ORIGINAL RESEARCH **Open access**

Model Entropy and Scientific Information Loss in Compressed Representations of Materials

Sanjay Kulkarni[1], Meenal Joshi[1*], Rohan Patil[2], Aniket Deshmukh[2]

Abstract

Compressed representations—such as handcrafted descriptors, autoencoder embeddings, and graph-neural-network latent spaces—have become indispensable in artificial-intelligence-driven materials science because they enable scalable property prediction from high-dimensional atomic configurations. Yet the very act of compression, while optimizing statistical correlation with target properties, systematically discards information whose scientific value lies outside mere predictive utility. This theoretical analysis applies information-theoretic principles from Shannon and Cover and Thomas to examine how dimensionality reduction in materials representations affects the retention of scientifically relevant content. Drawing on the concept of model entropy introduced by S. S., the paper introduces "model entropy" as a quantitative lens for assessing the information content preserved in any compressed materials representation. It articulates a core theoretical claim: compression optimized for predictive accuracy maximizes statistical information but can erode scientific information—mechanistic, causal, and counterfactual structures essential for understanding, explanation, and extrapolation. A typology of five distinct information-loss mechanisms is developed, each illustrated with representative materials-science scenarios. The analysis culminates in concrete implications for representation design and scientific inference, arguing that future materials AI must move beyond accuracy-centric evaluation toward explicit auditing and preservation of scientific information. By distinguishing statistical signal from epistemic content, this work offers a conceptual framework for building representations that serve both prediction and discovery without hidden epistemic costs.

Keywords Materials informatics, Representation learning, Model entropy, Scientific information loss, Compressed representations, Information theory

*Correspondence:
Meenal Joshi
meenal.joshi@gmail.com
[1] Department of Intelligent Materials Systems, Savitribai Phule Pune University, Pune, India
[2] Department of Computational Materials Analytics, IIT Bombay, Mumbai, India

Introduction

Materials artificial intelligence routinely converts complex atomic structures into compact vector or latent-space representations for tasks such as property prediction, phase classification, and inverse design. This compression is necessary for tractability and is usually treated as harmless so long as predictive accuracy remains high. This paper challenges that assumption by asking a deeper question: what information is irretrievably lost when materials are represented as low-dimensional vectors, and do those losses matter for scientific understanding rather than only prediction?

Standard information theory, developed by Shannon and systematized by Cover and Thomas, offers tools for measuring uncertainty and shared information [1-3]. In materials AI, however, these tools have largely been used to evaluate statistical performance rather than the epistemic consequences of retained versus discarded content. Surveys by Butler *et al.* [4] and Schmidt *et al.* [5] show how compressed representations have accelerated

discovery across solid-state materials, molecular systems, and high-entropy alloys. Yet, they pay little attention to whether compression itself obstructs mechanistic insight. This paper, therefore, distinguishes between two forms of information. Statistical information is the variance or correlation a representation preserves with respect to a target variable; it improves cross-validation scores. Scientific information, by contrast, consists of mechanistic, causal, geometric, and counterfactual structure that enables explanation, extrapolation, and hypothesis generation.

Table 1 formalizes the conceptual distinction between statistical and scientific information, highlighting their divergent behavior under compression and their unequal visibility to standard evaluation metrics."

Table 1. Structural comparison between statistical and scientific information in compressed materials representations

Dimension	Statistical information	Scientific information	Behavio under compress
Definition	Correlation with target variable (Y)	Mechanistic, causal, geometric, and counterfactual structure	Selective preserve
Measurement	Mutual information I(X;Y) [2]	Model entropy [3] (residual epistemic content)	Not explic optimize
Role in models	Optimized during training	Emergent, not enforced	Frequent discarde
Sensitivity to compression	Robust to dimensionality reduction	Fragile to dimensional bottlenecks	Rapid degradati below threshol
Relation to causality	May include spurious correlations	Encodes causal pathways	Not distinguish by loss function
Generalisation capacity	Strong in-distribution	Enables extrapolation	Degrade under ov compress
Dependence on architecture	Captured by any predictive model	Requires preservation of structure and relations	Architectu depende
Failure mode	Overfitting or noise capture	Mechanistic blindness	Hidden fr the accura metrics

Handcrafted descriptors compress crystal structures into symmetry-invariant fingerprints, variational autoencoders learn manifolds that discard atomic detail, and graph networks pool local environments into embeddings that can suppress long-range periodicity. In each case, predictive value is emphasized, while the scientific consequences of compression remain underexplored. This is not merely a technical issue but an epistemological one. If a representation optimized for one task discards information essential for another, then claims of "universal" or "general-purpose" embeddings may overstate their scientific scope.

Information Theoretic Foundations

Information theory provides a language for describing what is preserved and what is lost when an object is mapped into a lower-dimensional space. Its starting point is entropy, introduced by Shannon as a measure of uncertainty or average surprise in a random variable [1]. In materials science, entropy can be understood as the unpredictability of the full atomic configuration space before compression: high entropy reflects a wide range of chemically distinct arrangements, whereas low entropy reflects a collapse of many distinct configurations into similar vectors. Model entropy, as introduced by Liu *et al.* [3] and extended here, refers to the residual uncertainty or information richness retained in the compressed descriptor, embedding, or latent vector after dimensionality reduction. Unlike standard Shannon entropy applied to raw data, model entropy is sensitive to the interpretability of the retained dimensions, not just their statistical variability.

Mutual information, formalized by Cover and Thomas, measures how much knowing one variable reduces uncertainty about another [2]. Applied to a materials representation X and a target property Y, it indicates how

much predictive signal they share. Yet it does not reveal whether that signal reflects mechanistic causation or mere correlation. A representation may have high mutual information with formation energy while still discarding the geometric motifs that explain why that energy is low. This limitation is central to evaluating whether compression preserves scientifically useful content.

Rate-distortion theory, also from Cover and Thomas, formalizes the trade-off between compression and fidelity [2]. For any allowable level of distortion, there is a minimum rate, or number of bits and dimensions, needed to represent the source. In materials AI, rate corresponds to descriptor or latent-space dimensionality, while distortion is usually measured by predictive error. Lower rates, therefore, tend to increase distortion, but the framework does not specify what kind of information is being distorted. This paper argues that existing uses of rate-distortion logic in materials informatics focus almost entirely on statistical distortion while neglecting scientific distortion.

Together, entropy, mutual information, and rate-distortion theory provide the conceptual basis for this paper. Shannon's insight [1] that information can be quantified independently of meaning is powerful, but in materials science, it can also encourage the mistaken view that any reduction in entropy is harmless. Cover and Thomas organized these ideas into a framework that this study adapts to an epistemic setting [2]. If a materials representation is treated as a communication channel from full atomic structure to scientific interpreter, the key question is not only how many bits survive, but whether they still support mechanistic decoding. Model entropy is therefore the central diagnostic. It measures the information content of the representation itself, independent of downstream prediction. When it falls below the level needed for mechanistic reasoning, a representation may remain statistically accurate yet become scientifically impoverished. This reframing prepares the ground for analyzing current compression practices in materials AI and the information they discard.

Representation Compression in Materials AI

Three dominant families of compression underpin materials representation learning, each embodying a distinct strategy for reducing the complexity of atomic-scale information while preserving predictive utility. One widely adopted approach relies on handcrafted descriptors that transform atomic coordinates, lattice vectors, and chemical identities into fixed-length representations invariant to translation, rotation, and permutation. This invariance is achieved by abstracting away explicit positional detail and retaining aggregated statistical summaries, such as radial distribution functions or symmetry-based features. While this transformation enables computational tractability and facilitates model training, it does so by collapsing a multitude of structurally distinct configurations into indistinguishable representations, thereby eroding the specificity required to differentiate mechanistically meaningful variations [2-6].

A different mode of compression emerges in learned nonlinear embeddings, particularly those derived from autoencoders and variational autoencoders, as demonstrated by Gómez-Bombarelli *et al.* [7]. Here, high-dimensional inputs—including crystal graphs, electron densities, or molecular string encodings—are projected into a constrained latent space through an encoder network. The presence of a bottleneck enforces dimensional reduction, ensuring that only features deemed statistically salient for reconstruction are retained. Although the decoder can approximate the original input, the compression process selectively filters out information that does not contribute strongly to reconstruction fidelity. As a result, features that may be critical for mechanistic interpretation—such as defect topologies or subtle long-range ordering—are frequently attenuated or lost altogether, even when they play a decisive role in governing material behavior.

A third family of approaches is instantiated in graph neural networks and transformer-based architectures, exemplified by Chen *et al.* [6], where compression is distributed across iterative aggregation processes. In these models, node-level representations are constructed through message passing that integrates local atomic environments, followed by pooling operations that condense the entire structure into a fixed-size embedding. At each stage, information that cannot be efficiently summarized is discarded, and attention mechanisms further modulate this process by amplifying certain interactions while diminishing others. The resulting representation reflects a hierarchy of selective emphasis, in which only a subset of relational features is preserved in a form accessible to downstream prediction tasks.

Despite their methodological differences, these paradigms converge on a shared justification rooted in efficiency. Dimensionality reduction enhances training speed, improves generalization, and enables integration across diverse modeling pipelines. Yet this convergence also reveals a deeper pattern in how information is treated. Whether through averaging in descriptor-based methods, bottleneck constraints in latent-variable models, or iterative aggregation and selection in graph and transformer architectures, each approach systematically filters the input space according to criteria of statistical relevance. What is retained is that which maximizes mutual information with the training objective. At the same time, features that are weakly correlated with immediate predictive targets—despite their potential scientific significance—are progressively excluded. In this way, compression operates not merely as a technical necessity but as an implicit epistemic filter, privileging statistical fidelity over mechanistic completeness. The consequence is a consistent asymmetry: materials AI systems become increasingly proficient at prediction while leaving the question of scientific fidelity largely unexamined, thereby motivating the theoretical claims that follow.

Figure 1 illustrates how compression transforms full atomic configurations into low-dimensional representations, systematically preserving statistical information while discarding distinct categories of scientific information.

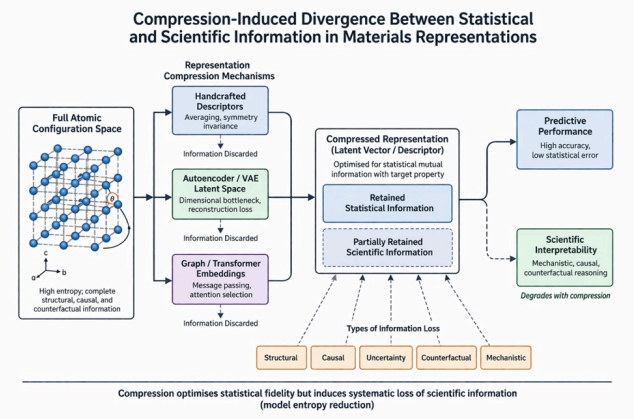

Figure 1. How compression transforms full atomic configurations into low-dimensional representations, systematically preserving statistical information while discarding distinct categories of scientific information.

Theoretical Claim: Compression Losses

Scientific Information

Compression optimized for predictive accuracy preserves statistical information—correlations with a target variable—but systematically discards scientific information, defined as mechanistic, causal, and counterfactual structure required for explanation and extrapolation.

Statistical and scientific information are not perfectly aligned; high mutual information with a target does not ensure preservation of scientifically relevant structure.

These claims follow from three arguments. First, any finite-dimensional representation projects a high-dimensional configuration space onto a lower-dimensional subspace. Rate-distortion theory [2] guarantees information loss, but optimization preserves only target-relevant variance. Dimensions orthogonal to prediction—yet essential for mechanism—are not retained. Second, materials exhibit emergent phenomena (e.g., defects, phase transitions, magnetic ordering) that depend on global or higher-order correlations often missed by compressed summaries. A model may predict accurately while erasing the structure that explains the phenomenon. Third, scientific reasoning relies on counterfactuals—how properties change under perturbations such as strain or substitution. Representations optimized for in-distribution prediction need not preserve the geometry required for reliable extrapolation.

Conceptually, statistical and scientific information can be viewed as overlapping but distinct sets [8-14]. Their intersection supports both prediction and explanation, but each also contains unique content: statistical-only information includes spurious correlations, while scientific-only information includes mechanistic detail and counterfactual structure. Model entropy [3] measures the total retained information, not just the predictive subset. When compression reduces representations to the intersection alone, predictive accuracy may remain high while scientific interpretability collapses.

These claims show that information loss under compression is not incidental but inherent to current optimization objectives, motivating the typology developed next.

Types of Information Loss

Table 2 consolidates the five compression-induced information-loss mechanisms into a unified analytical framework, linking each loss type to its underlying mechanism and its specific epistemic consequence in materials science.

Table 2. Typology of compression-induced scientific information loss and its epistemic consequences

Loss type	Compression mechanism	What is lost	Material exampl
Structural	Averaging and invariance constraints	Bond angles, topology, and spatial relations	Radial descripto removin lattice distortion high-entro alloys
Causal	Correlation-based optimization	Directional cause–and–effect chains	Coordinati number retained orbital hybridizati lost
Uncertainty	Variance suppression, latent regularisation	Epistemic and aleatoric uncertainty	DFT nois collapsed latent embeddin
Counterfactual	Dimensionality reduction	Perturbation directions (strain and doping)	Latent space lac strain-respons axes
Mechanistic	Bottlenecks and local aggregation	Higher-order/global physical mechanisms	Loss of long-rang charge ordering solids
Cross-Type Interaction	Combined compression effects	Coupled dependencies across scales	Perovski tilting + band structure collapse

Type 1: Structural information loss

This occurs when compression erases geometric or topological relationships between atoms that are not statistically salient for the chosen target. Handcrafted descriptors that average over radial shells, for example, discard precise bond-angle distributions even though those angles may govern phonon spectra or defect migration barriers. In high-entropy alloys, the local lattice distortion patterns that stabilize solid-solution phases can vanish once the representation collapses to composition-only vectors. Scientifically, this loss matters because structure–property relationships lie at the heart of materials design; without them, inverse design becomes blind to the atomic motifs that actually enable targeted functionality [15-17].

Type 2: Causal information loss

Compression discards directional cause–and–effect linkages when it retains only correlative statistics. Graph-neural-network embeddings that pool messages symmetrically may preserve that a certain coordination number correlates with hardness, yet lose the information that the coordination arises from specific orbital hybridization. In perovskite oxides, the causal chain from A-site cation size to octahedral tilting to band-gap opening can be severed if the representation collapses tilting angles into a single scalar. The scientific consequence is that downstream interpretability methods applied to such representations risk attributing causality to variables that are merely downstream effects, undermining hypothesis generation.

Type 3: Uncertainty information loss

When latent spaces are regularised toward Gaussian priors or when descriptors discard variance estimates, the representation loses information about epistemic or aleatoric uncertainty. Autoencoders trained on noisy density-functional-theory data may encode only mean properties while erasing the spread that signals metastability or synthesis difficulty. In the context of high-entropy ceramics, the representation may predict an average phase stability accurately yet hide the configurational entropy fluctuations that determine whether the phase can actually be quenched. Scientifically, this loss impairs risk-aware decision-making and prevents researchers from distinguishing robust trends from data artifacts.

Type 4: Counterfactual information loss

Aggressive dimensionality reduction removes the geometric degrees of freedom needed to simulate perturbations. A low-dimensional embedding optimized for equilibrium properties may lack the latent directions corresponding to applied strain or chemical substitution, rendering counterfactual queries ("what if we dope with 5 % Zr?") unreliable. In molecular materials design, variational autoencoders that collapse conformational flexibility into a handful of latent variables lose the ability to reason about how a small torsional change would alter reactivity. The scientific cost is diminished capacity for virtual screening and hypothesis testing outside the observed distribution.

Type 5: Mechanistic information loss

This is the most fundamental form: compression discards the underlying physical or chemical mechanisms when those mechanisms are encoded in higher-order or long-range correlations that the chosen architecture cannot preserve. Transformer attention that focuses exclusively on short-range interactions may retain compositional statistics while erasing the long-range charge ordering that explains colossal magnetoresistance. In solid-state systems, the link between electronic band structure and atomic topology can vanish once the representation is forced into a bottleneck that prioritizes local energetics. Without mechanistic information, materials AI delivers accurate predictions whose explanations remain opaque, stalling the feedback loop between computation and theory that has historically driven the field forward.

Each type arises directly from the rate-distortion optimization [2] that favors statistical fidelity over epistemic completeness. Collectively, they demonstrate that information loss in compressed materials representations is neither uniform nor benign; it is patterned, diagnosable, and consequential for the scientific enterprise.

Derived Properties and Corollaries

The theoretical claims from section 4, when considered alongside the five-part typology introduced in section 5, give rise to a set of corollaries that fundamentally reorient how compressed materials representations should be evaluated. Drawing on concepts from Shannon entropy, rate-distortion theory, and model-entropy diagnostics [1–3], these corollaries expose a structural limitation in prevailing approaches: representations optimized for statistical fidelity can achieve high predictive performance while remaining epistemically incomplete.

At the core of this reframing lies the recognition that predictive accuracy, taken in isolation, provides an insufficient criterion for judging representational quality. Compression procedures are designed to minimize statistical loss relative to a target, not to preserve the full spectrum of scientifically meaningful information. As a result, features that are weakly correlated with the prediction objective—yet central to mechanistic or causal understanding—may be systematically eliminated. Under such conditions, claims regarding the adequacy of a representation cannot be grounded solely in measures of fit; they require an explicit examination of the scientific content retained after compression. Absent such scrutiny, materials AI risks converging on representations that are numerically precise yet conceptually impoverished [18-21].

This limitation becomes more pronounced when considering the transferability of representations across tasks. Alignment with a specific prediction objective induces a selective filtering of information, privileging features that contribute to performance within that narrowly defined context. While this specialization enhances accuracy for the original task, it simultaneously erodes the capacity of the representation to support alternative scientific inquiries. What appears as robustness within one evaluative frame may therefore conceal fragility in another, explaining the persistent difficulty of constructing genuinely universal representations. The notion of universality remains elusive not because of insufficient model complexity, but because task-specific compression irreversibly discards information required for broader applicability.

A further implication emerges from the relationship between compression and the intrinsic dimensionality of the underlying scientific manifold. When dimensionality reduction is pushed beyond this intrinsic threshold, rate-distortion theory guarantees that information deemed statistically irrelevant to the target will be eliminated. Yet it is precisely within these low-signal dimensions that mechanistically significant structure often resides. The consequence is a form of representational brittleness, in which models perform reliably within the confines of their training distribution but fail under extrapolation, where the discarded dimensions become decisive. Importantly, this failure mode remains largely invisible to conventional predictive metrics, which do not capture the loss of latent scientific structure.

Taken together, these corollaries necessitate a shift from single-objective evaluation toward a more comprehensive framework of information accounting. Rather than optimizing exclusively for predictive accuracy, each stage of compression should be calibrated against thresholds that preserve scientifically relevant information. Model-entropy diagnostics provide one such mechanism, offering a way to quantify the extent to which representational capacity has been reduced relative to the complexity of the underlying system [3]. In practice, this entails the parallel use of two complementary metrics: one capturing statistical mutual information with the prediction target, and another assessing residual model entropy in relation to broader mechanistic probes. Through this dual perspective, evaluation moves beyond surface-level performance toward a deeper assessment of whether compressed representations retain the informational structure necessary for genuine scientific understanding [22-25].

Relation to Existing Concepts

The framework of model entropy and scientific information loss connects to—and significantly extends—four established concepts in information theory and machine learning for materials science. Rate-distortion theory, first articulated by Shannon and fully systematized by Cover and Thomas, already quantifies the inevitable trade-off between compression rate and allowable distortion [1, 2]. The present analysis refines this classical trade-off by partitioning distortion itself into two qualitatively different components: statistical distortion (measured by predictive error) and scientific distortion (measured by erosion of mechanistic, causal, and counterfactual content). Conventional applications within materials informatics have focused almost exclusively on the former; here, the latter is shown to occupy an equally critical region of the rate-distortion surface. A representation may therefore occupy an attractive operating point for statistical fidelity while simultaneously residing in an unacceptable regime for scientific fidelity.

The notion of sufficient statistics, again drawn from Cover and Thomas, provides a second natural anchor point [2]. A statistic is sufficient if it captures all information relevant to a given inference task. In materials AI, sufficiency has almost always been assessed solely with respect to narrow predictive objectives. The typology developed in Section 5 demonstrates that many widely used compressed representations are statistically sufficient yet scientifically insufficient: they preserve everything required for accurate regression while discarding the causal chains and counterfactual gradients required for broader scientific inference. Model entropy [3], therefore, functions as a practical diagnostic that reveals whether a representation has crossed the threshold from merely sufficient to genuinely informative for scientific purposes [24-29].

Disentanglement, a concept popularised in generative modeling and explicitly applied within autoencoder frameworks for molecules [7], seeks to isolate independent factors of variation into distinct latent dimensions. While disentanglement undeniably improves statistical interpretability inside the retained manifold, it does not automatically guarantee preservation of scientific information that lies outside those factors. A perfectly disentangled latent space may still collapse mechanistic couplings—such as the interplay between local coordination geometry and global band-structure topology in perovskites—that are statistically redundant yet causally essential. The present framework, therefore, positions disentanglement as a necessary but far from sufficient condition for scientific information retention.

Finally, the growing body of work on interpretability in materials AI, surveyed in detail by Butler *et al.* [4] and Schmidt *et al.* [5], typically applies post-hoc explanation techniques (saliency maps, attention weights, feature-importance rankings) to already-compressed representations. The analysis here exposes a fundamental upstream limitation. Once compression has removed mechanistic information through any of the five loss types, downstream interpretability methods can only illuminate the surviving statistical signal. Attention maps or saliency scores thereby risk becoming misleading precisely when model entropy [3] has been driven below the threshold required for mechanistic decoding. The framework thus elevates scientific information preservation from a desirable downstream refinement to a non-negotiable prerequisite for any credible interpretability pipeline.

By explicitly linking model entropy to these four pillars—rate-distortion theory, sufficient statistics, disentanglement, and interpretability—the present work supplies a unified conceptual vocabulary that bridges classical information theory with the distinctive epistemic demands of materials science. It does not supplant existing tools but adds the missing epistemic layer that determines whether those tools ultimately advance discovery or merely statistical approximation.

Implications for Representation Design

The theoretical claims, corollaries, and typology together translate into four concrete principles that must guide the future design of materials representations if scientific information is to be preserved alongside statistical fidelity.

Task-Specific information requirements

Before any compression, the designer must explicitly enumerate the scientifically necessary information required by each intended downstream task. For defect-migration studies, this includes counterfactual strain gradients and energy-barrier landscapes; for phase-stability prediction, it includes configurational-entropy fluctuations and long-range ordering motifs. Only after these requirements have been cataloged can the latent dimensionality, architecture, and regularisation strategy be chosen so that the scientific crescent is actively protected rather than inadvertently discarded.

Information auditing

Every published representation must be accompanied by an explicit audit that reports both its statistical mutual information with the target property and its residual model entropy relative to a suite of mechanistic probes. Estimators grounded in the foundational literature [1, 2] can be repurposed for this purpose, allowing practitioners to quantify the size of the scientific-only crescent that survives compression. The graph-network framework of Chen *et al.* [6] could, for example, be extended with such an audit layer instead of being evaluated on predictive accuracy alone.

Multi-objective compression

Optimization objectives should be expanded beyond predictive loss to include explicit regularisation terms that penalize erosion of model entropy [3]. This may take the form of auxiliary reconstruction losses that encourage retention of higher-order correlations or mutual-information maximization terms applied to a set of mechanistic descriptors. The outcome is a Pareto front on which designers can deliberately select operating points that balance statistical and scientific fidelity rather than optimizing one at the expense of the other.

Conservative compression

Dimensionality reduction should proceed only as far as required to satisfy immediate computational constraints—never further. Over-compression, defined as pushing latent dimensionality below the intrinsic scientific manifold, should be flagged as epistemically risky. In practice, this means establishing minimum model-entropy thresholds calibrated against representative materials families and rejecting candidate architectures that violate those thresholds even when they yield marginally superior accuracy.

Adoption of these four principles does not eliminate compression; it renders compression epistemically transparent, measurable, and therefore controllable. Representation design thereby evolves from an ad-hoc engineering practice into a principled information-accounting discipline that ensures every vector or latent space deployed in materials AI serves the dual imperatives of prediction and deepened mechanistic understanding.

Implications for Scientific Inference

When materials scientists draw inferences from compressed representations, the information losses cataloged in this analysis impose three distinct epistemic hazards that must be acknowledged explicitly. First, causal claims become vulnerable to type-2 (causal information loss): a graph-neural-network embedding may correctly predict that higher coordination number correlates with increased hardness, yet the representation no longer encodes the orbital-hybridization pathway that actually causes the correlation. Any subsequent mechanistic interpretation, therefore, risks reversing cause and effect. Second, generalization claims are undermined by type-4 (counterfactual information loss) and type-5 (mechanistic information loss): the geometric degrees of freedom and higher-order correlations required for reliable extrapolation outside the training distribution may have been discarded, rendering confident assertions about unseen chemical spaces statistically supported but scientifically ungrounded. Third, downstream interpretability techniques inherit precisely the omissions they attempt to illuminate; saliency maps or attention weights can only highlight surviving statistical dimensions, not the erased scientific ones.

These hazards do not invalidate the use of compressed representations but require that every scientific inference drawn from them carry an explicit qualification. A formation-energy prediction, however accurate, must now be

accompanied by a caveat that the supporting representation may lack the geometric or causal structure needed to explain the prediction. Likewise, any claim of "representation learning" must be tempered by the recognition that what has been learned is often only the statistical intersection rather than the full scientific union. The framework, therefore, calls for a new community reporting standard; every materials-AI study should include a concise model-entropy statement [3] that quantifies the scientific information retained, thereby allowing readers to calibrate the epistemic weight of the reported conclusions.

In the broader epistemology of computational materials science, this shift moves the field away from an implicit positivism—where predictive success is treated as self-certifying—toward a more reflexive stance that acknowledges representational choices themselves as carriers of epistemic cost. Scientific inference thereby becomes more robust precisely because its limitations are made visible and accountable rather than hidden inside the compression step.

Conclusion

This theoretical analysis has articulated a core claim: compression optimized for predictive accuracy in materials artificial intelligence maximizes statistical information while systematically risking the loss of scientific information—mechanistic, causal, counterfactual, and structural content essential for understanding, explanation, and extrapolation. By introducing model entropy as a diagnostic of residual information content within any descriptor, embedding, or latent space, the paper supplies a conceptual yardstick that distinguishes representations that merely predict from those that also illuminate. The typology of five information-loss types, the three derived corollaries, and the four design principles demonstrate that these losses are patterned, diagnosable, and consequential rather than incidental.

The distinction between statistical and scientific information, therefore, emerges as a foundational epistemic consideration for the entire field. Future representation learning must move beyond accuracy-centric evaluation toward explicit auditing and multi-objective optimization that safeguards the scientific crescent. Only by treating compression as an information-accounting act—rather than a neutral engineering convenience—can artificial intelligence for materials science fulfill its dual promise of rapid discovery and deepened mechanistic insight. The framework advanced here offers the theoretical scaffolding required to realize that promise without hidden epistemic costs.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 15 Sep 2022 Revised: 15 Oct 2022 Accepted: 26 Dec 2022
Published online: 18 January 2023

References

Floyd K, Schrodt P, Erbert LA, Scharp KM. Exploring communication theory: Making sense of us. London: Routledge; 2022.

Lapidoth A, Narayan P. New mathematical techniques in information theory. Oberwolfach Rep. 2023;19(1):683-707.

Liu Q, Tan Z, Chen D, Chu Q, Dai X, Chen Y, et al. Reduce information loss in transformers for pluralistic image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE; 2022. p. 11347-57.

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.

Trinh C, Tbatou Y, Lasala S, Herbinet O, Meimaroglou D. On the development of descriptor-based machine learning models for thermodynamic properties: Part 1—From data collection to model construction: Understanding of the methods and their effects. Processes. 2023;11(12):3325.

Davariashtiyani A, Kadkhodaie Z, Kadkhodaei S. Predicting synthesizability of crystalline materials via deep learning. Commun Mater. 2021;2(1):115.

Langer MF, Goeßmann A, Rupp M. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. npj Comput Mater. 2022;8(1):41.

Haghighatlari M, Li J, Heidar-Zadeh F, Liu Y, Guan X, Head-Gordon T. Learning to make chemical predictions: The interplay of feature representation, data, and machine learning methods. Chem. 2020;6(7):1527-42. https://doi.org/10.1016/j.chempr.2020.05.014.

Zhang Y, Wen C, Wang C, Antonov S, Xue D, Bai Y, et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. Acta Mater. 2020;185:528-39.

Kaufmann K, Maryanovsky D, Mellor WM, Zhu C, Rosengarten AS, Harrington TJ, et al. Discovery of high-entropy ceramics via machine learning. npj Comput Mater. 2020;6(1):42.

Huang EW, Lee WJ, Singh SS, Kumar P, Lee CY, Lam TN, et al. Machine-learning and high-throughput studies for high-entropy materials. Mater Sci Eng R Rep. 2022;147:100645.

Dai D, Xu T, Wei X, Ding G, Xu Y, Zhang J, et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. Comput Mater Sci. 2020;175:109618.

Rao Z, Tung PY, Xie R, Wei Y, Zhang H, Ferrari A, et al. Machine learning-enabled high-entropy alloy discovery. Science. 2022;378(6615):78-85.

Zhang J, Xu B, Xiong Y, Ma S, Wang Z, Wu Z, et al. Design high-entropy carbide ceramics from machine learning. npj Comput Mater. 2022;8(1):5.

Machaka R, Motsi GT, Raganya LM, Radingoana PM, Chikosha S. Machine learning-based prediction of phases in high-entropy alloys: A data article. Data Brief. 2021;38:107346. https://doi.org/10.1016/j.dib.2021.107346.

Wen C, Zhang Y, Wang C, Xue D, Bai Y, Antonov S, et al. Machine learning assisted design of high entropy alloys with desired property. Acta Mater. 2019;170:109-17.

Kaufmann K, Vecchio KS. Searching for high entropy alloys: A machine learning approach. Acta Mater. 2020;198:178-222.

De Breuck PP, Hautier G, Rignanese GM. Materials property prediction for limited datasets enabled by feature selection and joint learning with MODNet. npj Comput Mater. 2021;7(1):83.

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. npj Comput Mater. 2017;3(1):54.

Jain A, Bligaard T. Atomic-position independent descriptor for machine learning of material properties. Phys Rev B. 2018;98(21):214112.

Huang G, Guo Y, Chen Y, Nie Z. Application of machine learning in material synthesis and property prediction. Materials. 2023;16(17):5977.

Tavazza F, DeCost B, Choudhary K. Uncertainty prediction for machine learning models of material properties. ACS Omega. 2021;6(48):32431-40.

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

Zhou Q, Lu S, Wu Y, Wang J. Property-oriented material design based on a data-driven machine learning technique. J Phys Chem Lett. 2020;11(10):3920-7.

Zhang J, Zhu Z, Xiang XD, Zhang K, Huang S, Zhong C, et al. Machine learning prediction of superconducting critical temperature through the structural descriptor. J Phys Chem C. 2022;126(20):8922-7.

Seko A, Hayashi H, Nakayama K, Takahashi A, Tanaka I. Representation of compounds for machine-learning prediction of physical properties. Phys Rev B. 2017;95(14):144110.