J, Qin X, Zhang J, Zou W, Shakouri M, et al. Machine learning-driven material intelligence research and development. Nano Res. 2025;18(3):949-59. https://doi.org/10.26599/NR.2025.94908095.

Al-kfairy M, Mustafa D, Kshetri N, Insiew M, Alfandi O. Ethical challenges and solutions of generative AI: an interdisciplinary perspective. Informatics. 2024;11(3):58. https://doi.org/10.3390/informatics11030058.

Uddin M, Arfeen SU, Alanazi F, Hussain S, Mazhar T, Rahman MA. A critical analysis of generative AI: challenges, opportunities, and future research directions. Arch Comput Methods Eng. 2025;32(4):1-25.

Lekadir K, Frangi AF, Porras AR, Glocker B, Cintas CC, Langlotz CP, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. BMJ. 2025;388:e081554. https://doi.org/10.1136/bmj-2024-081554.

Ueda D, Kakinuma T, Fujita S, Kamishima Y, Yanagawa M, Sato J, et al. Fairness of artificial intelligence in healthcare: review and recommendations. Jpn J Radiol. 2024;42(1):3-15. https://doi.org/10.1007/s11604-023-01474-3.

Prabhu DF, Gurupur V, Stone A, Trader E. Integrating artificial intelligence, electronic health records, and wearables for predictive, patient-centered decision support in healthcare. Healthcare. 2025;13(21):2753. https://doi.org/10.3390/healthcare13212753.

Chowdhury SZ, Stevens S, Wu C, Woodward C, Andrews T, Ashall-Payne L, et al. An age-old problem or an old-age problem? A UK survey of attitudes, historical use and recommendations by healthcare professionals to use healthcare apps. BMC Geriatr. 2023;23(1):110. https://doi.org/10.1186/s12877-023-03772-x.

Fang S, Hu YH. Open the door to the atomic world by single-molecule atomic force microscopy. Matter. 2021;4(4):1189-223.

Awuni S, Adarkwah F, Ofori BD, Purwestri RC, Huertas Bernal DC, Hajek M. Managing the challenges of climate change mitigation and adaptation strategies in Ghana. Heliyon. 2023;9(5):e15491. https://doi.org/10.1016/j.heliyon.2023.e15491.

Zhang Y, Xie S, Zeng Z, Tang BZ. Functional scaffolds from AIE building blocks. Matter. 2020;3(6):1862-92. https://doi.org/10.1016/j.matt.2020.09.017.

Ghaffarian S, Taghikhah FR, Maier HR. Explainable artificial intelligence in disaster risk management: achievements and prospective futures. Int J Disaster Risk Reduct. 2023;98:104123. https://doi.org/10.1016/j.ijdrr.2023.104123.

GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2023;402(10397):203-34. https://doi.org/10.1016/S0140-6736(23)01301-6.

Bhore SS, Natraj NA, Hallur GG. Bayesian-driven autonomous defense adaptive consensus optimisation for blockchain networks. Sci Rep. 2025;15(1):2158. https://doi.org/10.1038/s41598-025-31929-8.

Silver P, Furey J, Heiman-Patterson T. Gastrointestinal symptoms in ALS: evidence for enteric nervous system involvement and clinical implications. Muscle Nerve. 2024;70(1):3-9.

Meijboom K, Ansodaria AV, Bamidele N, Eisenberg J, Sontheimer E, Brown R. Advanced base and prime editing strategies to correct common ALS-causing SOD1 mutations. Muscle Nerve. 2025;71(S1):S1-S97.

Zheng H, Luo Z, He K, Zhou W, Kong Z, Dong J, et al. KT-LLM: an evidence-grounded and sequence text framework for auditable kidney transplant modeling. npj Digit Med. 2025;8(1):23. https://doi.org/10.1038/s41746-025-02323-5.

Chakraborty S, Björk J, Dahlqvist M, Rosen J, Heintz F. A survey of AI-supported materials informatics. Comput Sci Rev. 2025;59:100845. https://doi.org/10.1016/j.cosrev.2025.100845.

Uddin M, Rahman MA, Hussain S, Alanazi F, Mazhar T, Arfeen SU. A critical analysis of generative AI: challenges, opportunities, and future research directions. Arch Comput Methods Eng. 2025;32(2):1-20.