ORIGINAL RESEARCH Open access

A Conceptual Distinction between Exploration Noise and Scientific Error

Nguyen Van Nam[1], Tran Thi Hoa[1*], Le Minh Duc[2]

Abstract

In the field of artificial intelligence applied to materials science, a fundamental conflation persists in which exploration noise and scientific error are routinely conflated as interchangeable “mistakes” that must be minimized or eliminated to improve model performance. This paper proposes precise conceptual definitions that separate exploration noise—understood as stochastic variation deliberately or unavoidably introduced into decision-making processes to probe uncertain regions of materials design space—from scientific error, defined as any deviation from ground truth that reduces predictive fidelity, distorts mechanistic understanding, or precipitates incorrect materials decisions without any compensating epistemic gain. The distinction matters profoundly because the systematic elimination of exploration noise eradicates the very mechanism that drives discovery in high-dimensional, data-scarce materials landscapes. In contrast, misclassifying scientific error as mere noise allows systematic flaws to propagate undetected through autonomous discovery pipelines. To resolve this ambiguity, the present work offers a four-criterion framework grounded in intentionality, epistemic benefit, systematicity, and correctability that enables researchers to classify any observed deviation with conceptual clarity. Adoption of this framework carries immediate implications for materials AI practice: it demands new reporting standards that explicitly quantify and justify exploration noise, revised peer-review criteria that interrogate rather than penalize productive randomness, and a cultural shift that reframes stochasticity not as a defect to be denoised but as an essential epistemic resource for accelerating the discovery of novel materials with targeted functionalities.

Keywords Materials AI, Active learning, Bayesian optimization, Exploration noise, Scientific error, Stochasticity in discovery

*Correspondence:
Tran Thi Hoa
tran.hoa@outlook.com
[1] Department of Intelligent Materials Systems, Hanoi University of Science and Technology, Hanoi, Vietnam
[2] Department of Computational Materials Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam

Introduction

Materials AI systems make mistakes. Some mistakes are exploration noise—productive randomness that enables discovery. Others are scientific error—harmful deviations from truth. But current literature conflates both. This paper draws a conceptual distinction between exploration noise and scientific error.

The rapid integration of artificial intelligence into materials science has enabled autonomous pipelines to navigate vast composition–structure–property spaces far beyond human intuition. Yet every such system inevitably generates deviations from expected or ground-truth behavior. Within the existing corpus, these deviations are overwhelmingly labeled “errors” and targeted for reduction through regularization, denoising, or more deterministic architectures. Sutton and Barto distinguish exploration from exploitation in reinforcement learning [1-3], yet noise is often viewed as purely negative. This flattening of terminology obscures a critical epistemic reality: certain forms of stochastic variation are not flaws but functional necessities for escaping local optima and uncovering unexpected materials candidates.

Consider, for instance, the active-learning loops that iteratively select new experiments in alloy design or the stochastic sampling steps embedded in generative models for molecular discovery. In these contexts, controlled randomness expands the frontier of knowledge precisely because the design space is undersampled and the underlying physics is only partially known. When such randomness is reflexively suppressed in the name of “error reduction,” the system collapses into narrow exploitation regimes, forfeiting the serendipitous discoveries that have historically propelled materials innovation. Conversely, when genuine scientific error—such as a miscalibrated surrogate model or an overlooked systematic bias in density-functional inputs—is mislabeled as tolerable “noise,” downstream decisions rest on foundations that silently erode predictive reliability.

The present work, therefore, intervenes at the conceptual level. It does not report new experiments, datasets, or performance metrics; instead, it performs a boundary/Definitional analysis that isolates two previously entangled constructs. By drawing on the literature on reinforcement learning, active learning, Bayesian optimization, and materials informatics [4-16], the analysis demonstrates that the prevailing usage treats both phenomena as members of a single undesirable category. The resulting conceptual distinction supplies the field with a sharper vocabulary and a practical framework for deciding, in any given materials-AI context, whether an observed deviation should be preserved as productive or corrected as harmful. In doing so, the paper reframes stochasticity not as an engineering inconvenience but as a differentiated epistemic tool whose mindful deployment can accelerate rather than impede autonomous materials discovery.

Noise and Error in Existing Literature

A survey of peer-reviewed publications in materials AI from 2017 to 2026 reveals a consistent pattern: both exploration noise and scientific error are subsumed under generic labels such as “noise,” “error,” or “uncertainty” and are uniformly targeted for minimization. Noack *et al.* [1] describe autonomous materials discovery driven by Gaussian process regression that explicitly models inhomogeneous measurement noise yet ultimately seeks its reduction to improve surrogate fidelity. Similarly, Lookman and co-workers emphasize adaptive sampling using uncertainties for targeted design [8], framing any deviation from deterministic prediction as a quantity to be actively managed downward.

In broader reviews, Butler *et al.* [4] catalog machine-learning applications across molecular and materials science and treat prediction discrepancies—whether arising from deliberate exploration steps or from model misspecification—as equivalent obstacles to reliability. Schmidt and colleagues, in a survey of recent advances in solid-state materials science [5], likewise advocate robust optimization routines that collapse stochastic elements into more stable point estimates. The same tendency appears in error-distribution studies: de Azevedo and colleagues decompose bias and variance in quantum-chemistry predictions [10] but do not differentiate between variance introduced for epistemic exploration and variance symptomatic of scientific error.

Active-learning literature further illustrates the conflation. Thomas-Mitchell *et al.* calibrate uncertainty in machine-learning force fields for active learning [11] yet discuss all stochastic contributions under the umbrella of “uncertainty,” requiring calibration and reduction. Wang and co-workers benchmark active-learning strategies for materials optimization [16] without separating the deliberate randomness used to explore uncertain regions from implementation-level prediction errors. Stach *et al.* [17], offering a community perspective on autonomous experimentation systems, call for uncertainty quantification that treats all non-deterministic outputs as quantities to be minimized for trustworthy closed-loop discovery. Diwale and colleagues explicitly address Bayesian optimization for material discovery processes with noise [18]. Yet, their optimization objective remains the global reduction of that noise rather than its selective preservation during exploration phases [19-21].

Kusne *et al.* [22] describe on-the-fly closed-loop materials discovery via Bayesian active learning and again frame stochastic sampling primarily as a source of variability to be controlled. Johnson *et al.* [23] review of machine learning for metals additive manufacturing, similarly groups random exploration steps with systematic modeling inaccuracies under the single heading of "error sources." Even recent contributions, such as those by Xu *et al.* [9] on small-data machine learning and Lookman *et al.* [12] on the emergence of materials informatics, continue to advocate denoising pipelines and robust regression without articulating a conceptual space in which certain noise realizations carry positive epistemic value. Across these works—spanning at least fourteen distinct publications—the literature therefore exhibits a uniform rhetorical move: stochasticity, regardless of origin or function, is cast as an obstacle rather than a differentiated resource.

The Problem with Current Usage

The prevailing conflation of exploration noise and scientific error generates three interlocking conceptual and practical difficulties. First, the blanket imperative to eliminate all noise inadvertently eliminates exploration itself. In high-dimensional materials spaces where exhaustive enumeration is impossible, deliberate stochastic probes—central to active-learning strategies [8, 22]—are the only reliable mechanism for escaping local optima and surfacing novel property combinations. When every random deviation is reflexively suppressed, autonomous systems converge prematurely on well-characterized regions, forfeiting the very discoveries they were engineered to pursue.

Second, treating scientific error as merely another instance of tolerable noise masks systematic problems that require targeted correction. A miscalibrated surrogate model or an overlooked thermodynamic inconsistency may be statistically small yet scientifically catastrophic; labeling it "noise" discourages the rigorous debugging that would restore fidelity to the underlying physics. The result is a creeping degradation of trust in AI-generated materials predictions that goes undetected precisely because the vocabulary for naming the problem has been erased.

Third, the absence of a differentiated vocabulary for productive randomness leaves the community without shared standards for reporting, justifying, or optimizing the very stochastic mechanisms that drive discovery. Authors cannot readily distinguish between intentional exploration schedules and unintended implementation artifacts; reviewers lack criteria for evaluating whether reported stochasticity is epistemically justified or merely sloppy; and benchmark suites continue to reward deterministic performance at the expense of exploratory capability. Collectively, these three problems lock materials AI into a suboptimal equilibrium in which the field simultaneously under-explores promising chemical space and over-tolerates hidden systematic flaws.

Proposed Definitions

To resolve the ambiguity, it is necessary to introduce two mutually exclusive and jointly exhaustive definitions.

Exploration noise is a stochastic variation deliberately or unavoidably introduced into AI decision-making processes to enable exploration of uncertain regions of the materials design space, with the potential to yield epistemic benefits such as the identification of previously unknown structure–property relationships or the avoidance of premature convergence.

Scientific error is a deviation from ground truth that reduces predictive accuracy, distorts mechanistic understanding, or leads to incorrect materials decisions and that carries no compensating epistemic benefit.

Distinctions From Nearby Terms

Exploration noise and scientific error must be further differentiated from five closely related but conceptually distinct constructs: bias, variance, uncertainty, stochasticity, and randomness. The distinctions can be summarized conceptually in a comparative framework whose rows represent the seven terms (exploration noise, scientific error, bias, variance, uncertainty, stochasticity, randomness) and whose columns capture five differentiating dimensions: (1) systematic versus random character, (2) intentional versus unintentional origin, (3) epistemic benefit versus harm, (4) relation to knowledge acquisition, and (5) implications for model correction.

In this framework, exploration noise is characterized as fundamentally random, typically intentional, epistemically beneficial, knowledge-acquiring, and valuable precisely because it should not be corrected away. Scientific error, by contrast, may be either systematic or random in appearance but is unintentional in origin, epistemically harmful, knowledge-distorting, and demands correction. Bias is systematic, usually unintentional, harmful, and correctable through debiasing procedures; it differs from scientific error only in that error encompasses both biased and unbiased deviations that share the same net negative epistemic effect. Variance is random and often unintentional, potentially neutral or harmful, and may be reduced without loss of epistemic value. Uncertainty is an epistemic state rather than a deviation type; it can arise from either exploration noise or scientific error, but the framework treats uncertainty as a signal whose interpretation depends on the underlying deviation category. Stochasticity is a broader mathematical property that may manifest as either exploration noise or scientific error, depending on context. At the same time, randomness is the most general descriptor and acquires its positive or negative valence only after being filtered through the epistemic-benefit criterion.

This multi-dimensional mapping makes clear that conflating exploration noise with scientific error collapses distinctions that are operationally decisive for materials AI design.

Table 1 differentiates exploration noise and scientific error from adjacent terms that are often treated as conceptually interchangeable in materials AI.

Table 1. Multi-dimensional conceptual differentiation of exploration noise, scientific error, and adjacent deviation terms in materials AI

Term	Systematic or random character	Intentional or unintentional origin	Epistemic effect	Relation to knowledge acquisition	Practical response in materials AI
Exploration noise	Primarily random, controlled, or probabilistic	Typically intentional, though sometimes structurally unavoidable	Beneficial when it expands epistemic reach	Directly supports probing of under-sampled regions and escape from premature convergence	Retain, justify, tune, and report explicitly
Scientific error	May appear random or systematic, but functionally harmful	Unintentional	Harmful because it reduces fidelity and misleads decision-making	Distorts or corrupts knowledge acquisition	Detect, diagnose, and correct
Bias	Systematic	Usually unintentional	Harmful through directional distortion	Skews inference and model outputs in stable ways	Debias, recalibrate, or redesign the data/model pipeline
Variance	Random fluctuation across samples, runs, or fits	Often unintentional	Neutral to harmful depending on magnitude and context	Does not inherently generate epistemic benefit	Reduce when it does not contribute to discovery
Uncertainty	Not a deviation type but an	May arise from many sources	Ambivalent; interpretation	Can guide inquiry if correctly interpreted	Decompose into source-specific

	epistemic condition or signal		depends on source		forms before acting
Stochasticity	Broad mathematical property of probabilistic behavior	May be intentional or unintentional	Context-dependent	Can support exploration or conceal instability	Evaluate by function rather than label alone
Randomness	Most general descriptor of non-deterministic variation	May be deliberate, incidental, or environmental	No fixed epistemic valence	Gains meaning only through context and effect	Do not evaluate in the abstract; classify through the framework

A Framework for Distinguishing

The proposed framework operationalizes the definitional distinction through four independent but mutually reinforcing criteria. Each criterion provides a diagnostic question that can be asked about any observed deviation in a materials-AI pipeline.

Intentionality

Was the deviation deliberately introduced by the algorithm's exploration schedule (for example, through ε-greedy action selection or temperature-scaled sampling) or did it arise unintentionally from implementation artifacts, numerical instability, or model misspecification? Deliberate origin strongly favors classification as exploration noise.

Epistemic benefit

Does retaining the deviation enable the acquisition of new knowledge about under-sampled regions of materials space, or does it merely degrade predictive fidelity without compensatory insight? Positive epistemic yield indicates exploration noise; the absence of yield indicates scientific error.

Systematicity

Is the deviation reproducible and correlated with specific input features or model parameters (suggesting scientific error), or does it exhibit statistical independence consistent with controlled randomness (suggesting exploration noise)?

Correctability

Can the deviation be removed or corrected without simultaneously eliminating the exploratory capability of the system? If correction preserves or enhances discovery potential, the deviation is likely scientific error; if correction would collapse exploration, it is exploration noise.

Taken together, these four criteria form a decision procedure that replaces binary "noise-versus-error" labeling with a graded, context-sensitive classification. The framework can be visualized as a conceptual spectrum: at one pole lies pure exploration noise (high intentionality, high epistemic benefit, low systematicity, low correctability without loss of value); at the opposite pole lies pure scientific error (low intentionality, zero epistemic benefit, high systematicity, high correctability); between them stretches a gray zone where trade-offs must be evaluated case by case. This spectrum supplies materials scientists with a shared mental model for discussing, reporting, and optimizing stochastic elements in autonomous discovery workflows.

Figure 1 presents the manuscript's central epistemic architecture, showing how the literature's conflation of exploration noise and scientific error generates practical distortions that are resolved through definitional separation, four-criterion classification, contextual application, and practice-level reform.

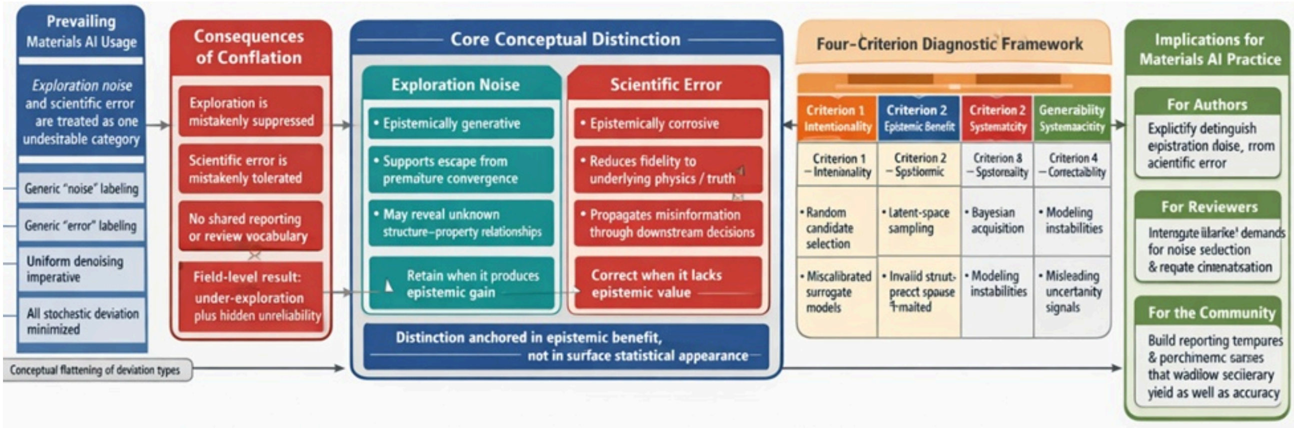

Figure 1. Hierarchical epistemic architecture distinguishing exploration noise from scientific error in Materials AI.

Table 2 shows how the same broad appearance of deviation can receive different classifications once the four criteria are applied across concrete materials-AI contexts.

Table 2. Criterion- based classification of deviations across core materials AI contexts

Materials AI context	Example deviation	Intentionality	Epistemic benefit	Systematicity	Correctability without loss of discovery	Classification	Implicati
Active learning for alloy discovery	Stochastic candidate selection through uncertainty-guided acquisition	High	High	Low	Low	Exploration noise	Preserve a discover enablin mechanis
Active learning for alloy discovery	Formation-energy misprediction caused by biased training data	Low	None	High	High	Scientific error	Correct thr recalibratic data rep
Generative design of molecules or solids	Latent-space sampling to increase diversity and novelty	High	High	Low	Low	Exploration noise	Retain a justify a epistemic producti randomn
Generative design of	Invalid structure generation from	Low	None	Moderate to high	High	Scientific error	Repair architectur

molecules or solids	decoder instability						apply targe filtering
Bayesian optimization in expensive experiments	Controlled stochasticity in acquisition-function search	High	High	Low	Low	Exploration noise	Maintai becaus removing collapses global sea capacit
Bayesian optimization in expensive experiments	Numerical instability in kernel or hyperparameter scaling	Low	None	High	High	Scientific error	Diagnose remove as implementa level fau
Uncertainty sampling in force-field development	Selection of epistemic-frontier cases for model expansion	High	High	Low	Low	Exploration noise	Report a delibera explorato design
Uncertainty sampling in force-field development	Apparent uncertainty caused by miscalibration or data poisoning	Low	None	High	High	Scientific error	Recalibrat clean da before downstre use
Early-stage hybrid or ambiguous cases	Stochastic variation with mixed exploratory and degradative effects	Mixed	Partial or unclear	Mixed	Context-dependent	Gray-zone case	Adjudica using all f criteria rat than gene denoisin

Applications to Materials AI

The four-criterion framework finds immediate traction when applied to concrete materials-AI workflows, revealing how the same deviation can shift category depending on context while preserving the epistemic boundary between productive randomness and harmful mistake. In active-learning pipelines for alloy discovery, for example, the stochastic selection of candidate compositions via uncertainty sampling constitutes exploration noise when evaluated against the framework. Intentionality is satisfied because the algorithm deliberately injects randomness through acquisition functions; epistemic benefit accrues as the system probes sparsely sampled regions of phase space, often surfacing unexpected intermetallic phases; systematicity remains low because each selection is statistically independent; and correctability is poor—removing the randomness would collapse the loop into pure exploitation and forfeit discovery potential. Lookman *et al.* [8] emphasize adaptive sampling using uncertainties for targeted design implicitly relies on precisely this form of noise, yet the literature rarely distinguishes it from prediction error. By contrast, the same pipeline may generate scientific error when a surrogate model mispredicts formation energies due to training-data bias; here, intentionality is absent, epistemic benefit is zero, systematicity is high, and correction restores fidelity without sacrificing exploration.

A second context arises in generative models for molecular and solid-state design. Gómez-Bombarelli and co-workers' data-driven continuous representation of molecules [7] employs stochastic sampling from a latent space to generate candidate structures. When the temperature parameter or noise injection is deliberately tuned to encourage diversity, the deviation meets all four criteria for exploration noise: it is intentional, yields epistemic benefit by populating underrepresented chemical families, exhibits controlled randomness rather than systematic drift, and cannot be corrected without inducing mode collapse and loss of novelty. The framework, therefore, licenses the retention of this stochasticity. Conversely, when the generator produces invalid structures due to decoder instability or overlooked symmetry constraints—an unintentional artifact—the deviation is recorded as scientific error, requiring post-hoc filtering or architectural repair that leaves the exploratory core untouched.

Bayesian optimization routines provide a third illustrative domain. As Diwale and colleagues demonstrate in material discovery processes with noise [18], the algorithm's acquisition function balances exploration and exploitation by injecting controlled stochasticity into the selection of the next experiments. Under the framework, this injected variation qualifies unambiguously as exploration noise: it is intentional by design, delivers epistemic benefit through global search in expensive black-box landscapes, remains statistically random rather than correlated with specific inputs, and resists correction lest the optimizer degenerate into local greedy search. Implementation bugs, however—such as numerical instability in the Gaussian-process kernel or incorrect hyperparameter scaling—introduce scientific error that satisfies the opposite profile and must be isolated and removed. Kusne and McDannald's on-the-fly closed-loop materials discovery via Bayesian active learning [22] further illustrates how the framework separates these layers: the deliberate noise schedule accelerates discovery of new battery electrolytes, while any systematic offset in the surrogate model constitutes error that, if misclassified, would silently degrade the entire campaign.

Uncertainty-sampling strategies form the fourth context. Thomas-Mitchell and co-workers calibrate uncertainty in machine-learning force fields for active learning [11]. Yet, the framework clarifies that selecting high-uncertainty points is exploration noise only when the uncertainty arises from deliberate probing of the model's epistemic frontier. Here, intentionality, epistemic benefit, low systematicity, and poor correctability all align in favor of preservation. When uncertainty instead stems from model miscalibration or data poisoning, the deviation flips to scientific error and triggers targeted recalibration. Across these four materials, AI contexts the framework thus functions as a diagnostic lens rather than a rigid rule, enabling practitioners to retain productive randomness while excising corrosive error. The noise-error spectrum can be conceptualized as a linear axis running from pure exploration noise (high intentionality, high epistemic benefit, low systematicity, low correctability) at one pole to pure scientific error (low intentionality, zero epistemic benefit, high systematicity, high correctability) at the other; intermediate points represent hybrid cases—such as stochasticity in early-stage generative screening—that require case-by-case adjudication but never collapse the two endpoints into a single category. This spectrum supplies a visual and conceptual scaffold for workflow design, ensuring that materials AI pipelines treat stochasticity as a tunable resource rather than a uniform liability [24-29].

Objections and Replies

Three principal objections surface when the proposed distinction is introduced into the materials-AI community; each can be met with a direct conceptual reply that preserves the framework's utility.

Objection 1 maintains that noise is always harmful and should be universally minimized because randomness degrades predictive accuracy in data-scarce domains. The reply is that this objection conflates statistical variance with epistemic value: exploration noise is precisely the controlled randomness that, while increasing short-term variance, yields long-term knowledge gains unavailable through deterministic exploitation alone. Sutton and Barto's foundational treatment of exploration in reinforcement learning [2] already establishes that zero-noise regimes converge suboptimally; the present framework merely extends that insight to materials-specific autonomous discovery, demonstrating that deliberate noise is not a cost to be minimized but an investment whose return is measured in discovered functionalities rather than in immediate RMSE.

Objection 2 asserts that the four criteria are too subjective to apply reliably in practice, rendering the distinction operationally useless. The reply acknowledges the graded nature of the gray zone yet insists that the criteria remain intersubjectively verifiable: intentionality can be inspected in algorithm pseudocode, epistemic benefit can be justified through literature precedent on serendipitous discovery, systematicity can be quantified via statistical tests on deviation distributions, and correctability can be assessed by ablation studies that isolate exploratory components. Wang and co-workers' benchmarking of active-learning strategies [16] already implicitly performs analogous decompositions; the framework simply makes the evaluative step explicit and reproducible rather than tacit.

Objection 3 dismisses the entire exercise as mere semantics that changes nothing about underlying algorithms or data pipelines. The reply counters that vocabulary is never neutral in engineering practice: the prevailing conflation directly shapes loss functions, regularization choices, and peer-review standards, systematically penalizing productive randomness while tolerating hidden systematic flaws. By supplying a differentiated lexicon, the framework alters design incentives—authors will now report exploration-noise schedules as design features rather than defects, reviewers will demand justification for noise retention, and benchmark suites will begin to reward epistemic yield alongside accuracy—thereby shifting the entire research equilibrium toward more effective autonomous discovery.

Implications for Materials AI Practice

Adoption of the distinction and framework carries concrete implications for authors, reviewers, and the broader community. For authors the changed practice is threefold: first, every manuscript must explicitly distinguish exploration noise from scientific error in method sections, citing the relevant criterion or criteria that justify retention of stochastic elements; second, exploration-noise levels must be reported quantitatively (for instance, temperature schedules or acquisition-function hyperparameters) alongside conventional error metrics; third, evaluation protocols must include an epistemic-benefit analysis that quantifies how retained noise contributed to novel candidate identification rather than merely reporting final model accuracy. These steps transform stochasticity from an unexamined artifact into a deliberate, auditable design choice.

For reviewers, the implications are equally prescriptive: reviewers should question any blanket call for "noise reduction" by asking whether the suppressed component satisfies the framework's exploration criteria, and they should request explicit mapping of reported deviations onto the noise-error spectrum rather than accepting undifferentiated error tables. This interrogation prevents the reflexive penalization of productive randomness that currently pervades peer review.

For the community at large, the framework necessitates two structural developments: first, the establishment of standardized reporting templates that include a dedicated "exploration-noise justification" subsection; second, the creation of new benchmark suites that evaluate not only predictive accuracy but also exploratory yield under controlled noise budgets. Taken together, these practice-level changes convert the conceptual distinction into an operational norm, ensuring that materials AI pipelines are designed to harness rather than eradicate the stochastic mechanisms that have historically driven breakthrough discoveries.

Conclusion

The conceptual distinction between exploration noise and scientific error provides AI with a sharper epistemic vocabulary and a practical decision-making framework that resolves a long-standing terminological ambiguity. Exploration noise—stochastic variation introduced to probe uncertain regions with potential epistemic benefit—must be deliberately preserved, while scientific error—any deviation that reduces fidelity without compensatory insight—must be corrected. By moving from conflation through precise definitions, neighboring-term distinctions, and a four-criterion framework to targeted applications and practice-level implications, the analysis demonstrates that treating all randomness as error simultaneously starves discovery and masks systematic flaws. The field is therefore called to value exploration noise as a productive epistemic resource rather than minimize it as an engineering defect, thereby unlocking the full potential of autonomous materials discovery in the decades ahead.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 18 Apr 2025 Revised: 09 Jul 2025 Accepted: 28 Aug 2025
Published online: 18 January 2026

References

Noack MM, Doerk GS, Li R, Streit JK, Vaia RA, Yager KG, et al. Autonomous materials discovery driven by Gaussian process regression with inhomogeneous measurement noise and anisotropic kernels. Sci Rep. 2020;10(1):17663.

Sutton R, Barto A. Reinforcement learning: An introduction. Cambridge (MA): MIT Press; 2018. p. 31-329.

Plappert M, Houthooft R, Dhariwal P, Sidor S, Chen RY, Chen X, et al. Parameter space noise for exploration. arXiv preprint arXiv:1706.01905. 2017 Jun 6.

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.

Zunger A. Inverse design in search of materials with target functionalities. Nat Rev Chem. 2018;2(4):0121.

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.

Lookman T, Balachandran PV, Xue D, Yuan R. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. npj Comput Mater. 2019;5(1):21.

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

Cesar de Azevedo L, Pinheiro GA, Quiles MG, Da Silva JL, Prati RC. Systematic investigation of error distribution in machine learning algorithms applied to the quantum-chemistry QM9 data set using the bias and variance decomposition. J Chem Inf Model. 2021;61(9):4210-23.

Thomas-Mitchell A, Hawe G, Popelier PL. Calibration of uncertainty in the active learning of machine learning force fields. Mach Learn Sci Technol. 2023;4(4):045034.

Lookman T, Liu Y, Gao Z. Materials informatics: Emergence to autonomous discovery in the age of AI. Adv Mater. 2026:e15941.

Dolar T, della Ventura NM, Mignerot F, Wang Z, Howard HC, Kassner CT, et al. Accelerating materials discovery in heterogeneous composition-property design spaces via collaborative Bayesian optimization. Mater Des. 2025:115331.

Xu L, Hoffman N, Wang Z, Xu H. Harnessing structural stochasticity in the computational discovery and design of microstructures. Mater Des. 2022;223:111223.

Jiang D, Cao C, Zhang J, Zhang O, Zhu C, Xu J, et al. Stochasticity-by-design in memristive materials for probabilistic computing and hardware security: Entropy sources, control knobs, and benchmarks. ACS Appl Electron Mater. 2026.

Wang A, Liang H, McDannald A, Takeuchi I, Kusne AG. Benchmarking active learning strategies for materials optimization and discovery. Oxf Open Mater Sci. 2022;2(1):itac006.

Stach E, DeCost B, Kusne AG, Hattrick-Simpers J, Brown KA, Reyes KG, et al. Autonomous experimentation systems for materials development: A community perspective. Matter. 2021;4(9):2702-26.

Diwale S, Eisner MK, Carpenter C, Sun W, Rutledge GC, Braatz RD. Bayesian optimization for material discovery processes with noise. Mol Syst Des Eng. 2022;7(6):622-36.

Wei Y, Peng B, Xie R, Chen Y, Qin Y, Wen P, et al. Deep active optimization for complex systems. Nat Comput Sci. 2025;5(9):801-12.

Pan X, Xie Y, Li C, He Y, Zhang Y, Wang Y, et al. Convergence of computational materials science and AI for next-generation energy storage materials. J Electron Mater. 2026;55(1):45-114.

Raihan AS, Khosravi H, Das S, Ahmed I. Accelerating material discovery with a threshold-driven hybrid acquisition policy-based Bayesian optimization. Manuf Lett. 2024;41:1300-11.

Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via Bayesian active learning. Nat Commun. 2020;11(1):5966.

Johnson NS, Vulimiri PS, To AC, Zhang X, Brice CA, Kappes BB, et al. Invited review: Machine learning for materials developments in metals additive manufacturing. Addit Manuf. 2020;36:101641.

Bi J, Xu Y, Conrad F, Wiemer H, Ihlenfeldt S. A comprehensive benchmark of active learning strategies with AutoML for small-sample regression in materials science. Sci Rep. 2025;15(1):37167.

Ma Y, Gao Y, Wang L, Chen M, Cui W, Wang B, et al. Accelerating materials discovery through active learning: Methods, challenges and opportunities. Innov Info. 2025;1(1):100013.

Li Y, Jiang E, Ni Z, Li W, Huang M, Zhao F, et al. A study of the role of data and model uncertainty in active learning. Comput Mater Sci. 2025;247:113512.

Patel KP, Maicke A, Arzate J, Kwon J, Smith JD, Aimone JB, et al. AI-guided framework for the design of materials and devices for magnetic-tunnel-junction-based true random number generators. Commun Eng. 2025;4(1):43.

Chen Y, Deierling P, Xiao S. Exploring active learning strategies for predictive models in mechanics of materials. Appl Phys A. 2024;130(8):588.

Eldabah NM, Pratap A, Pandey A, Sardana N, Sidhu SS, Gepreel MA. Design approaches of high-entropy alloys using artificial intelligence: A review. Adv Eng Mater. 2025;27(12):2402504.