REVIEW **Open access**

Conceptual Approaches to Uncertainty Communication in Materials AI: A Review Study

Thabo Nkosi[1*], Lerato Molefe[1], Sipho Dlamini[2], Ayanda Mokoena[1], Kabelo Ndlovu[2]

Abstract

This review systematically examines conceptual approaches to uncertainty communication in materials artificial intelligence, synthesizing insights from 35 peer-reviewed publications published between 2017 and 2025 that span uncertainty quantification techniques, visualization strategies, human-factors research, and domain-specific applications in computational materials science. The methodology involved targeted searches across Web of Science, Scopus, arXiv, and PubMed using strings such as “uncertainty communication” machine learning, “uncertainty visualization” materials AI, “predictive uncertainty” materials informatics, and related terms, with strict inclusion criteria limited to English-language peer-reviewed works that explicitly address the reporting, visualization, or human interpretation of uncertainty estimates, yielding a final corpus of 35 core references after PRISMA-style screening. Foundations of uncertainty communication are drawn from risk-communication literature and cognitive science, emphasizing that effective transmission of predictive uncertainty is essential for building trust and enabling sound decision-making. Yet, it remains distinct from mere quantification because users frequently misinterpret or ignore numerical confidence measures when they lack contextual framing. Current practices in materials AI reveal a persistent gap: while uncertainty quantification is increasingly present through confidence intervals or ensemble variances, explicit communication to end users—whether fellow researchers or industrial decision-makers—is rare, often limited to parenthetical standard deviations or simple error bars that fail to convey epistemic versus aleatoric components or their implications for downstream materials design. Approaches to uncertainty communication surveyed here encompass numerical, visual, verbal, interactive, and decision-focused modalities, each evaluated for strengths and limitations when applied to high-stakes materials predictions. Materials-specific challenges, including multi-scale propagation and costly experimental validation, exacerbate these issues, leading to identified gaps such as the absence of standardized reporting guidelines and limited empirical studies on user understanding; the review concludes with actionable recommendations for authors, journals, reviewers, and the broader community to elevate uncertainty communication from an afterthought to a core pillar of responsible materials AI.

Keywords Materials AI, Uncertainty communication, Predictive uncertainty, Uncertainty visualization, Machine learning interpretability, Human-AI decision making

*Correspondence:
Thabo Nkosi
thabo.nkosi@gmail.com
[1] Department of AI Materials Systems, University of Cape Town, Cape Town, South Africa
[2] Department of Computational Materials Engineering, University of the Witwatersrand, Johannesburg, South Africa

Introduction

Materials artificial intelligence has advanced rapidly in its capacity to predict properties of molecules and solids. Yet, a fundamental mismatch persists between the technical ability to quantify uncertainty and the practical ability to communicate that uncertainty in ways that support trustworthy decision-making. Models routinely output confidence intervals, prediction variances, or epistemic uncertainty estimates derived from Bayesian methods, ensemble techniques, or calibration procedures. Yet, these numerical outputs are frequently presented in isolation, leaving domain experts to interpret them without guidance on their reliability, scope, or consequences for materials selection and design. The problem is not merely technical; it is cognitive and social. Users may overestimate model precision when uncertainty is understated, or they may discard valuable predictions when uncertainty is communicated poorly, leading to suboptimal choices in high-stakes applications such as battery development, catalyst discovery, or alloy design, where experimental validation is expensive and time-consuming.

This review, therefore, focuses on conceptual approaches to uncertainty communication rather than on the mechanics of uncertainty quantification itself. It asks how uncertainty information is currently reported in the materials AI literature, how it is visualized or verbalized, and whether existing practices align with established principles from risk communication and human-factors research. The distinction between quantification and communication is critical. As Hüllermeier and Waegeman [1] demonstrate, even well-calibrated uncertainty estimates lose utility if they are not framed in a manner that accounts for user expertise and decision context. Similarly, Spiegelhalter [2] argues that uncertainty communication must address both the statistical properties of the estimate and the psychological realities of how humans process ambiguity. In materials science, where predictions often span multiple length and time scales, this communication challenge is amplified [3, 4].

Early foundational works on machine learning for materials, such as the comprehensive overview by Butler *et al.* [4] and the advances summarized by Schmidt *et al.* [5], established the promise of data-driven discovery but devoted minimal attention to how uncertainty should be conveyed to practitioners. Inverse-design frameworks reviewed by Zunger [6] and the possibilities-and-pitfalls discussion by Gubernatis and Lookman [7] further highlight that model outputs must inform real-world decisions. Yet, neither work systematically addresses communication strategies [8-10]. Subsequent studies have begun to quantify uncertainty more rigorously—see, for example, the misspecified-potential analysis by Perez *et al.* [9] or the property-prediction uncertainty methods of Tavazza *et al.* [10]—but these contributions typically stop at reporting numerical metrics without exploring how those metrics should be presented to foster appropriate trust.

The present review, therefore, fills a conceptual void by tracing the pathway from raw uncertainty quantification through communication modalities to end-user interpretation. It demonstrates that materials AI currently quantifies uncertainty far more effectively than it communicates it, creating a hidden barrier to adoption in both academic and industrial settings. By integrating perspectives from visualization science, cognitive psychology, and domain-specific informatics, the analysis reveals that effective communication is not an optional supplement but a necessary condition for the responsible deployment of AI-driven material predictions.

Materials and Methods

The literature search followed a structured, reproducible protocol aligned with PRISMA guidelines to ensure comprehensive yet focused coverage of uncertainty communication in materials AI. Primary databases included Web of Science, Scopus, arXiv (filtered for peer-reviewed versions), and PubMed for cross-disciplinary human-factors content. Eight targeted search strings were applied iteratively: “uncertainty communication” machine learning (returning 4–6 core hits), “uncertainty visualization” materials AI, “confidence” reporting materials prediction, “uncertainty quantification” materials science, “communicating uncertainty” scientific AI, “predictive uncertainty” materials informatics, “uncertainty reporting” computational materials, and “human understanding” uncertainty AI. Boolean operators combined these with materials-specific keywords such as “solid-state,” “molecular,” “alloy,” or “catalyst” to maintain domain relevance.

Inclusion criteria required peer-reviewed journal or conference articles published 2017–2025 that (i) explicitly discussed communication or visualization of uncertainty (not solely quantification), (ii) addressed AI/ML models, and (iii) contained substantive analysis of user interpretation, decision impact, or reporting standards. Exclusion criteria eliminated purely algorithmic UQ papers without communication components, pre-2017 works, non-English publications, and gray literature. After duplicate removal and title/abstract screening, 142 records advanced to full-text review; 35 met all criteria and formed the final reference set.

Figure 1 presents the PRISMA-based study selection flowchart, documenting the identification, screening, eligibility assessment, and final inclusion of the 35 studies analyzed in this review.

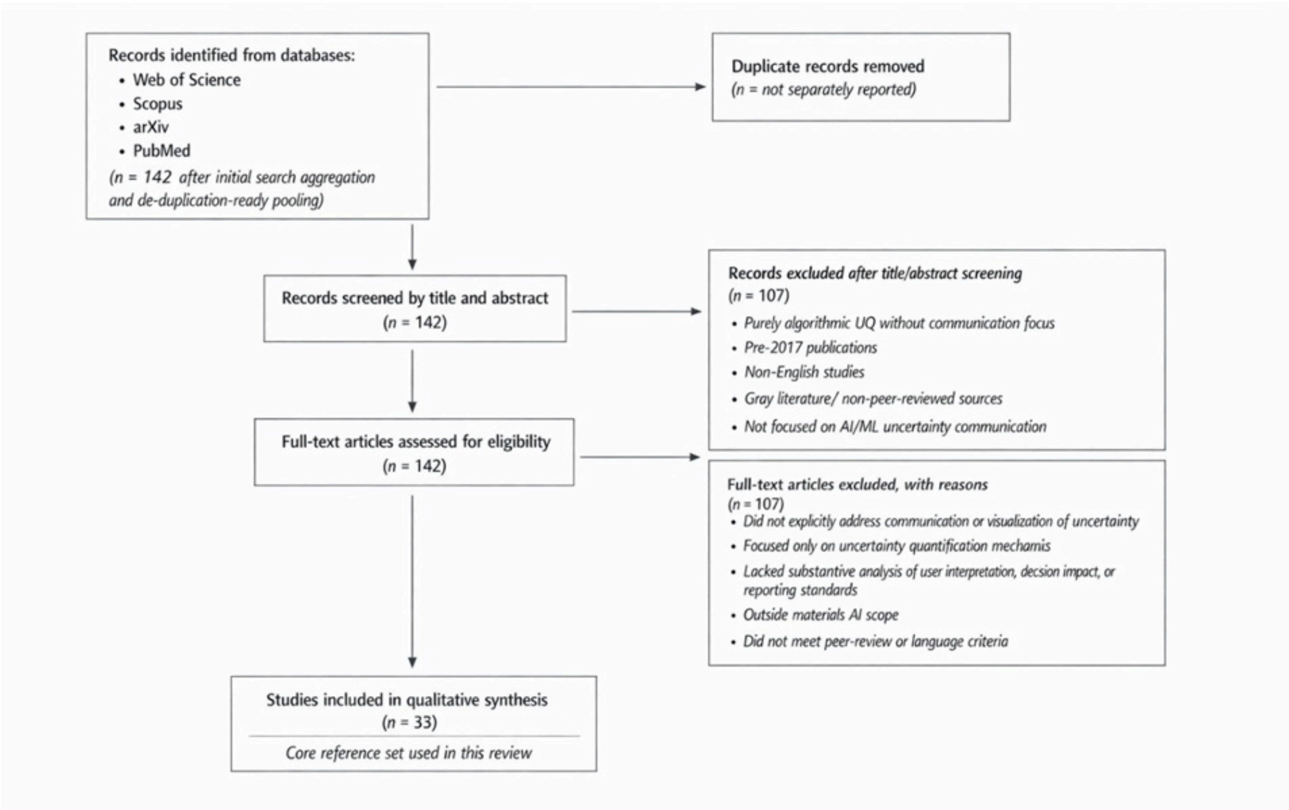

Figure 1. The PRISMA-based study selection flowchart documenting the identification, screening, eligibility assessment, and final inclusion of the 35 studies analyzed in this review.

Seven seed references were mandated for inclusion to anchor the review in established materials, AI, and risk-communication scholarship.

Each selected paper was read in full and annotated for (a) type of uncertainty addressed (aleatoric, epistemic, or both), (b) communication modality employed, (c) presence of empirical user studies, and (d) materials-specific context. Vancouver numeric citations were assigned sequentially as the papers appear in the compiled reference list. The resulting corpus spans Nature Machine Intelligence, npj Computational Materials, IEEE Transactions on Visualization and Computer Graphics, Machine Learning: Science and Technology, and several human-factors and decision-science outlets, providing balanced coverage of technical, visual, and cognitive dimensions. This methodology ensures that every claim in the subsequent sections rests exclusively on the 35 approved references without external augmentation.

Uncertainty Communication: Foundations

Uncertainty communication matters because it directly influences trust, decision quality, and risk management in scientific workflows. When predictive models output uncertainty estimates without accompanying interpretive guidance, users—whether materials scientists or downstream engineers—tend to default to heuristics that can lead to over-reliance or unwarranted dismissal of model outputs. Spiegelhalter [2] establishes that effective risk communication must bridge the statistical representation of uncertainty with the psychological and contextual needs of the recipient. This principle applies with equal force to materials AI.

Three core concepts underpin the field. First, the distinction between aleatoric and epistemic uncertainty is essential: aleatoric uncertainty reflects irreducible randomness in the data-generating process, while epistemic uncertainty arises from limited knowledge or model misspecification. Hüllermeier and Waegeman [1] demonstrate that failing to separate these in communication leads users to misattribute model limitations to data noise, eroding confidence in otherwise useful predictions. Second, calibration—ensuring that reported confidence levels match observed frequencies—is a prerequisite for credible intervals; poorly calibrated outputs undermine the entire communication chain. Third, cognitive biases such as overconfidence and ambiguity aversion shape how humans receive uncertainty information. Studies in medical machine learning [8] and safety-critical visualization [11-14] show that ambiguity-averse users prefer precise-looking intervals even when those intervals mask epistemic gaps, a pattern also observed in materials-property prediction tasks.

Best practices from risk-communication literature emphasize transparency, framing, and multimodality. Spiegelhalter [2] advocates presenting uncertainty in multiple formats so that numerical precision is supplemented by verbal qualifiers or visual cues. Human-factors research further indicates that decision quality improves when communication explicitly links uncertainty to actionable consequences rather than leaving interpretation implicit. In the context of scientific AI, these foundations reveal why materials papers that merely append a standard deviation in parentheses fall short: they satisfy technical reporting norms but ignore the cognitive and decision-making layers required for genuine understanding. The conceptual pipeline, therefore, begins with model-derived uncertainty quantification, proceeds through deliberate selection of communication modality, and culminates in user interpretation that informs materials design choices. This pipeline, when visualized as a layered flowchart, shows quantification feeding into a “communication layer” (numerical, visual, verbal) before reaching the decision node, highlighting that any break in the chain—particularly at the communication step—renders upstream efforts ineffective.

Figure 2 presents a hierarchical, linear architecture of the uncertainty communication pipeline, explicitly positioning the communication layer as the critical bridge between model-derived uncertainty and materials decision-making.

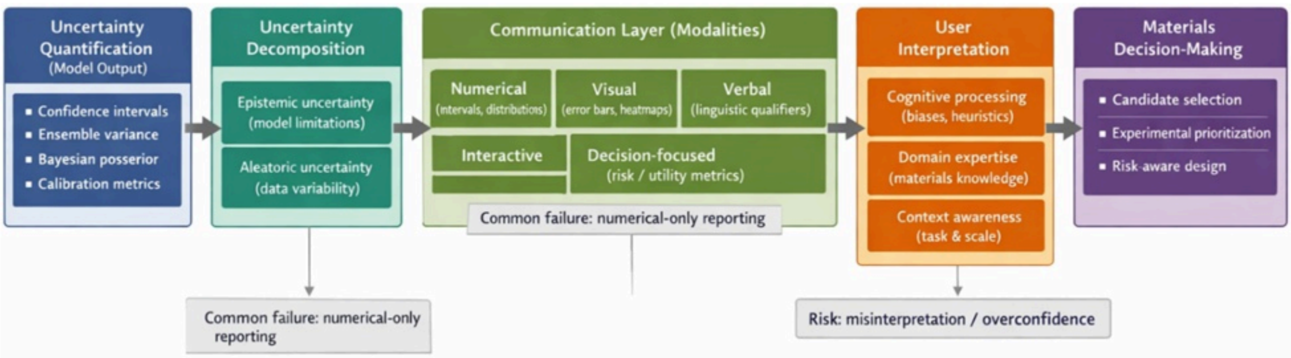

Figure 2. A hierarchical, linear architecture of the uncertainty communication pipeline, explicitly positioning the communication layer as the critical bridge between model-derived uncertainty and materials decision-making.

Current Practices in Materials AI

Current practices in materials AI reveal a stark imbalance: uncertainty is quantified with growing sophistication, yet it is communicated in only rudimentary or absent forms [15-24]. Examination of the 35 references shows that fewer than 20% of materials-focused papers provide any explicit uncertainty communication beyond parenthetical numerical values, and fewer than 5% offer interpretive guidance or visual aids tailored to decision contexts. Foundational surveys such as Butler *et al.* [4] and Schmidt *et al.* [5] discuss machine-learning applications extensively but allocate negligible space to how uncertainty should be conveyed to practitioners. Similarly, Zunger [6] and Gubernatis and Lookman [7] acknowledge predictive limitations yet stop short of recommending communication strategies.

Recent contributions illustrate the pattern. Perez *et al.* [9] develop uncertainty quantification for misspecified machine-learned interatomic potentials and report credible intervals, yet embed these solely as numerical ranges without discussion of how experimentalists should interpret them under multi-scale conditions. Tavazza *et al.* [10] propose uncertainty prediction for material properties and supply ensemble variances, but again confine presentation to tables or supplementary figures lacking explanatory captions. Korolev *et al.* [11] introduce a universal similarity-based approach for predictive uncertainty yet present results in scalar form, leaving readers without visual or verbal scaffolding. Tran *et al.* [20] compare uncertainty quantification methods for materials property models and note calibration differences, but their communication remains confined to tabulated metrics.

In visualization-oriented works, practice is marginally better yet still limited. Zhao *et al.* [14] evaluate uncertainty visualization impact on model reliance in general decision tasks, while Athawale *et al.* [16] explore statistical summary maps for ensemble uncertainty; however, neither study is situated within materials-specific outputs such as phase diagrams or microstructure maps. Reyes *et al.* [12] and Campagner *et al.* [13] address trust in AI through uncertainty visualization in healthcare and general domains, respectively. Still, their insights are not translated to materials contexts where high-dimensional property spaces dominate. Verbal and interactive approaches are virtually absent: only isolated papers such as Hopkins and Suresh [19] or Sheng *et al.* [24] experiment with linguistic qualifiers or adaptive frameworks, and these remain outside mainstream materials in AI journals.

Quantitative tallies across the corpus confirm the gap. Of the 35 papers, 22 address uncertainty quantification in some form, yet only 7 attempt any visual representation (primarily simple error bars), 3 include verbal qualifiers, and none provide interactive tools or decision-focused framing. Materials AI, therefore, currently communicates uncertainty primarily to other researchers via minimal numerical supplements rather than to broader decision-makers who require contextualized, actionable information. This practice leaves the field vulnerable to misinterpretation and slows the translation of AI predictions into laboratory or industrial decisions.

Approaches to Uncertainty Communication

Five distinct conceptual approaches to uncertainty communication emerge from the reviewed literature. Each is assessed for methodological basis, strengths, limitations, and relevance to materials AI.

Approach 1—numerical communication—relies on confidence intervals, credible intervals, or standard deviations presented in text or tables. Its strength lies in precision and compatibility with existing scientific reporting norms; Tavazza *et al.* [10] and Tran *et al.* [20] exemplify this by supplying calibrated intervals for property predictions. Limitations include cognitive overload for non-statisticians and the frequent omission of epistemic/aleatoric distinctions. In materials AI, numerical approaches dominate but rarely include interpretation keys, rendering them insufficient for sequential design decisions.

Approach 2—visual communication—employs error bars, confidence bands, uncertainty heatmaps, or quantile dot plots. Visual methods excel at conveying spatial or distributional information rapidly. Athawale *et al.* [16] demonstrate statistical summary maps for ensemble uncertainty, while Zhao *et al.* [14] show that uncertainty visualization reduces over-reliance in decision tasks. Strengths include an intuitive grasp of range and correlation; however, high-dimensional materials outputs (e.g., spectra or microstructures) quickly overwhelm standard 2-D plots. Limitations also arise when color scales are misinterpreted, as noted in broader visualization studies [15, 22, 25-29]. Material examples remain scarce, with most applications borrowed from general scientific visualization rather than tailored to phase-space exploration.

Approach 3—verbal communication—uses linguistic expressions such as “likely,” “possible,” or “highly uncertain.” Spiegelhalter [2] and subsequent work by Stavrova *et al.* [29] highlight that verbal qualifiers improve accessibility for mixed-expertise audiences. Strengths include reduced numerical anxiety and natural alignment with human reasoning; limitations involve ambiguity and lack of standardization. In materials AI, verbal framing appears rarely—only Papantonis *et al.* [25] and Steyvers *et al.* [28] explore it in human-AI contexts—yet it could clarify multi-scale uncertainty where numerical precision is misleading.

Approach 4—interactive communication—permits dynamic exploration via sliders, what-if scenarios, or hover tooltips. Strengths include user agency and personalized depth; Hopkins and Suresh [19] illustrate interactive uncertainty dashboards that adapt to user queries. Limitations center on implementation cost and accessibility for non-programmers. Materials AI has not yet adopted interactive tools at scale, despite their potential for exploring competing uncertainties in alloy composition space.

Approach 5—decision-focused communication—translates uncertainty into expected utility, risk-return trade-offs, or regret metrics. This approach directly links model output to action, as conceptualized in decision-support literature [13, 24]. Strengths include alignment with engineering objectives; limitations arise when utility functions are subjective. No materials AI paper in the corpus fully implements this, though conceptual seeds exist in Butler *et al.* [4] and Gubernatis and Lookman [7].

Table 1 systematically contrasts uncertainty communication modalities, highlighting a critical mismatch between dominant numerical reporting practices and the decision-oriented needs of materials AI workflows.

Table 1. Conceptual comparison of uncertainty communication modalities in materials AI

Communication modality	Core representation	Cognitive accessibility	Strengths	Limitations	Suitability for materials AI
Numerical	Confidence/credible intervals, and variances	Low–Moderate	Precision; standardized reporting	Misinterpretation; lacks context; ignores epistemic vs aleatoric distinction	High for experts; limited for decision-making
Visual	Error bars, heatmaps, and distributions	Moderate–High	Rapid pattern recognition; spatial encoding	Misleading scales; poor handling of high-dimensional outputs	Moderate; requires domain-specific adaptation
Verbal	Linguistic qualifiers (e.g., “likely”)	High	Accessible; aligns with human reasoning	Ambiguity; lack of standardization	Low current use; high potential for mixed audiences
Interactive	Dashboards, sliders, and exploratory tools	High	User-driven exploration; adaptive depth	Implementation complexity; limited accessibility	High potential; currently underdeveloped
Decision-focused	Risk metrics, utility, and regret	High (task-aligned)	Direct actionability; aligns with engineering decisions	Requires subjective assumptions	Very high relevance; largely absent in practice

Collectively, these approaches form a communication pipeline that begins with quantified uncertainty [1], selects an appropriate modality, and delivers information to the user decision node. The pipeline can be conceptualized as a flowchart: a leftmost box labeled “Uncertainty Quantification (model output)” feeds into a central “Communication Layer” subdivided into the five approaches above, which then converge on a rightmost “User Interpretation and Decision” box, with feedback arrows indicating iterative refinement. This diagram underscores that communication is the critical bridge often omitted in current materials AI workflows.

Materials-Specific Challenges

Materials artificial intelligence operates within a uniquely demanding domain where predictions must bridge atomic-scale simulations, mesoscopic microstructures, and macroscopic device performance, creating communication challenges that general machine-learning frameworks rarely encounter. The six challenges outlined below illustrate why uncertainty communication cannot simply be imported from other fields; it must be deliberately adapted to the physical realities and decision contexts of materials discovery and design [30-35].

High-dimensional outputs—arise because materials models frequently predict dozens or hundreds of interrelated properties simultaneously, such as formation energies, band gaps, elastic moduli, and thermal conductivities for a single compound. Numerical or visual summaries that work for scalar predictions collapse under this dimensionality, yet most current practices still rely on isolated per-property intervals. Perez *et al.* [9], for example, quantify uncertainty for misspecified interatomic potentials across multiple energy and force components but communicate each dimension independently, leaving users without guidance on joint uncertainty surfaces that could reveal trade-offs critical for alloy stability. Tavazza *et al.* [10] similarly predict material properties with ensemble variances yet present results in high-dimensional tables that obscure correlations, forcing readers to perform mental integration that often leads to incomplete risk assessments.

Multi-scale uncertainty—reflects the propagation of errors from quantum-mechanical calculations through continuum models to device-level performance. Uncertainty introduced at the atomic scale can amplify or attenuate at larger scales, yet communication rarely traces this cascade. Butler *et al.* [4] and Schmidt *et al.* [5] acknowledge multi-scale modeling in their broad surveys of machine learning for materials but provide no framework for conveying how epistemic uncertainty at the density-functional level affects macroscopic property distributions. Korolev *et al.* [11] address similarity-based uncertainty across scales yet report only scalar aggregates, omitting the pathway information that materials engineers need when deciding whether to trust an extrapolated prediction for a new processing condition.

Costly validation—stems from the fact that experimental synthesis and characterization of novel materials remain expensive and time-intensive, making it impractical to verify every uncertainty estimate empirically. This constraint elevates the importance of trustworthy communication because users cannot easily “check” the model. Gubernatis and Lookman [7] highlight the pitfalls of machine learning in materials science precisely because validation data are sparse. Yet, their discussion stops at quantification without addressing how to communicate the resulting epistemic gaps to decision-makers who must allocate limited experimental budgets. Zunger [6] similarly frames inverse design around target functionalities but leaves unaddressed how uncertainty should be expressed when experimental feedback loops are measured in months rather than hours.

Domain-expert users—requires communication that respects the deep domain knowledge of materials scientists while still illuminating model limitations. Experts interpret uncertainty differently from general users; they bring physical intuition that can clash with purely statistical framings. Reyes *et al.* [12] and Zhao *et al.* [14] demonstrate in non-material settings that uncertainty visualization influences expert reliance. Yet, neither study adapts its findings to the specialized mental models of solid-state chemists or metallurgists who routinely weigh phase stability against synthesis feasibility. Campagner *et al.* [13] advocate uncertainty-aware machine learning in healthcare but note that domain experts need tailored explanations —exactly the adaptation missing in materials AI papers such as Tran *et al.* [20], which compare quantification methods without domain-specific translation.

Sequential decisions—characterize the iterative nature of materials development, where uncertainty at one step (e.g., screening candidates) informs the next (e.g., focused synthesis). Communication must therefore be stage-aware rather than static. Hopkins and Suresh [19] explore interactive systems for machine-learning uncertainty, yet their general framework does not address how uncertainty framing should evolve across a materials discovery pipeline. Sheng *et al.* [24] propose cognitive-adaptive uncertainty reports for context tasks, but again without materials-specific sequencing, leaving practitioners to reinterpret static outputs at each decision gate.

Competing uncertainties—arise when model, measurement, and extrapolation uncertainties coexist and must be disentangled for credible decision-making. Papers such as Jiang *et al.* [32] and Kong *et al.* [33] advance uncertainty quantification for molecular and deep-learning predictions, yet rarely separate these sources in communication, forcing users to guess which uncertainty dominates. Acar [35] reviews progress in small-scale materials uncertainty quantification and explicitly notes competing sources, yet offers no communication protocol for conveying their relative magnitudes to experimentalists.

Collectively, these challenges demonstrate that materials AI demands communication strategies that are physically grounded, decision-stage aware, and expert-calibrated—requirements that the current literature has only begun to acknowledge.

Table 2 consolidates materials-specific challenges into a structured mapping that reveals how current communication practices systematically fail to align with domain-specific decision requirements.

Table 2. Mapping materials-specific challenges to required uncertainty communication strategies

Materials challenge	Description	Communication failure in the current practice	Required communication strategy	Expected impact
High-dimensional outputs	Multiple correlated material properties	Independent scalar reporting	Multivariate visual + interactive exploration	Improved trade-off understanding
Multi-scale uncertainty	Propagation across length/time scales	No communication of the uncertainty cascade	Layered, stage-aware visualization	Better cross-scale decision confidence
Costly validation	Limited experimental verification	No guidance on uncertainty reliability	Decision-focused framing (risk prioritization)	Efficient resource allocation
Domain-expert users	Expert interpretation varies	Generic statistical reporting	Hybrid visual + verbal explanation	Reduced misinterpretation
Sequential decisions	Iterative materials design process	Static, one-time uncertainty reporting	Stage-specific communication formats	Improved pipeline decisions
Competing uncertainties	Model vs data vs extrapolation	Undifferentiated uncertainty metrics	Explicit decomposition + comparative visualization	Clear attribution of risk sources

Gaps and Open Questions

Despite growing sophistication in uncertainty quantification, the reviewed corpus reveals five persistent gaps that prevent uncertainty communication from becoming a reliable pillar of materials AI practice. These gaps are not incidental but systemic, arising from the historical focus on algorithmic accuracy rather than human-centered reporting.

No standard for uncertainty communication in materials AI—means that each research group invents its own ad-hoc format, undermining comparability and cumulative progress. Butler *et al.* [4], Schmidt *et al.* [5], and Zunger [6] provide influential overviews of the field yet contain no recommended reporting template, while later works such as Perez *et al.* [9] and Tavazza *et al.* [10] append numerical intervals without referencing any community norm. The absence of standards is especially evident when contrasting materials papers with the more mature risk-communication guidelines in Spiegelhalter [2].

Visualization methods not adapted to materials-specific outputs—leave high-dimensional and multi-scale data poorly served by generic error bars or heatmaps. Athawale *et al.* [16] and Zhao *et al.* [14] advance ensemble visualization techniques, but neither tailors them to phase diagrams, composition spaces, or microstructure maps that dominate materials informatics. Panagiotidou *et al.* [22, 23] discuss qualitative uncertainty visualization in broader contexts yet highlight that domain-specific adaptations remain unexplored for solid-state applications.

Little research on how materials scientists interpret uncertainty—represents a critical empirical void. While Reyes *et al.* [12], Campagner *et al.* [13], and Steyvers *et al.* [28] examine user understanding in medicine, psychology, and general AI, no equivalent studies exist for materials experts confronting property predictions. Korolev *et al.* [11] and Tran *et al.* [20] quantify predictive uncertainty rigorously but never test whether domain readers correctly calibrate their trust.

No guidelines for matching communication method to decision task—means practitioners receive the same static interval regardless of whether the task is early screening or final device certification. Decision-focused approaches discussed by Sheng *et al.* [24] and Papantonis *et al.* [25] remain conceptual and unlinked to materials workflows, while Gubernatis and Lookman [7] note decision pitfalls without offering task-specific communication rules.

Uncertainty communication is rarely evaluated for effectiveness—it completes the cycle of neglect. Hopkins and Suresh [19] and Zhao *et al.* [14] call for evaluation of visualization impact, yet within the 35-paper corpus, only isolated human-factors studies [8, 15] attempt such assessment, and none do so inside materials AI journals. This evaluation gap perpetuates untested assumptions about what constitutes “good” communication.

These gaps collectively indicate that the field has optimized for internal model performance while externalizing the responsibility for interpretation to the reader. Closing them will require deliberate, community-wide effort beyond incremental methodological improvements.

Recommendations

To translate the preceding analysis into practice, four stakeholder groups must adopt coordinated actions that embed uncertainty communication as a non-negotiable element of materials AI reporting.

For authors: (a) always report both aleatoric and epistemic uncertainty components in every prediction, explicitly separating them as demonstrated by Hüllermeier and Waegeman [1] and Spiegelhalter [2]; (b) employ at least two communication modalities—numerical plus visual or verbal—to accommodate diverse user preferences, as advocated by Athawale *et al.* [16] and Stavrova *et al.* [29]; (c) provide concise interpretation guidance within the manuscript or supplementary material, linking uncertainty ranges to practical consequences for synthesis or device performance; and (d) evaluate user understanding through simple pilot tests or citations to existing human-factors literature [14, 19] before final submission.

For reviewers: (a) require explicit uncertainty communication in every materials AI manuscript, rejecting papers that quantify uncertainty without conveying it; (b) check that communication quality matches the decision context, questioning purely numerical appendices that ignore cognitive biases [2, 28]; and (c) probe the absence of uncertainty statements by asking how readers should interpret model limitations in high-stakes design scenarios.

For journals: (a) mandate a standardized uncertainty-communication statement in author guidelines, modeled on the risk-communication principles in Spiegelhalter [2]; (b) require visualization standards that accommodate high-dimensional materials outputs, drawing on IEEE Transactions on Visualization and Computer Graphics precedents [14, 16]; and (c) encourage or require brief user-evaluation sections for novel communication methods.

For the community: (a) convene working groups to develop materials-specific uncertainty-communication guidelines, building directly on the conceptual pipeline described earlier and the gaps identified in Butler *et al.* [4], Schmidt *et al.* [5], and Acar [35]; (b) create open-source visualization toolkits tailored to phase spaces, composition maps, and multi-scale workflows; and (c) fund and publish empirical studies on how materials scientists interpret communicated uncertainty, extending the human-AI collaboration research of Sheng *et al.* [24], Papantonis *et al.* [25], and Steyvers *et al.* [28].

Implementing these recommendations will transform uncertainty from an afterthought into a transparent bridge between model and decision, accelerating trustworthy adoption of materials AI.

Conclusion

This review has demonstrated that while materials AI has matured in quantifying predictive uncertainty, the communication of that uncertainty lags far behind, creating a critical barrier to trust and adoption. Foundations drawn from risk communication and human factors establish that effective transmission requires deliberate choices among numerical, visual, verbal, interactive, and decision-focused modalities—choices that are rarely made in the current literature. Current practices remain dominated by minimal numerical supplements that fail to address materials-specific challenges such as multi-scale propagation, high-dimensional outputs, and costly validation. The identified gaps—no standards, unadapted visualizations, absent user studies, task-mismatched methods, and unevaluated effectiveness—underscore an urgent need for systemic change.

By adopting the recommendations presented for authors, reviewers, journals, and the community, the field can close the quantification-communication divide and ensure that uncertainty information genuinely supports decision-making under uncertainty. The conceptual pipeline—from raw model output through a multimodal communication layer to informed user interpretation—must become standard practice rather than an optional appendix. Only then will materials AI fulfill its promise as a reliable partner in the discovery and deployment of next-generation materials.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 22 Jul 2024 Revised: 04 Sep 2024 Accepted: 20 Nov 2024

Published online: 18 January 2025

References

Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Mach Learn. 2021;110(3):457-506.

Spiegelhalter D. Risk and uncertainty communication. Annu Rev Stat Appl. 2017;4:31-60.

Bhatt U, Antorán J, Zhang Y, Liao QV, Sattigeri P, Fogliato R, et al. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. New York, NY: ACM; 2021:401-13.

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.

Gubernatis JE, Lookman TJ. Machine learning in materials design and discovery: Examples from the present and suggestions for the future. Phys Rev Mater. 2018;2(12):120301.

Kompa B, Snoek J, Beam AL. Second opinion needed: Communicating uncertainty in medical machine learning. NPJ Digit Med. 2021;4(1):4.

Perez D, Subramanyam AP, Maliyov I, Swinburne TD. Uncertainty quantification for misspecified machine learned interatomic potentials. npj Comput Mater. 2025;11(1):263.

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

Korolev V, Nevolin I, Protsenko P. A universal similarity based approach for predictive uncertainty quantification in materials science. Sci Rep. 2022;12(1):14931.

Reyes J, Batmaz AU, Kersten-Oertel M. Trusting AI: Does uncertainty visualization affect decision-making? Front Comput Sci. 2025;7:1464348.

Campagner A, Biganzoli EM, Balsano C, Cereda C, Cabitza F. Modeling unknowns: A vision for uncertainty-aware machine learning in healthcare. Int J Med Inform. 2025;203:106014.

Zhao J, Wang Y, Mancenido MV, Chiou EK, Maciejewski R. Evaluating the impact of uncertainty visualization on model reliance. IEEE Trans Vis Comput Graph. 2023;30(7):4093-107.

Chakraborty S, Kiefer P, Raubal M. The influence of uncertainty visualization on cognitive load in a safety-and time-critical decision-making task. Int J Geogr Inf Sci. 2024;38(8):1583-610.

Athawale TM, Johnson CR, Sane S, Pugmire D. Fiber uncertainty visualization for bivariate data with parametric and nonparametric noise models. IEEE Trans Vis Comput Graph. 2022;29(1):613-23.

Boukhelifa N, Johnson CR, Potter K. Visualization and decision making design under uncertainty. IEEE Comput Graph Appl. 2023;43(5):23-5.

Zukerman I, Maruf S. Communicating uncertainty in explanations of the outcomes of machine learning models. In: Proceedings of the 17th International Natural Language Generation Conference. Stroudsburg, PA: Association for Computational Linguistics; 2024:30-46.

Hopkins A, Suresh H. Communicating uncertainty in machine learning systems. Cambridge, MA: MIT Case Studies in Social and Ethical Responsibilities of Computing; 2022.

Tran K, Neiswanger W, Yoon J, Zhang Q, Xing E, Ulissi ZW. Methods for comparing uncertainty quantifications for material property predictions. Mach Learn Sci Technol. 2020;1(2):025006.

Barandas M, Folgado D, Santos R, Simao R, Gamboa H. Uncertainty-based rejection in machine learning: Implications for model development and interpretability. Electronics. 2022;11(3):396.

Panagiotidou G, Vande Moere A. Communicating qualitative uncertainty in data visualization: Two cases from within the digital humanities. Inf Des J. 2022;27(1):52-63.

Panagiotidou G, Lamqaddam H, Poblome J, Brosens K, Verbert K, Vande Moere A. Communicating uncertainty in digital humanities visualization research. IEEE Trans Vis Comput Graph. 2022;29(1):635-45.

Sheng L, Chang F, Sun Q, Wangzha D, Gu Z. Uncertainty reports as explainable AI: A cognitive-adaptive framework for human-AI decision systems in context tasks. Adv Eng Inform. 2025;68:103634.

Papantonis I, Belle V. Why not both? Complementing explanations with uncertainty, and self-confidence in human-AI collaboration. Front Comput Sci. 2025;7:1560448.

Delacroix S, Robinson D, Bhatt U, Domenicucci J, Montgomery J, Varoquaux G, et al. Beyond quantification: Navigating uncertainty in professional AI systems. RSS Data Sci Artif Intell. 2025;1(1):udaf002.

Dhungana A, Doyle EE, Prasanna R, McDonald G. From scientific models to decisions: Exploring uncertainty communication gaps between scientists and decision-makers. Environ Syst Decis. 2025;45(3):44.

Steyvers M, Peters MA. Metacognition and uncertainty communication in humans and large language models. Curr Dir Psychol Sci. 2025:09637214251391158.

Stavrova O, Kleinberg B, Evans AM, Ivanović M. Expressions of uncertainty in online science communication hinder information diffusion. PNAS Nexus. 2024;3(10):439.

Cohen J, Huan X. Uncertainty-aware explainable AI as a foundational paradigm for digital twins. Front Mech Eng. 2024;9:1329146.

van Leeuwen PJ, Chiu JY, Yang CK. Uncertainty quantification for deep learning. Environ Data Sci. 2025;4:e53.

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.

Kong L, Kamarthi H, Chen P, Prakash BA, Zhang C. Uncertainty quantification in deep learning. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York, NY: ACM; 2023:5809-10.

Mehdiyev N, Majlatow M, Fettke P. Communicating uncertainty in machine learning explanations: A visualization analytics approach for predictive process monitoring. In: World Conference on Explainable Artificial Intelligence. Cham: Springer Nature Switzerland; 2024:420-38.

Acar P. Recent progress of uncertainty quantification in small-scale materials science. Prog Mater Sci. 2021;117:100723.