

ORIGINAL RESEARCH

Open access

Separating Epistemic from Aleatoric Uncertainty in Materials Graph Networks for High-Risk Predictions

Chinedu Okafor^{1*}, Amina Bello¹

Abstract

High-risk materials predictions in fields such as nuclear reactor design, aerospace component qualification, and advanced energy storage systems demand not only accurate property forecasts but also uncertainty estimates that can be trusted under regulatory scrutiny. Materials graph neural networks have emerged as powerful tools for property prediction across composition and structure spaces, yet the overwhelming majority output a single scalar uncertainty that conflates two fundamentally different sources: epistemic uncertainty arising from incomplete or sparse training data and aleatoric uncertainty arising from irreducible noise in the underlying data-generation process itself. This conceptual framework paper demonstrates that failure to separate these uncertainty types leads to flawed risk assessments and inefficient decision-making. Epistemic uncertainty signals the need for additional data collection or targeted active learning, whereas aleatoric uncertainty signals the need for higher-fidelity data sources such as improved density-functional approximations or experimental validation. The proposed modular framework equips materials graph networks with dedicated estimators for each uncertainty type, an aggregator that applies risk-adjusted safety margins, a decision gate that triggers appropriate remediation actions, and a calibration certificate required for deployment in safety-critical contexts. Operational criteria are defined to validate the separation in practice, including data-scaling behavior, source-identification tests, and active-learning efficiency. By enabling precise mapping of uncertainty to actionable protocols, the framework bridges a critical gap between current uncertainty-quantification practices in graph neural networks and the stringent reliability requirements of high-risk materials engineering. It builds directly on foundational distinctions introduced for deep learning while adapting them to the unique challenges of crystal-graph representations and multi-fidelity materials data. The result is a conceptual blueprint that supports safer model deployment, more efficient discovery campaigns, and clearer pathways to regulatory approval.

Keywords Epistemic uncertainty, Aleatoric uncertainty, Uncertainty decomposition, Materials graph networks, High-risk predictions, Safety-critical materials

*Correspondence:

Chinedu Okafor
chinedu.okafor@gmail.com

¹ Department of Data-Driven Materials Engineering, Faculty of Engineering, University of Lagos, Lagos, Nigeria

Introduction

A nuclear reactor design team relies on a machine-learning-predicted formation energy for a novel cladding material. The model returns a value with an uncertainty of ± 0.05 eV per atom. The team must decide whether this uncertainty is tolerable for safety certification. Is the ± 0.05 eV per atom the result of insufficient training data covering the relevant composition space (epistemic uncertainty), or does it reflect irreducible errors inherent to the density-functional theory calculations used to generate the training labels (aleatoric uncertainty)? The distinction is not academic [1]. If the uncertainty is epistemic, the rational next step is to acquire more data through targeted calculations or experiments. If the uncertainty is aleatoric, the rational next step is to improve the fidelity of the data source itself—perhaps by moving to a

higher-level theory or incorporating experimental benchmarks [2]. In high-risk domains such as nuclear engineering, aerospace, and grid-scale energy storage, this diagnostic capability is safety-critical [3].

Materials graph neural networks have become the dominant architecture for property prediction because they naturally encode crystal symmetry and local atomic environments [4, 5]. Equivariant graph networks, for example, deliver state-of-the-art accuracy on interatomic potentials and formation energies while remaining data-efficient [4]. Yet nearly all existing implementations collapse epistemic and aleatoric contributions into a single uncertainty figure [6, 7]. This mixing prevents practitioners from knowing whether a high-uncertainty prediction can be improved by simply enlarging the training set or whether the underlying data-generation pipeline has reached its noise floor [8].

The present conceptual framework addresses this limitation directly. It provides a modular architecture that separates the two uncertainty types within materials graph networks, defines operational criteria for verifying the separation, and links the decomposed uncertainties to concrete decision protocols tailored for high-risk applications [9, 10]. The framework draws on established uncertainty distinctions from Bayesian deep learning and extends them to the materials domain, where data are generated by heterogeneous computational and experimental sources and where safety margins must be defensible to regulators [1, 11].

Previous work on Gaussian process regression for materials has long acknowledged the presence of both uncertainty types, yet practical implementations typically treat noise as homoscedastic and inseparable [2]. Deep-ensemble and Monte-Carlo-dropout approaches applied to graph networks similarly report total variance without decomposition [11]. Even recent attempts to quantify uncertainty in graph neural networks for molecular or materials properties remain limited to total uncertainty or focus on calibration alone [12, 13]. None provide the explicit separation required for high-stakes decision-making [14, 15].

The consequences of this gap are tangible. In aerospace alloy design, an over-confident model that hides large epistemic uncertainty may lead to premature certification of a component that later fails under extreme conditions [3]. In battery electrolyte screening, failure to recognize dominant aleatoric uncertainty may result in selection of a solvent whose predicted stability is limited by the accuracy of the reference calculations rather than by true physical behavior [16]. Regulatory bodies in nuclear and aerospace sectors increasingly demand evidence that model uncertainty has been reduced to acceptable levels before deployment; without separation, such evidence cannot be provided [17, 18].

Defining Epistemic and Aleatoric Uncertainty

Epistemic uncertainty is model uncertainty. It originates from limitations in the training data—insufficient coverage of composition space, missing structural prototypes, or incomplete sampling of thermodynamic conditions [1, 7]. Because it reflects what the model does not yet know, epistemic uncertainty can be reduced by acquiring more or better-targeted data. It is therefore also called reducible uncertainty. In a materials graph network, epistemic uncertainty is high when the crystal graph presented at inference lies far from the training distribution in the learned representation space [14, 19].

Aleatoric uncertainty is data uncertainty. It arises from inherent noise or variability in the data-generation process itself. In materials science this noise commonly stems from density-functional approximations, pseudopotential choices, numerical integration tolerances, or stochastic physical processes such as thermal vibrations [2, 16]. Because the noise is a property of the data source rather than of the model, aleatoric uncertainty cannot be reduced simply by adding more data generated by the same method. It can only be lowered by switching to a higher-fidelity data source [20].

For materials graph networks the distinction maps cleanly onto domain realities [21]. Epistemic sources include sparse sampling across the vast inorganic composition space, under-representation of complex defect structures, and limited exploration of pressure–temperature conditions [6]. Aleatoric sources include systematic errors introduced by the exchange-correlation functional, basis-set incompleteness, or experimental measurement scatter [2, 8]. A single total-

uncertainty value reported by a graph network obscures this separation and therefore obscures the correct remediation pathway [9].

Table 1 establishes a functional mapping between uncertainty types, their scaling behavior, and the corresponding intervention regimes required for safety-critical materials decision-making.

Table 1. Functional Decomposition of Uncertainty Types into Distinct Remediation Regimes in High-Risk Materials Prediction

Uncertainty Type	Origin Mechanism	Mathematical Behavior Under Data Scaling	Intervention Lever	Failure Mode if Misinterpreted	Regulatory Interpretation
Epistemic Uncertainty	Incomplete sampling of composition–structure space	Monotonically decreases with dataset expansion	Active learning; targeted sampling; dataset augmentation	Over-conservative rejection of viable materials due to assumed irreducibility	Must be demonstrably minimized prior to deployment
Aleatoric Uncertainty	Noise intrinsic to data-generation process (DFT approximations, experimental variability)	Asymptotically constant regardless of dataset size	Upgrade fidelity of data source; switch computational method; experimental validation	Underestimation of irreducible risk leading to unsafe deployment	Defines lower bound of achievable reliability
Mixed (Unseparated) Uncertainty	Conflation of epistemic and aleatoric sources in total variance	Non-interpretable scaling behavior	No valid intervention pathway	Ambiguous decision-making; inefficient resource allocation	Not acceptable for certification in high-risk domains
Dominant Epistemic Regime	Sparse or extrapolative prediction region	Rapid decay under targeted sampling	Focused data acquisition campaigns	Premature model rejection	Acceptable if reducibility is demonstrated
Dominant Aleatoric Regime	High intrinsic noise in reference data	Persistent variance floor	Improve physical modeling fidelity	False confidence in data-driven improvements	Requires explicit safety margin justification

The foundational distinction between epistemic and aleatoric uncertainty was articulated for computer-vision tasks and later adapted to deep learning through Bayesian approximations [1, 22]. When applied to materials, however, the definitions must be interpreted through the lens of heterogeneous data pipelines. A graph network trained exclusively on PBE-level calculations will exhibit aleatoric uncertainty that reflects PBE’s known limitations; adding more PBE data cannot shrink that component [2]. In contrast, the epistemic component will shrink as the network encounters more diverse crystal graphs [23].

High-risk predictions amplify the practical importance of the distinction. A safety margin that treats all uncertainty as irreducible over-designs components unnecessarily when the uncertainty is actually epistemic and therefore reducible. Conversely, treating all uncertainty as reducible leads to premature deployment when the dominant contribution is aleatoric and cannot be lowered without changing the reference method [3, 18]. The conceptual framework therefore

treats separation not as an optional refinement but as a prerequisite for trustworthy use in any context where human safety or environmental integrity is at stake [17].

Why Separation is Critical for High-Risk Predictions

Disentangling epistemic from aleatoric uncertainty yields a set of tightly coupled advantages that become decisive once model outputs guide high-risk decisions [8, 10, 24]. Divergent remediation pathways illustrate the point: joint elevation of both uncertainties signals the need for expanded training data, whereas low epistemic uncertainty alongside persistently high aleatoric noise indicates that the model has saturated the available distribution and that improvement depends on higher-fidelity measurements or calculations [2, 16]. When epistemic uncertainty dominates, targeted sampling within the existing chemical space remains effective, while uniformly low levels justify conditional trust within stated safety margins. Collapsing these regimes into a single undifferentiated signal obscures actionable distinctions and erodes interpretability [7].

A related implication concerns safety margins, which must be anchored to irreducible variability. In applications such as nuclear cladding or aerospace turbine blades, the aleatoric floor defines the minimum admissible buffer because it cannot be reduced without altering the underlying data-generating process [3]. Epistemic contributions, by contrast, can be systematically minimized prior to deployment through active learning, aligning uncertainty management with regulatory expectations that reducible components be exhausted [15, 20]. Beyond this immediate constraint, certification pathways depend on explicit decomposition, as regulators require clear differentiation between uncertainty that can be mitigated and uncertainty intrinsic to the computational or experimental pipeline [17, 18].

This distinction also reshapes data acquisition strategies. Prioritizing epistemic uncertainty accelerates convergence toward reliable models, whereas attempts to reduce aleatoric noise through additional sampling are inherently inefficient given its fixed origin [9, 11]. The consequence is a more disciplined allocation of experimental or computational resources in high-risk campaigns [23]. At the level of model evaluation, identical aggregate uncertainty can conceal fundamentally different risk profiles: models dominated by epistemic uncertainty remain improvable through data expansion, whereas those constrained by aleatoric noise demand changes in reference methodology [8, 10]. The capacity to align uncertainty structure with project constraints transforms uncertainty from a limiting factor into an operational decision variable, rendering its separation indispensable for credible deployment in safety-critical materials engineering [14].

Existing Methods and their Limitations for High-Risk Contexts

Contemporary uncertainty-quantification strategies for graph neural networks can be grouped into several methodological families, yet each entangles epistemic and aleatoric contributions in ways that limit applicability in high-risk materials prediction [7, 14, 25]. Ensemble approaches, which infer uncertainty from variance across independently trained models, provide scalability and conceptual simplicity but fail to distinguish between disagreement arising from model insufficiency and variability inherent to the data [11]. This ambiguity undermines both safety-margin construction and the design of targeted learning strategies [11, 20].

Bayesian neural networks offer a probabilistic alternative by approximating weight posteriors, typically through variational inference or Monte Carlo dropout. In practice, however, the resulting predictive variance still conflates uncertainty sources, often exhibits calibration deficiencies, and incurs computational costs that scale poorly with the graph sizes characteristic of inorganic materials systems [22, 26, 27]. These limitations become particularly restrictive in regulatory settings, where explicit evidence of reducible uncertainty reduction is required but cannot be isolated within such formulations [3].

Gaussian process regression retains a privileged position due to its formal treatment of predictive variance, yet its standard assumption of homoscedastic noise conflicts with the pronounced heteroscedasticity observed across compositional and structural regimes in materials datasets [2, 8]. Misalignment between assumed and actual noise structure leads to systematic miscalibration, especially in extrapolative regimes central to high-risk discovery. Learned-aleatoric models partially address this issue by introducing input-dependent noise estimates, but the learned parameter absorbs both data and model uncertainty, preserving the core ambiguity that complicates downstream decision-making [1, 12, 16].

Even recent attempts to enforce explicit decomposition within graph-based architectures remain incomplete with respect to high-risk requirements [6, 9, 13]. The absence of standardized validation criteria for separation, coupled with a lack of integration between uncertainty estimates and risk-adjusted decision mechanisms, limits their operational utility. Without embedded links to regulatory thresholds or safety multipliers, these approaches fall short of providing the comprehensive framework necessary for safety-critical deployment, motivating the need for a more systematically integrated solution.

Modular Framework for High-Risk Predictions

The proposed framework integrates six interdependent components within a unified materials graph network architecture, ensuring that uncertainty estimation and decision-making remain structurally aligned [6, 10]. At its core, a shared feature extractor maps the input crystal graph—representing atoms as nodes and interactions as edges—into a common latent space, thereby enforcing informational consistency across all downstream modules [4, 7].

Building on this representation, epistemic uncertainty is captured through an ensemble-based module in which multiple graph networks, diversified by initialization, data subsampling, or architectural variation, quantify the sensitivity of predictions to training data limitations [11, 20]. Reliable deployment requires that this ensemble exhibits sufficient diversity to stabilize variance estimates under repeated training conditions [15]. In parallel, aleatoric uncertainty is inferred by leveraging multi-fidelity reference data, allowing a dedicated sub-network to learn systematic discrepancies across computational approximations and experimental measurements, thereby isolating the irreducible noise floor [2, 8, 16].

These estimates are subsequently combined within an aggregation layer that introduces a domain-specific safety multiplier, calibrated to reflect acceptable risk levels in applications where conservative bounds are mandatory [3, 17]. The framework then operationalizes uncertainty through a decision gate that evaluates epistemic and aleatoric components against predefined thresholds. Exceedance of epistemic limits triggers targeted data acquisition via active learning, whereas elevated aleatoric levels indicate fundamental inadequacy of the data source for the intended application [15, 16, 23]. Only when both criteria are satisfied does the system authorize deployment, accompanied by the computed safety margin.

Prior to use, the framework produces a calibration certificate comprising reliability diagnostics, coverage guarantees, and, where necessary, conformal prediction adjustments, thereby generating auditable evidence aligned with regulatory expectations [8, 10]. The model is shown in **figure 1**.

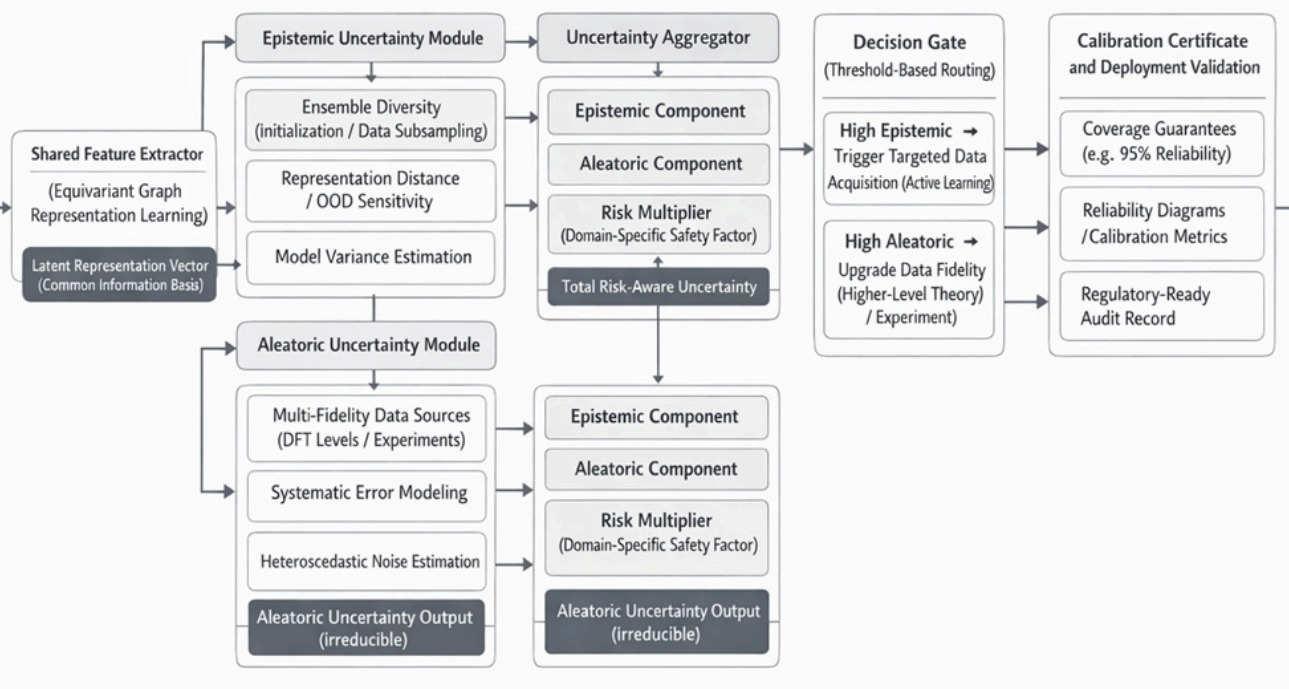


Figure 1. The full modular architecture, illustrating the strict separation of epistemic and aleatoric uncertainty pathways and their integration into risk-aware decision routing for safety-critical deployment.

The modular design ensures that each component can be implemented independently while preserving end-to-end traceability. Because the framework is architecture-agnostic, it can be retrofitted to existing equivariant graph networks or future message-passing variants [4, 28]. Its explicit separation of uncertainty types, risk-aware aggregation, and decision-gate logic together provide the conceptual foundation required for high-risk materials predictions [14, 15].

Operational Criteria for Separation in High-Risk Contexts

To ensure that epistemic and aleatoric uncertainty are meaningfully separated rather than merely declared, the framework defines five operational criteria that must be satisfied before any materials graph network is certified for high-risk deployment [7, 9]. These criteria are testable in principle and provide auditable evidence for regulators and practitioners alike [8, 10].

Table 2 formalizes uncertainty separation as a certification problem by translating conceptual criteria into testable diagnostic signatures required for regulatory approval.

Table 2. Operational Validation Criteria as a Certification Framework for Uncertainty Separation in Materials Graph Networks

Criterion	Test Structure	Expected Diagnostic Signature	Failure Signal	Implication for Deployment
Data-Scaling Test	Incrementally increase dataset size	Epistemic ↓ toward zero; Aleatoric ≈ constant	Both decrease or both remain high	Separation invalid; model not certifiable
Source-Identification Test	Same inputs with different data sources	Aleatoric varies significantly; Epistemic	Epistemic fluctuates with data source	Noise attribution failure

		stable		
Active-Learning Efficiency Test	Compare epistemic-guided vs random sampling	≥50% faster error reduction using epistemic guidance	No efficiency gain	Epistemic estimate not actionable
Noise-Floor Convergence Test	Large-scale training limit	Total uncertainty → aleatoric floor	Total continues decreasing below noise	Aleatoric underestimation
Calibration Test	Coverage validation across datasets	Empirical coverage matches predicted intervals	Under/over-coverage	Unreliable uncertainty bounds
Decision-Gate Consistency	Evaluate routing decisions across scenarios	Correct mapping to remediation actions	Misrouting (e.g., data collection when aleatoric dominates)	Unsafe operational behavior

As the training set expands, epistemic uncertainty is expected to diminish asymptotically toward zero, whereas the aleatoric component remains effectively invariant [15, 20]. In high-risk settings, this expectation becomes a strict requirement: once data sufficiency is achieved, epistemic uncertainty must fall below 0.1 times the aleatoric floor [7, 11]. Such convergence substantiates the reducibility of model uncertainty and indicates that continued data acquisition remains a productive intervention. A related diagnostic emerges when identical crystal graphs are annotated using distinct data-generation protocols, such as PBE and SCAN density functionals, where aleatoric uncertainty should diverge markedly while epistemic estimates remain comparable [2, 8]. Enforcing a separation factor exceeding five ensures that variability is attributed to the data source rather than mischaracterized as model incompleteness [16].

This logic extends directly to learning dynamics, where query strategies guided solely by epistemic uncertainty must achieve at least a 50 % faster reduction in prediction error relative to random sampling [15, 23]. Such differential efficiency confirms that the reducible component has been correctly isolated for targeted refinement [20]. As the training regime approaches saturation, total predictive uncertainty should converge to the independently estimated aleatoric floor [2, 9], which, in safety-critical applications, must remain demonstrably below the domain-specific threshold [3]. Beyond convergence, reliability hinges on calibration: predicted uncertainty intervals are required to attain nominal coverage—such as 95 %—across both in-distribution and mildly out-of-distribution samples [8, 10]. The integration of conformal prediction further strengthens this guarantee by providing distribution-free coverage that remains externally verifiable [14]. Together, these conditions transform uncertainty decomposition into a verifiable engineering specification grounded in the principles of Kendall and Gal [1] while extending their applicability to heterogeneous, multi-fidelity materials data pipelines [2, 16], thereby ensuring that epistemic uncertainty can be systematically suppressed prior to deployment and that aleatoric uncertainty is correctly interpreted as the irreducible limit imposed by the data source [17, 18].

Relation to Decision-Making under Uncertainty

The separation of uncertainty components maps directly onto operational decision protocols governing active learning, risk evaluation, and certification processes [15, 23, 24]. Within active-learning loops, epistemic uncertainty functions as the acquisition signal, prioritizing crystal graphs whose inclusion most effectively reduces model ignorance [15, 20]. As this component declines below its safety threshold, further data collection ceases to be justified on epistemic grounds, whereas aleatoric uncertainty remains excluded from acquisition logic due to its irreducible nature [2, 16]. This distinction becomes consequential in risk assessment, where the interaction between low and high regimes of each uncertainty type defines deployment readiness: elevated epistemic uncertainty precludes use irrespective of aleatoric magnitude, while low epistemic but high aleatoric uncertainty indicates that the model has saturated available knowledge yet remains constrained by data quality, necessitating expanded safety margins [3]. When both components are minimal, predictions can be deployed under standard safety assumptions [17]. Such differentiation avoids both excessive conservatism and unwarranted confidence in predictions limited by intrinsic data noise [18].

Regulatory evaluation is correspondingly streamlined through the provision of auditable evidence aligned with certification requirements [3, 17]. Demonstrating that epistemic uncertainty has been reduced to negligible levels via documented active-learning procedures establishes model completeness [15, 20], while the residual aleatoric component defines the irreducible reliability bound [2]. Calibration certificates then verify that total uncertainty, adjusted by risk-dependent multipliers, satisfies domain-specific acceptability criteria [8, 10]. By disentangling these components, the framework replaces opaque scalar uncertainty with a transparent and verifiable risk narrative [14, 29].

Relation to Existing Safety Frameworks

The modular architecture is intended to augment, rather than displace, established safety methodologies, enabling integrated certification pathways [3, 17]. Conformal prediction contributes distribution-free coverage guarantees particularly suited to high-risk environments [8, 10], and when applied to already decomposed uncertainty, produces tighter and more interpretable prediction sets due to explicit isolation of the aleatoric noise floor [14]. Bayesian deep-learning approaches retain theoretical rigor in uncertainty quantification but do not inherently provide decomposition [1, 22]; embedding dedicated epistemic and aleatoric modules downstream of shared representations extends these methods into operationally actionable forms [7], preserving probabilistic foundations while enabling separation required for safety validation [20].

A complementary relationship also arises with out-of-distribution detection, which signals epistemic vulnerability arising from data sparsity [26, 27]. The epistemic uncertainty module aligns naturally with such signals, as both reflect limitations in learned representations [14]. Under these conditions, high epistemic uncertainty and strong out-of-distribution indicators converge on the same corrective mechanism—targeted data acquisition—while aleatoric estimates remain unaffected [23]. This alignment is particularly salient in materials discovery, where extrapolation beyond existing compositional space is routine [13]. By integrating separation with these established approaches, the framework combines statistical guarantees with diagnostically precise uncertainty attribution [3, 17].

Implications for Materials GNN Design for High-Risk Applications

The proposed framework introduces consequential design and governance implications across development, deployment, and regulation [4, 14, 28]. Model construction must shift toward explicitly modular architectures in which feature extraction is decoupled from uncertainty estimation, allowing epistemic and aleatoric components to evolve independently and be validated against operational criteria [6, 7]. This requirement extends to reporting standards, where calibration certificates and validation outcomes accompany model dissemination [9, 10], and only systems satisfying all criteria warrant classification as suitable for high-risk contexts [8].

Deployment practices are similarly reconfigured, as the presence of decomposed uncertainties becomes a prerequisite for application in safety-critical domains such as nuclear systems, aerospace, and energy storage [17, 18]. The framework enables unambiguous diagnosis of performance limitations: elevated epistemic uncertainty motivates targeted data expansion, whereas elevated aleatoric uncertainty indicates the need for higher-fidelity measurement or simulation protocols [2, 16]. Safety margins can therefore be calibrated transparently in accordance with application-specific risk tiers [10].

For regulatory bodies, the framework establishes a standardized evidentiary structure for certification, encompassing decomposed uncertainty outputs, decision thresholds, and calibration verification [18]. Approval processes can thus be anchored in demonstrable reduction of epistemic uncertainty below prescribed limits alongside validation that the aleatoric floor remains acceptable under operational conditions [8]. This increased granularity reduces interpretive

ambiguity, accelerates review timelines, and strengthens confidence in machine-learning-assisted materials qualification [14].

Conclusion

High-risk materials predictions require separating epistemic from aleatoric uncertainty in materials graph networks. The conceptual framework presented here meets that requirement through a modular architecture comprising a shared feature extractor, dedicated epistemic and aleatoric modules, an aggregator with risk-adjusted safety margins, a decision gate that routes to appropriate remediation, and a calibration certificate for regulatory scrutiny. Five operational criteria—data scaling, source identification, active-learning efficiency, noise-floor convergence, and interval calibration—provide auditable verification that separation has been achieved.

The framework directly informs decision-making under uncertainty by prioritising epistemic reduction via active learning, anchoring safety margins to the aleatoric floor, and furnishing regulators with a transparent risk narrative . It integrates seamlessly with conformal prediction, Bayesian deep learning, and out-of-distribution detection while overcoming the core limitation shared by all prior methods: the inability to distinguish reducible from irreducible uncertainty in safety-critical contexts.

By adopting this separation-aware approach the materials-science community can move from opaque total-uncertainty scalars to trustworthy, actionable uncertainty diagnostics. The result will be safer deployment of graph-network predictions, more efficient discovery campaigns, and clearer regulatory pathways for the next generation of high-performance materials in nuclear, aerospace, and energy-storage applications. The conceptual blueprint is now available; its implementation across laboratories and regulatory frameworks is the logical next step.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 27 May 2023

Revised: 06 Oct 2023

Accepted: 09 Dec 2023

Published online: 18 January 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv Neural Inf Process Syst.* 2017;30:5574-84.
<https://doi.org/10.5555/3295222.3295309>.
- Deringer VL, Bartók AP, Bernstein N, Wilkins DM, Ceriotti M, Csányi G. Gaussian process regression for materials and molecules. *Chem Rev.* 2021;121(16):10073-141.
<https://doi.org/10.1021/acs.chemrev.1c00022>.
- Burton S, Herd B. Addressing uncertainty in the safety assurance of machine-learning. *Front Comput Sci.* 2023;5:1132580.
<https://doi.org/10.3389/fcomp.2023.1132580>.
- Batzner S, Musaelian A, Sun L, Geiger M, Mailoa JP, Kornbluth M, et al. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat Commun.* 2022;13(1):2453.
<https://doi.org/10.1038/s41467-022-29939-5>.
- Corso G, Stärk H, Jegelka S, Jaakkola T, Barzilay R. Graph neural networks. *Nat Rev Methods Primers.* 2024;4(1):17.
<https://doi.org/10.1038/s43586-024-00294-7>.
- Jiang S, Qin S, Van Lehn RC, Balaprakash P, Zavala VM. Uncertainty quantification for molecular property predictions with graph neural architecture search. *Digit Discov.* 2024;3(8):1534-53.
<https://doi.org/10.1039/D4DD00088A>.
- Munikoti S, Agarwal D, Das L, Natarajan B. A general framework for quantifying aleatoric and epistemic uncertainty in graph neural networks. *Neurocomputing.* 2023;521:1-10.
<https://doi.org/10.1016/j.neucom.2022.11.049>.
- Gruich CJ, Madhavan V, Wang Y, Goldsmith BR. Clarifying trust of materials property predictions using neural networks with distribution-specific uncertainty quantification. *Mach Learn Sci Technol.* 2023;4(2):025019.
<https://doi.org/10.1088/2632-2153/accace>.
- Musielewicz J, Lan J, Uyttendaele M, Kitchin JR. Improved uncertainty estimation of graph neural network potentials using engineered latent space distances. *J Phys Chem C.* 2024;128(49):20799-810.
<https://doi.org/10.1021/acs.jpcc.4c04972>.
- Varivoda D, Dong R, Omee SS, Hu J. Materials property prediction with uncertainty quantification: A benchmark study. *Appl Phys Rev.* 2023;10(2):021409.
<https://doi.org/10.1063/5.0133528>.
- Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Adv Neural Inf Process Syst.* 2017;30:6402-13.
<https://doi.org/10.5555/3295222.3295387>.
- Kwon Y, Lee D, Choi YS, Kang S. Uncertainty-aware prediction of chemical reaction yields with graph neural networks. *J Cheminform.* 2022;14(1):2.
<https://doi.org/10.1186/s13321-021-00579-z>.

Wang S, Yue H, Yuan X. Accelerating polymer discovery with uncertainty-guided PGCNN: Explainable AI for predicting properties and mechanistic insights. *J Chem Inf Model*. 2024;64(14):5500-9.
<https://doi.org/10.1021/acs.jcim.4c00555>.

Wang F, Liu Y, Liu K, Wang Y, Medya S, Yu PS. Uncertainty in graph neural networks: A survey. *arXiv [Preprint]*. 2024. arXiv:2403.07185

Ying R, Bourgeois D, You J, Zitnik M, Leskovec J. GNNExplainer: Generating explanations for graph neural networks. *Adv Neural Inf Process Syst*. 2019;32:9240-51.
<https://doi.org/10.5555/3454287.3455116>.

Soleimany AP, Amini A, Goldman S, Rus D, Bhatia SN, Coley CW. Evidential deep learning for guided molecular property prediction and discovery. *ACS Cent Sci*. 2021;7(8):1356-67.
<https://doi.org/10.1021/acscentsci.1c00546>.

Soize C. Uncertainty quantification: An accelerated course with advanced applications in computational engineering. Cham: Springer International Publishing; 2017. xxii, 329 p.
<https://doi.org/10.1007/978-3-319-54339-0>.

Dawood T, Chen C, Sidhu BS, Ruijsink B, Gould J, Porter B, et al. Uncertainty aware training to improve deep learning model calibration for classification of cardiac MR images. *Med Image Anal*. 2023;88:102861.
<https://doi.org/10.1016/j.media.2023.102861>.

Fuchsgruber D, Wollschläger T, Günnemann S. Energy-based epistemic uncertainty for graph neural networks. *Adv Neural Inf Process Syst*. 2024;37:34378-428.
<https://doi.org/10.52202/079017-1084>.

Busk J, Schmidt MN, Winther O, Vegge T, Jørgensen PB. Graph neural network interatomic potential ensembles with calibrated aleatoric and epistemic uncertainty on energy and forces. *Phys Chem Chem Phys*. 2023;25(37):25828-37.
<https://doi.org/10.1039/D3CP02143B>.

Wu L, Cui P, Pei J, Zhao L, Guo X. Graph neural networks: Foundation, frontiers and applications. In: *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery; 2022. p. 4840-1.
<https://doi.org/10.1145/3534678.3542609>.

Li Y, Gal Y. Dropout inference in Bayesian neural networks with alpha-divergences. In: *Proceedings of the 34th International Conference on Machine Learning*. *Proc Mach Learn Res*. 2017;70:2052-61.

Stadler M, Charpentier B, Geisler S, Zügner D, Günnemann S. Graph posterior network: Bayesian predictive uncertainty for node classification. *Adv Neural Inf Process Syst*. 2021;34:18033-48.
<https://doi.org/10.5555/3540261.3541641>.

Alzate-Mejía N, Santos-Boada G, de Almeida-Amazonas JR. Decision-making under uncertainty for the deployment of future hyperconnected networks: A survey. *Sensors (Basel)*. 2021;21(11):3791.
<https://doi.org/10.3390/s21113791>.

Stocker S, Gasteiger J, Becker F, Günnemann S, Margraf JT. How robust are modern graph neural network potentials in long and hot molecular dynamics simulations? *Mach Learn Sci Technol*. 2022;3(4):045010.
<https://doi.org/10.1088/2632-2153/ac9955>.

Zhao X, Chen F, Hu S, Cho JH. Uncertainty aware semi-supervised learning on graph data. *Adv Neural Inf Process Syst*. 2020;33:12827-36.
<https://doi.org/10.5555/3495724.3496800>.

Zhao X, Chen F, Hu S, Cho JH. Uncertainty-aware prediction for graph neural networks. OpenReview [Preprint]. 2019.

Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst.* 2021;32(1):4-24.

<https://doi.org/10.1109/TNNLS.2020.2978386>.

Jiang J, Ling C, Li H, Bai G, Zhao X, Zhao L. Quantifying uncertainty in graph neural network explanations. *Front Big Data.* 2024;7:1392662.

<https://doi.org/10.3389/fdata.2024.1392662>.