ORIGINAL RESEARCH **Open access**

Conceptual Limits of Analogy-Based Reasoning in Generative Materials Models

Carlos Ramirez[1*], Elena Torres[2], Pablo Ortega[1], Sofia Mendes[3]

Abstract

Generative materials models, including variational autoencoders, generative adversarial networks, and diffusion models, have become central to modern artificial intelligence for materials science. Yet, their pervasive reliance on analogy-based reasoning remains largely unexamined and conceptually undertheorized. These models routinely treat latent-space interpolation, transfer learning, and structural substitution as forms of analogical mapping—assuming that what holds between known materials will hold for novel ones—without acknowledging the fundamental epistemological limits of such reasoning. This critical critique identifies four interlocking problems that undermine the reliability of analogy-driven generation: analogy functioning as a substitute for genuine physical understanding, the propagation of false analogies, boundary blindness to domains where analogies break, and the reification of statistical correlations into ontological claims. The consequences of these unacknowledged limits extend beyond technical inaccuracy to wasted experimental resources, overconfident predictions, and a subtle distortion of scientific understanding in materials discovery. Rather than abandoning analogy entirely, this paper argues for hybrid frameworks that explicitly bind analogical transfer with physical invariants, causal verification, and uncertainty quantification. By confronting these conceptual limits head-on, the field can move toward more robust, epistemologically grounded generative models that augment rather than replace mechanistic insight.

Keywords Analogy-based reasoning, Generative materials models, Conceptual limits, False analogies, Boundary blindness, Analogy reification

*Correspondence:
Carlos Ramirez
carlos.ramirez@gmail.com
[1] Department of Intelligent Materials Engineering, University of Barcelona, Barcelona, Spain
[2] Department of Materials Data Analytics, University of Lisbon, Lisbon, Portugal
[3] Department of AI Materials Systems, University of Porto, Porto, Portugal

Introduction

Generative materials models—variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models—now occupy a central position in artificial intelligence for materials science because they promise to accelerate discovery by proposing novel compounds from patterns latent in existing data. At their core, these architectures depend on analogical reasoning. They construct high-dimensional latent spaces in which structural or chemical similarity is encoded as geometric proximity, so that interpolation between two known points is presumed to yield a chemically meaningful third material. Transfer learning similarly imports knowledge from one chemical domain to another on the assumption that the analogy between domains is sufficiently strong. Inverse design workflows further embed this logic: materials that share structural motifs are expected to share property profiles, and element substitution is treated as a straightforward analogical operation. Yet this dependence on analogy is rarely subjected to sustained conceptual scrutiny. The prevailing literature celebrates the practical successes of latent-space navigation and chemical analogy while treating the underlying reasoning mechanism as self-evidently valid [1-9].

The present critique begins from the recognition that analogy is not merely a convenient computational heuristic; it is a form of relational mapping whose validity depends on deep structural alignment between source and target domains [1, 2]. When generative models perform interpolation or substitution, they enact a version of structure-mapping theory without the safeguards that cognitive scientists have long identified as essential. The result is a systematic blind spot: models generate candidates that appear plausible precisely because they preserve surface-level analogies, yet they lack any internal representation of the physical constraints that would invalidate those analogies. This paper, therefore, undertakes a targeted conceptual critique rather than an empirical benchmark. It argues that the unexamined reliance on analogy introduces fundamental limits that cannot be overcome by simply scaling data or model size. Instead, these limits are epistemic: they concern what generative models can legitimately claim to know and what they can safely propose.

Figure 1 conceptualizes the manuscript's central argument by tracing how analogy-based operations in generative materials models translate into four distinct epistemic failure modes, their downstream discovery consequences, and the architectural safeguards required to bound them.

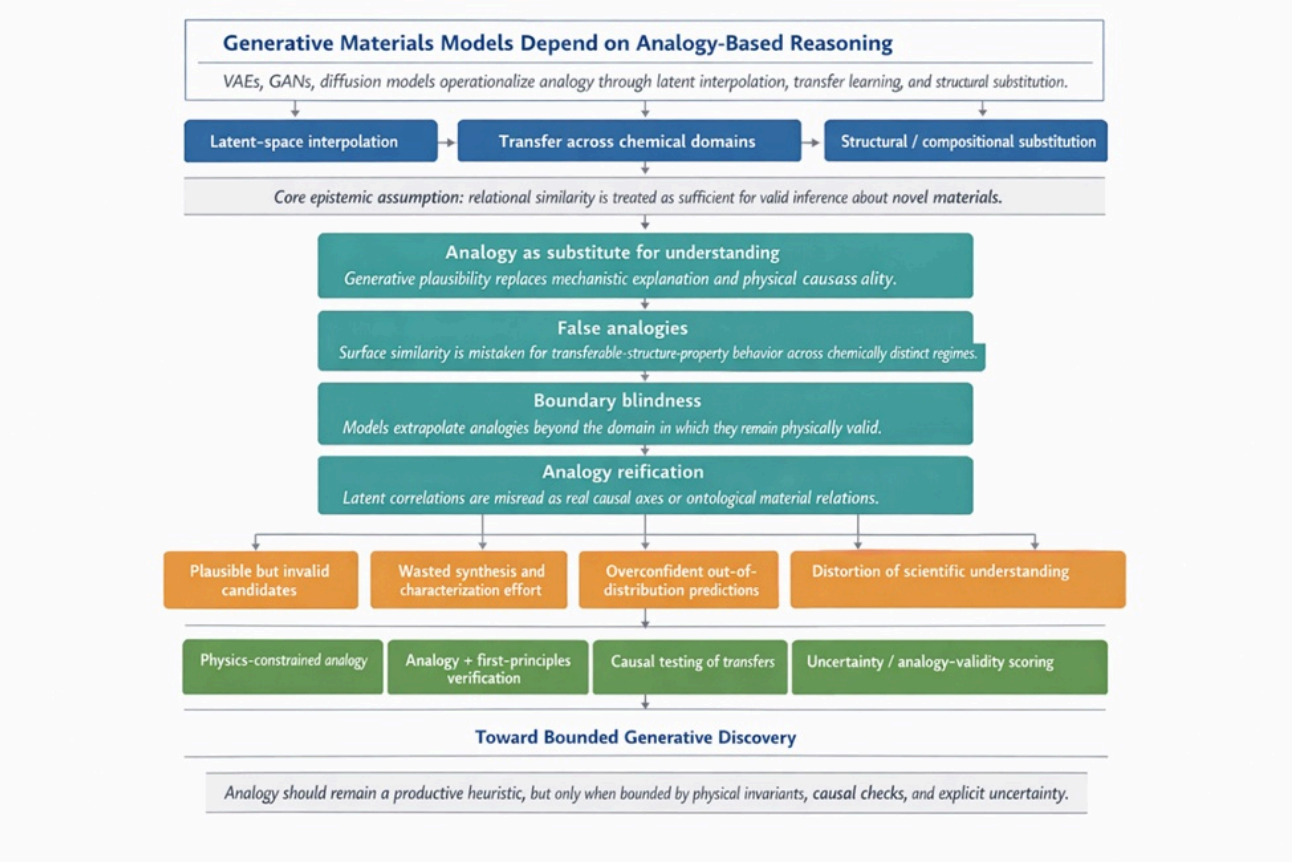

Figure 1. The manuscript's central argument is to trace how analogy-based operations in generative materials models translate into four distinct epistemic failure modes, their downstream discovery consequences, and the architectural safeguards required to bound them.

The critique unfolds in four focused points. First, analogy frequently functions as a substitute for genuine mechanistic understanding, allowing models to produce superficially valid outputs without encoding underlying physics. Second, false analogies proliferate when structural similarity is mistaken for property similarity, especially across chemically distinct regimes. Third, generative architectures exhibit boundary blindness: they extrapolate analogies indefinitely without mechanisms to detect when those analogies cease to hold. Fourth, analogy reification occurs when statistical correlations in latent space are elevated to ontological status, leading researchers to treat model-derived "axes" as real causal

directions. These problems are not peripheral bugs but intrinsic consequences of how analogy operates in data-driven generative pipelines.

The stakes are high for materials discovery. Over-reliance on unexamined analogy risks flooding experimental pipelines with plausible-but-invalid candidates, squandering synthesis and characterization resources on compounds that fail basic physical tests. It also fosters overconfidence in predictions that lie far outside training distributions and subtly distorts the scientific community's collective understanding of structure–property relationships [10-15]. By mapping the conceptual terrain and exposing these limits, this critique seeks to reorient the field toward generative frameworks that treat analogy as a powerful but strictly bounded tool rather than an implicit epistemology. The following sections first document the pervasiveness of analogy in current models, acknowledge its genuine strengths, and then develop each critique point in depth before turning to consequences and alternatives.

Analogy in Generative Materials Models

Analogy permeates generative materials modeling at every architectural level. The most direct manifestation appears in latent-space interpolation. In the seminal work of Gómez-Bombarelli and colleagues, molecules are encoded into a continuous latent representation where linear paths between known points correspond to chemically meaningful transitions; the model therefore treats the space between two known structures as an analogical continuum that can be sampled to propose new materials [7]. Graph neural networks extend this logic to crystalline systems by embedding local atomic environments such that similar coordination motifs map to nearby embeddings, enabling the network to reason analogically across crystal structures [6, 16-23]. Diffusion models similarly learn to denoise toward regions of high probability density by implicitly comparing noisy candidates to training examples through analogical similarity.

Transfer learning provides a second major conduit for analogical reasoning. Models pretrained on large organic molecular datasets are fine-tuned on inorganic or hybrid materials under the assumption that chemical knowledge transfers via structural or compositional analogies [4, 5, 18]. This transfer is not mere parameter initialization; it rests on the premise that the relational structure learned in one domain—bonding patterns, coordination preferences, electronic correlations—maps meaningfully onto another. Chemical analogy appears explicitly in substitutional design strategies. Davies and co-workers formalized materials discovery by chemical analogy, demonstrating how oxidation-state patterns and structural templates from known compounds can be reused to predict new compositions [8]. Ihalage and Hao [9] further operationalized analogical discovery by constructing material fingerprints that allow disordered perovskite oxides to be identified through hidden structural analogies.

Inverse design workflows embed analogy even more deeply. Given a target property profile, generative models propose structures by retrieving and modifying known motifs on the assumption that structural similarity implies property similarity [21, 24-28]. Element substitution routines treat the periodic table as an analogical map: replacing Pb with Sn in a perovskite lattice is performed because the two cations occupy analogous positions in terms of size, charge, and coordination chemistry [8]. Even unsupervised embedding approaches, such as those that learn word-like vectors for materials concepts from the literature, rely on distributional analogies: compounds that co-occur in similar contexts are presumed to share latent similarities that can be exploited for discovery [15].

Table 1 consolidates the manuscript's central theoretical contribution by distinguishing the specific analogy-driven operations used in generative materials modeling from the hidden assumptions, characteristic failure modes, and inferential safeguards each operation requires.

Table 1. Analytical taxonomy of analogy-based reasoning failures in generative materials models.

Analogy-driven	Hidden epistemic	Failure mode	Typical manifestation in	Why the error is conceptually	What would be needed to

operation in generative modeling	assumption		materials AI	serious	discipline the inference
Latent-space interpolation between known materials	Geometric proximity corresponds to chemically meaningful continuity	Analogy as a substitute for understanding	A generated structure appears compositionally or structurally plausible but violates stability, symmetry, or electronic constraints	The model reproduces pattern regularity without representing the causal basis of the pattern	Embed physical invariants, stability filters, and mechanism-aware validation during generation
Transfer learning from one material domain to another	Relational structure learned in one chemical family remains valid in another	False analogy	Knowledge transferred from organics to inorganics, oxides to chalcogenides, or Pb-based to Sn-based systems, despite different bonding and defect physics	Similarity at the representation level is mistaken for equivalence at the mechanism level	Domain-shift diagnostics, counterfactual testing, and chemically explicit transfer criteria
Extrapolation beyond the training manifold	Learned analogies remain valid as they are extended outward	Boundary blindness	Confident proposals in sparsely sampled or unseen compositional spaces with no signal that the analogy has broken	The model lacks a representation of the limits of its own inferential scope	Out-of-distribution detection, analogy-validity boundaries, and uncertainty calibrated to physical regime change
Interpreting latent dimensions as property directions	Statistical regularities reflect real causal axes in matter	Analogy reification	A latent direction is described as a true “bandgap axis,” “stability axis,” or “design rule”	Correlation is elevated into ontology, distorting scientific interpretation and theory-building	Causal identification, experimental contradiction testing, and strict separation of heuristic features from mechanistic claims
Structural or compositional substitution	Shared template positions imply shared functional behavior	Compound propagation of multiple analogy errors	Element swaps preserve formal descriptors but alter defect chemistry, kinetics, or emergent behavior	It normalizes pattern-preserving but physically invalid proposal generation	Multi-level screening linking substitution logic to thermodynamic, electronic, and kinetic checks

These practices are not occasional; they constitute the default operating mode of modern generative pipelines. Kailkhura *et al*. emphasize explainable machine learning that leverages analogical reasoning for accelerated discovery, while Merchant and colleagues scale deep learning by implicitly assuming that analogies learned at smaller scales generalize to larger, more complex material spaces [10, 27]. Buehler’s work on generative materials modeling draws explicit parallels to cognitive frameworks of analogy drawn from nature [21]. Even small-data regimes, where labeled examples are scarce, turn to analogical transfer as a primary mechanism for generalization [18].

The literature surveyed here—spanning at least twelve distinct contributions from the approved reference set—reveals a consistent pattern: generative models do not merely use similarity metrics; they operationalize analogy as the core inference engine [1, 2, 11, 14, 16, 20]. Yet the conceptual machinery that makes these analogies valid or invalid is rarely

foregrounded. The next section acknowledges the genuine epistemic power that such reasoning confers before turning to the limits that the field has largely left unexamined.

The Power of Analogy

Despite the conceptual risks detailed later, analogy remains an undeniably powerful engine of generalization in data-scarce scientific domains. Materials science has always operated under severe data limitations; the space of possible compounds vastly exceeds what can be synthesized and characterized. Analogy allows generative models to extrapolate creatively from sparse examples by identifying relational structures that hold across different chemical contexts [2, 16]. Gentner's structure-mapping framework, though developed in cognitive science, illuminates why this works: when deep relational systems align—such as coordination geometries or electronic band structures—analogy can license inferences that go beyond surface features [1].

In practice, this power manifests as the ability to generate novel candidates that respect broad chemical regularities. A model trained on a handful of stable perovskites can propose thousands of compositionally related variants because it has captured the analogical invariance of the ABX_3 template [8, 9]. Transfer learning similarly capitalizes on analogy to bootstrap performance in low-data regimes: knowledge of organic photovoltaic motifs transfers analogically to hybrid organic–inorganic systems, dramatically reducing the need for exhaustive labeled datasets [4, 18]. Historically, materials discovery itself has been propelled by analogical leaps—think of the progression from silicon to III–V semiconductors or from oxide perovskites to halide variants—demonstrating that analogy is not an AI-specific artifact but a fundamental cognitive and scientific strategy [5, 25].

Moreover, analogy embeds domain knowledge efficiently. By encoding chemical similarity through latent dimensions or graph embeddings, models implicitly incorporate centuries of accumulated heuristic knowledge about periodic trends, bonding preferences, and structural motifs [6, 15]. This embodiment of domain knowledge is precisely what allows generative pipelines to propose synthetically plausible rather than merely mathematically possible materials. Kittur and colleagues have shown at larger scales how analogical innovation, when scaled via crowds and AI, can produce breakthroughs precisely because it leverages relational mappings that humans intuitively recognize [16].

The power of analogy, therefore, lies in its capacity to compress relational knowledge, support creative generation, and enable generalization where brute-force enumeration is impossible [11, 20]. Yet power is not a warrant for unlimited application. The very mechanisms that make analogy efficient—its focus on relational abstraction rather than exhaustive causal modeling—also introduce the vulnerabilities explored in the following critique points. Acknowledging the strengths does not license ignoring the limits; instead, it sharpens the need to understand when and why analogy fails in generative materials contexts.

Analogy as Substitute for Understanding

A fundamental limitation emerges when analogy ceases to function as a heuristic aid and instead assumes the role of explanation. In generative materials models, particularly those based on latent-space interpolation or structural substitution, outputs often appear chemically plausible precisely because they preserve surface-level regularities. Yet this apparent coherence can mask a deeper absence of mechanistic grounding. Models trained on structured datasets, such as perovskites, may generate new compositions by smoothly varying lattice parameters or atomic configurations, producing candidates that align with learned patterns while simultaneously violating fundamental constraints such as the Goldschmidt tolerance factor or phonon stability criteria [3, 7]. Under these conditions, what is learned is not the causal structure of the system but a mapping between proximal representations within the training manifold.

The difficulty is compounded by the interpretive habits of researchers, who frequently evaluate generated structures through visual or compositional similarity to known compounds. This practice reinforces the impression that the model

has captured meaningful physical relationships, even when it has merely reproduced analogical continuity [4, 21]. Insights from cognitive science underscore the distinction: genuine analogical reasoning involves not only mapping relational structure but also evaluating its validity within the target domain. Holyoak and Thagard emphasize that successful analogy requires systematic alignment and constraint checking, whereas generative models perform only the initial mapping step, omitting the evaluative phase entirely [2]. The result is a form of analogical reasoning without verification, in which proximity in representation space substitutes for physical plausibility.

This substitution becomes particularly consequential in inverse design contexts, where optimization targets further constrain the generation process. When a desired bandgap is specified, the model retrieves and modifies known motifs through analogical extension rather than engaging with the underlying electronic structure. The resulting candidates may satisfy the target within the model's internal metric, yet remain inconsistent with fundamental principles of crystal symmetry or electronic behavior [8, 28]. In this setting, analogy effectively replaces quantum-mechanical reasoning, yielding outputs that are internally coherent but externally invalid.

Crucially, this limitation is not contingent on model size or data volume but reflects a structural feature of correlation-driven architectures. Even the most advanced generative systems, including diffusion and transformer-based models, operate by learning statistical regularities rather than causal mechanisms [10, 25]. Their strength lies in reproducing patterns of co-occurrence, not in distinguishing between superficial similarity and physically necessary relations. Without explicit mechanisms to ground analogical mappings in physical law, generative models risk producing artifacts that satisfy statistical criteria while lacking scientific validity. In this sense, analogy becomes epistemically hollow when it replaces, rather than supports, understanding.

False Analogies and Materials

The limitations of analogy become more pronounced when considering cases in which relational mappings appear valid at the level of structure or composition but fail to hold at the level of emergent properties. Generative models often assume that structural similarity implies functional similarity. Yet, this assumption breaks down in many materials contexts where subtle differences in electronic structure or bonding lead to divergent behavior [8, 9]. The substitution of Pb^{2+} with Sn^{2+} in perovskites exemplifies this issue: while the overall ABX_3 architecture is preserved, the resulting materials exhibit markedly different band structures, defect dynamics, and stability profiles. Models that interpolate along this axis propagate an analogy that is geometrically consistent but physically misleading, generating candidates whose predicted properties rest on invalid relational assumptions [7, 27].

This problem is amplified near chemical boundaries, where shifts in elemental composition introduce qualitatively new regimes of behavior. A model trained predominantly on oxides may treat sulfides or selenides as straightforward analogs, overlooking the increased polarizability and altered bonding characteristics that fundamentally change their properties [15, 25]. Embedding approaches such as those developed by Tshitoyan *et al*. capture historical patterns of association within the literature, but these patterns reflect prior human categorizations rather than intrinsic physical equivalence [15]. When generative models inherit such embeddings, they risk reproducing and amplifying conceptual biases embedded in the training corpus.

The challenge becomes even more pronounced in systems characterized by disorder and emergent complexity. High-entropy alloys provide a case in which analogical reasoning suggests that properties can be inferred through extrapolation from simpler compositions. Yet, the interplay of configurational entropy, lattice distortion, and competing phases produces behaviors that defy such simplification [18, 28]. Generative models that blend known structures to propose new compositions, therefore, generate candidates whose stability and performance cannot be reliably inferred from analogical similarity alone.

Underlying these failures is the absence of causal scaffolding within generative architectures. While models capture statistical associations between descriptors and outcomes, they lack the capacity to evaluate whether these associations

would persist under counterfactual perturbations [14, 20]. As a result, false analogies become self-reinforcing: generated candidates are filtered based on similarity to known examples, and the literature accumulates instances that appear to validate the mapping. At the same time, counterexamples remain unexplored [3]. This feedback loop operates largely below the level of explicit awareness, allowing erroneous analogies to propagate unchecked. What emerges is not an occasional anomaly but a systematic vulnerability, rooted in the reliance on surface-level relational mapping in domains governed by deep causal structure.

Analogy Boundary Blindness

A further limitation arises from the inability of generative models to detect the boundaries within which analogies remain valid. Once relational mappings are encoded in latent space, they are treated as continuously extensible, with no internal mechanism to signal when extrapolation has entered a regime where underlying assumptions no longer hold [3, 7, 21]. This absence of boundary awareness is not merely a technical gap but reflects a deeper conceptual limitation: analogy, when implemented without evaluative constraints, lacks an intrinsic notion of scope. As a result, models continue to generate candidates along learned dimensions even when physical or chemical plausibility has been violated.

Comparisons with human cognition highlight this deficiency. Theories of analogical reasoning emphasize that mapping must be accompanied by assessment of relational alignment and domain constraints. Gentner's structure-mapping framework and the work of Holyoak and Thagard both stress that successful analogy depends on evaluating whether transferred relations remain consistent within the target system [1, 2]. Generative models, by contrast, lack such evaluative layers. A system trained on cubic perovskites may extend analogical transformations into regimes where symmetry-breaking distortions or coordination constraints invalidate the original mapping, yet no internal signal marks this transition [8, 9, 27].

Empirical examples illustrate the consequences of this blindness. In high-entropy oxides, extrapolation based on analogical blending assumes linear scaling of configurational entropy, ignoring competing enthalpic effects that limit stability beyond certain compositional thresholds [18, 28, 29]. Similarly, in two-dimensional materials discovery, models may generate new structures by extending patterns learned from van der Waals systems into regimes where bonding transitions alter electronic and structural properties in fundamental ways [6, 15, 25]. In each case, the model continues to operate as though the analogy remains valid, producing candidates that appear coherent within the latent representation but fail under physical scrutiny.

The problem becomes particularly acute in inverse design, where target properties may lie outside the convex hull of the training data. Even in such cases, models proceed through analogical retrieval and modification, generating outputs that maintain surface-level similarity while occupying entirely different physical regimes [10, 20]. Confidence metrics derived from reconstruction error or discriminator scores may remain high, as they reflect consistency within the learned representation rather than fidelity to underlying physics. This disconnect creates an illusion of reliability precisely where uncertainty should be greatest.

Importantly, increasing the dataset size does not resolve this issue. While broader training distributions may extend the range of valid analogies, they do not introduce mechanisms for detecting their limits. Without explicit boundary-detection strategies—whether through physics-informed constraints, uncertainty quantification, or hybrid symbolic reasoning—generative models remain blind to the conditions under which their analogies fail [3, 20]. Addressing this limitation, therefore, requires a shift in architectural design, moving beyond unbounded analogical extrapolation toward systems capable of recognizing and respecting the limits of their own representations.

Analogy Reification

A final concern arises when analogical structures encoded in latent space are treated as if they correspond directly to real material relationships. This process, which may be described as analogy reification, involves the elevation of statistical correlations into ontological claims about how materials behave. Latent dimensions that correlate with properties such as bandgap or pore size are often interpreted as genuine axes of variation, inviting the assumption that traversing these dimensions corresponds to causal manipulation of material characteristics [3, 7, 28]. In reality, these axes reflect patterns present in the training data rather than fundamental physical laws.

This misinterpretation operates at multiple levels. In variational autoencoders, disentangled latent factors are frequently described as independent property controls, yet they encode mixtures of influences shaped by data distribution, experimental bias, and historical research focus [7, 21]. Interpolating along such directions may produce structures that align with statistical expectations while violating thermodynamic or kinetic constraints that were never captured by the model [8, 9]. The apparent interpretability of these dimensions thus conceals their contingent and constructed nature.

The risk extends to broader analogical mappings, particularly when generative frameworks draw inspiration from domains such as biology. Efforts to translate biological design principles into materials generation often rely on superficial similarities, such as hierarchical structuring, while neglecting differences in energy scales and environmental conditions that fundamentally alter system behavior [21]. Similarly, literature-based embeddings treat co-occurrence patterns as indicators of causal relationships, despite the fact that such patterns may reflect sociological factors rather than physical equivalence [15, 25].

Over time, these reified analogies can become entrenched within the literature. Once a latent direction is labeled as representing a specific property, subsequent studies may cite it as evidence for an underlying relationship, even in the absence of experimental validation [4, 5, 10]. In data-scarce regimes, this effect is particularly pronounced, as limited training examples amplify incidental correlations that are then interpreted as general principles [18]. What begins as a heuristic tool for generation thus evolves into a conceptual framework that shapes how materials are understood.

The persistence of this process reflects a structural feature of generative models, which present their internal representations in ways that invite ontological interpretation. Without mechanisms to distinguish between statistical regularities and causal relationships, users are encouraged to treat latent geometry as a direct reflection of material reality [2, 14, 20]. Preventing this slippage requires maintaining a clear epistemological boundary, in which analogies are recognized as provisional constructs that support exploration rather than as definitive accounts of how materials systems operate.

Consequences for Materials Discovery

The conceptual limits of analogy-based reasoning manifest not only as theoretical concerns but as concrete disruptions within materials discovery pipelines. When generative models operate without explicit recognition of these limits, the resulting outputs often appear chemically coherent while failing to satisfy fundamental physical constraints. Structures generated through latent interpolation or motif preservation may align with known coordination patterns or compositional trends, yet violate thermodynamic stability or electronic consistency upon closer inspection [7, 8, 27]. Because such candidates retain a surface-level plausibility, they pass initial screening and enter experimental workflows, where their deficiencies are revealed only after considerable effort has been expended.

This dynamic introduces a broader inefficiency in experimental practice. Repeated reliance on analogical extrapolation directs synthesis efforts toward compounds whose predicted properties rest on invalid relational mappings, leading to cycles of unsuccessful experimentation [9, 15, 25]. Each failed attempt represents not only the consumption of material and computational resources but also the displacement of alternative research directions that may have yielded more substantive insight. Over time, this pattern accumulates into a structural inefficiency that undermines the very promise of acceleration associated with generative methodologies.

A related consequence emerges at the level of model interpretation, where confidence estimates fail to reflect the true limits of applicability. Generative systems trained on bounded datasets often extend analogical reasoning into regions far removed from the training manifold without appropriately increasing uncertainty. Probability scores and predictive metrics remain artificially stable, creating an impression of reliability even in regimes where the underlying assumptions no longer hold [10, 20, 28]. This miscalibration fosters overconfidence, encouraging researchers to make strong claims regarding the feasibility or performance of proposed materials without sufficient grounding in physical validation.

The cumulative effect of these tendencies extends into the domain of scientific understanding itself. As analogical correlations become embedded within the literature, they risk being reinterpreted as genuine structure–property relationships, shaping how subsequent studies conceptualize material behavior [3-5]. Over time, this process generates a feedback loop in which models trained on historically contingent patterns reinforce and amplify those same patterns, gradually transforming heuristic associations into accepted knowledge. What begins as a practical shortcut thus evolves into a source of conceptual distortion, influencing both experimental priorities and theoretical development.

These dynamics indicate that unexamined reliance on analogy does not merely introduce isolated errors but systematically reshapes the epistemic landscape of materials AI. The resulting system, while capable of generating large volumes of plausible candidates, risks functioning as an amplifier of error rather than a driver of discovery, with consequences that are both scientific and economic in scope.

Alternative Approaches

Addressing these limitations requires a reconfiguration of how analogy is integrated into generative workflows, shifting from an implicit organizing principle to a bounded and verifiable component of the modeling process. One avenue involves embedding physical constraints directly within generative architectures, ensuring that analogical interpolation is restricted to regions consistent with known invariants. By incorporating conditions such as charge neutrality, tolerance-factor limits, or stability criteria into latent-space sampling, models can prevent the generation of candidates that violate fundamental principles while preserving the flexibility of analogical exploration [6, 8, 25].

A complementary strategy emphasizes the integration of rapid verification mechanisms that subject-generated candidates to independent physical evaluation before downstream use. Techniques such as density-functional theory or molecular dynamics simulations can act as filters, distinguishing between structures that are merely plausible within the model's representation and those that are physically realizable [4, 5, 18]. This coupling of generation and validation introduces an external check on analogical reasoning, aligning model outputs with established physical laws.

Beyond these modifications, there is growing interest in architectures that move beyond purely correlational representations by incorporating elements of causal reasoning. Hybrid models that learn explicit relationships between variables alongside latent embeddings offer the possibility of testing whether analogical mappings remain valid under counterfactual perturbations [14, 20]. Such approaches begin to address the core limitation of analogy by introducing mechanisms capable of evaluating, rather than simply extending, relational patterns.

Another refinement involves making the limits of analogy visible to the user through explicit uncertainty quantification. By associating each generated candidate with a measure of its distance from validated regions of the representation space, models can provide a calibrated indication of how far an analogy has been extended beyond known regimes [10, 21, 28]. This form of uncertainty does not eliminate error but renders it interpretable, allowing researchers to prioritize candidates based on both potential and risk.

Finally, hybrid frameworks that combine generative models with rule-based or physics-informed components offer a pathway toward integrating analogy with mechanistic insight. By embedding domain knowledge at multiple stages—before generation, during sampling, and after candidate formation—such systems ensure that analogical reasoning

remains constrained by established scientific principles [3, 9, 27]. In this configuration, analogy serves as a source of creative variation rather than as a substitute for understanding.

Taken together, these approaches do not seek to eliminate analogy but to discipline its use, preserving its capacity to explore complex design spaces while introducing safeguards that maintain alignment with physical reality. Through this reorientation, generative materials AI can retain its generative power without incurring the epistemic costs associated with unbounded analogical reasoning.

Table 2 translates the critique into a constructive research agenda by contrasting unbounded analogy-driven generation with the design principles of a hybrid framework capable of supporting more trustworthy materials discovery.

Table 2. From unbounded analogical generation to epistemically grounded discovery: a design framework for hybrid generative materials systems

Design dimension	Unbounded analogy-driven generator	Epistemically grounded hybrid generator	Practical implication for materials discovery	Evaluation question for authors, reviewers, and readers
Basis of inference	Similarity, proximity, and pattern continuation	Similarity is bounded by physics, causality, and verification	Reduces the chance that plausible-looking candidates are treated as valid hypotheses	Does the model distinguish resemblance from physically warranted transfer?
Treatment of novelty	Novelty is rewarded if it remains near the learned analogical structure	Novelty is accepted only if it survives constraint checks and validation gates	Encourages discovery without confusing novelty with scientific legitimacy	Is novelty filtered by known invariants before being reported?
Handling of domain boundaries	Boundaries are implicit or ignored	Boundaries are explicitly modeled, estimated, and communicated	Prevents silent extrapolation into chemically or physically alien regimes	Does the system estimate where its analogy ceases to be trustworthy?
Interpretation of latent structure	Latent directions are often treated as meaningful design axes	Latent directions are treated as provisional heuristics unless causally validated	Limits ontological overclaiming and protects theory formation	Are latent features described as heuristic correlations or as demonstrated mechanisms?
Uncertainty reporting	Confidence is often internal to the model and poorly calibrated	Uncertainty includes analogy validity, out-of-distribution risk, and physics screening status	Makes downstream synthesis decisions more rational and less wasteful	Does uncertainty increase when the proposal moves away from validated regions?
Role of verification	Often post hoc and optional	Built-in verification loop using fast physics, simulation, or symbolic checks	Converts generation into a disciplined hypothesis pipeline rather than a raw idea generator	What verification step is mandatory before a candidate is presented as promising?
Scientific contribution	May accelerate candidate production	Aims to accelerate trustworthy scientific understanding	Shifts the field from pattern replication to	Does the architecture improve explanation, or only output volume?

			mechanism-aware discovery	
Ideal research posture	Analogy as default epistemology	Analogy as a bounded heuristic within a broader evidentiary framework	Reorients the field toward robust, defensible generative science	Is analogy explicitly framed as a tool with limits rather than as proof?

Conclusion

This study has demonstrated that analogy-based reasoning, while undeniably powerful, imposes fundamental conceptual limits on generative materials models that the field has largely left unexamined. From its role as a substitute for understanding, through the propagation of false analogies and boundary blindness, to the reification of statistical correlations into ontological claims, analogy introduces systematic vulnerabilities that cannot be overcome by scaling alone. The consequences—plausible but invalid candidates, wasted synthesis effort, overconfident predictions, and scientific misunderstanding—threaten to undermine the very promise of accelerated discovery. By proposing physics-constrained analogy, verification loops, causal modeling, uncertainty-aware generation, and hybrid frameworks, this paper charts a path toward epistemologically robust generative systems that treat analogy as a carefully bounded tool rather than an unacknowledged foundation. The future of artificial intelligence for materials science depends on explicit acknowledgment of these limits and the architectural innovations required to respect them. Only then can generative models fulfill their potential as genuine partners in scientific discovery rather than sophisticated pattern replicators.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 10 Dec 2022 Revised: 31 Jan 2023 Accepted: 25 Feb 2023
Published online: 18 July 2023

References

Puppin M, Nicholson CW, Monney C, Deng Y, Xian RP, Feldl J, et al. Excited-state band structure mapping. Phys Rev B. 2022;105(7):075417.

Weber M, Szigetvári Á, Halmai M, Keglevich P, Szántay C Jr. On the role of mental leaps in small-molecule structure elucidation by NMR spectroscopy. ARKIVOC. 2022;2022.

Kaushik A, Kaur P, Choudhary N, Priyanka. Stacking regularization in analogy-based software effort estimation. Soft Comput. 2022;26(3):1197-216.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.

Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):83.

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.

Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-76.

Davies DW, Butler KT, Isayev O, Walsh A. Materials discovery by chemical analogy: Role of oxidation states in structure prediction. Faraday Discuss. 2018;211:553-68.

Ihalage A, Hao Y. Analogical discovery of disordered perovskite oxides by crystal structure information hidden in unsupervised material fingerprints. npj Comput Mater. 2021;7(1):75.

Kailkhura B, Gallagher B, Kim S, Hiszpanski A, Han TY. Reliable and explainable machine-learning methods for accelerated material discovery. npj Comput Mater. 2019;5(1):108.

Han J, Shi F, Chen L, Childs PR. A computational tool for creative idea generation based on analogical reasoning and ontology. AI EDAM. 2018;32(4):462-77.

Thibaut JP, Glady Y, French RM. Understanding the what and when of analogical reasoning across analogy formats: An eye-tracking and machine learning approach. Cogn Sci. 2022;46(11):e13208.

Leonard K, Sepehri P, Cheri B, Kelly DM. Relational complexity influences analogical reasoning ability. iScience. 2023;26(4).

Combs K, Lu H, Bihl TJ. Transfer learning and analogical inference: A critical comparison of algorithms, methods, and applications. Algorithms. 2023;16(3):146.

Tshitoyan V, Dagdelen J, Weston L, Dunn A, Rong Z, Kononova O, et al. Unsupervised word embeddings capture latent knowledge from materials science literature. Nature. 2019;571(7763):95-8.

Kittur A, Yu L, Hope T, Chan J, Lifshitz-Assaf H, Gilon K, et al. Scaling up analogical innovation with crowds and AI. Proc Natl Acad Sci. 2019;116(6):1870-7.

Ichien N, Lu H, Holyoak KJ. Verbal analogy problem sets: An inventory of testing materials. Behav Res Methods. 2020;52(5):1803-16.

Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. npj Comput Mater. 2023;9(1):42.

Badra F, Sedki K, Ugon A. On the role of similarity in analogical transfer. In: International Conference on Case-Based Reasoning. Cham: Springer; 2018. p. 499-514.

Hüllermeier E. Towards analogy-based explanations in machine learning. In: International Conference on Modeling Decisions for Artificial Intelligence. Cham: Springer; 2020. p. 205-1.

Hsu YC, Yang Z, Buehler MJ. Generative design, manufacturing, and molecular modeling of 3D architected materials based on natural language input. APL Mater. 2022;10(4).

Fuhr AS, Sumpter BG. Deep generative models for materials discovery and machine learning-accelerated innovation. Front Mater. 2022;9:865270.

Fung V, Zhang J, Juarez E, Sumpter BG. Benchmarking graph neural networks for materials chemistry. npj Comput Mater. 2021;7(1):84.

Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol. 2019;188(12):2222-39.

Oliveira ON Jr, Oliveira MC. Materials discovery with machine learning and knowledge discovery. Front Chem. 2022;10:930369.

Han N, Shen Z, Zhao X, Chen R, Thakur VK. Perovskite oxides for oxygen transport: Chemistry and material horizons. Sci Total Environ. 2022;806:151213.

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

Menon D, Ranganathan R. A generative approach to materials discovery, design, and optimization. ACS Omega. 2022;7(30):25958-73.

Nakazato K. Ecological analogy for generative adversarial networks and diversity control. J Phys Complex. 2023;4(1):01LT01.