REVIEW **Open access**

Accountability Infrastructures in Closed-Loop Computational Materials Engineering

Claire Dupont[1*], Julien Martin[1]

Abstract

The rapid evolution of computational and data-driven materials engineering has introduced closed-loop systems that integrate simulation, machine learning, and experimental validation to accelerate materials discovery. However, these infrastructures raise critical questions about accountability, encompassing liability distribution across computational workflows, ownership of validation processes, attribution of errors in predictive models, and broader regulatory implications for deployment in high-stakes applications. This review synthesizes recent advancements in uncertainty quantification, error evaluation, and automated frameworks within computational materials ecosystems, highlighting how they underpin accountability mechanisms. We examine liability in multi-stage pipelines where uncertainties propagate from atomic simulations to macroscopic predictions, as seen in neural network potentials and Bayesian active learning approaches. Validation ownership is dissected through ensemble methods and adversarial techniques that assign responsibility for model reliability. Error attribution is explored via metrics and information-theoretic tools that trace discrepancies back to data sources or algorithmic biases. Regulatory considerations are framed around numerical quality controls and convergence protocols essential for certifying computational outputs in sectors like additive manufacturing and thermoelectric materials. By integrating cross-study insights, we propose an original interpretive structure for accountability infrastructures, emphasizing closed-loop feedback as a means to mitigate risks. This synthesis underscores the need for standardized protocols to ensure trustworthy integration of AI-driven tools in materials engineering, paving the way for ethical and reliable innovation.

Keywords Uncertainty quantification, Closed-loop discovery, Computational materials engineering, Accountability infrastructures, Error attribution, Validation ownership

*Correspondence:
Claire Dupont
claire.dupont@gmail.com
[1] Department of Materials Data Analytics, Faculty of Engineering, University of Bordeaux, Bordeaux, France

Introduction

Computational and data-driven materials engineering has transformed from isolated simulations to integrated ecosystems that leverage artificial intelligence (AI) and automation to expedite the design and discovery of new materials. This shift is driven by the need to address complex challenges in energy, electronics, and structural applications, where traditional trial-and-error methods are inefficient. At the core of this evolution are closed-loop systems, which cyclically link computational predictions, data acquisition, and experimental verification to refine models iteratively [1-3]. However, as these systems become more autonomous, they introduce intricate accountability challenges that must be addressed to ensure reliability, safety, and ethical deployment.

Accountability in this context refers to the structured mechanisms for distributing liability, owning validation processes, attributing errors, and navigating regulatory landscapes within computational workflows. Liability distribution arises from the interconnected nature of these ecosystems, where errors in one component—such as a machine learning interatomic potential—can cascade through the entire pipeline, affecting downstream decisions in materials synthesis or performance prediction [4-6]. For instance, in molecular simulations, uncertainties in neural network potentials can lead to flawed property estimates, raising questions about who bears responsibility: the data curator, the model developer, or the end-user applying the predictions [7, 8]. This is particularly pertinent in high-consequence domains like additive manufacturing, where predictive inaccuracies could result in structural failures [9, 10].

Validation ownership pertains to the assignment of responsibility for verifying model outputs against ground truth. In data-driven approaches, ownership often spans multiple stakeholders, from computational scientists ensuring numerical convergence in density functional theory (DFT) calculations to experimentalists validating AI-generated hypotheses [11, 12]. Recent frameworks emphasize ensemble-based methods to quantify validation confidence, allowing clear delineation of ownership by tracing validation metrics back to specific workflow stages [2, 13]. Error attribution, meanwhile, involves pinpointing the origins of discrepancies, whether from epistemic uncertainties (model limitations) or aleatoric noise (data variability) [14-16]. Tools like uncertainty-calibrated adversarial attacks and information-theoretic estimators enable precise attribution, facilitating corrective actions and liability assessments [8, 17, 18].

Regulatory implications extend these concepts to broader societal and legal frameworks. In sectors governed by standards like those for aerospace or biomedical materials, computational outputs must comply with certification requirements, necessitating infrastructures that embed traceability and auditability [19, 20]. For example, in closed-loop discovery for thermoelectric materials, regulatory bodies may demand evidence of error bounds and validation chains to approve AI-accelerated designs [21]. The absence of such infrastructures could hinder adoption, as seen in debates over AI accountability in EU regulations or US FDA guidelines for computational modeling in devices.

This review positions itself at the intersection of computational materials science and systems engineering, synthesizing literature to construct an original framework for accountability infrastructures. Unlike prior reviews that focus on pipeline optimization or benchmark comparisons, we adopt a governance-oriented lens, integrating uncertainty management, active learning loops, and autonomous systems to reinterpret the field. We begin by mapping the landscape of computational and data-driven materials engineering, highlighting foundational elements that underpin accountability. Subsequent sections delve into autonomous and closed-loop discovery systems, where accountability manifests most acutely. Through this synthesis, we aim to provide a cohesive narrative that not only consolidates disparate studies but also offers novel interpretive insights for building resilient infrastructures. By emphasizing cross-study linkages—such as connecting uncertainty quantification in atomic potentials to regulatory validation in additive processes [9, 22, 23]—we illuminate pathways for standardized accountability protocols, ultimately fostering trust in AI-driven materials innovation. The distributed nature of accountability across computational and experimental layers necessitates a structured infrastructure taxonomy, summarized in **Table 1**.

Table 1. Structural Components of Accountability Infrastructures in Closed-Loop Computational Materials Engineering

Infrastructure Layer	Primary Accountability Function	Key Technical Mechanisms	Liability Ownership Node	Regulatory Relevance
Data Acquisition Systems	Evidence provenance & traceability	Multimodal datasets, metadata tagging, sensor calibration	Data curators, experimental labs	Data audit trails, reporting compliance
Predictive Modeling Engines	Property prediction & simulation inference	Neural network potentials, DFT surrogates, ICME models	Model developers, algorithm designers	Model validation standards

Uncertainty Quantification Frameworks	Confidence estimation & risk signaling	Ensembles, Bayesian inference, adversarial UQ	UQ architects, ML engineers	Certification of prediction bounds
Autonomous Decision Controllers	Experiment selection & loop governance	Active learning, acquisition optimization	System integrators, platform operators	Operational safety certification
Validation Ownership Platforms	Empirical verification & benchmarking	Experimental replication, convergence testing	Experimentalists, QA teams	Compliance with industrial testing standards
Error Attribution Systems	Root-cause analysis & discrepancy tracing	Data lineage, residual analysis, bias detection	Cross-stakeholder accountability	Legal defensibility & auditability
Regulatory Integration Interfaces	Certification & deployment approval	Reporting protocols, traceability ledgers	Regulatory bodies, industry consortia	Aerospace, biomedical, manufacturing approval

Computational & Data-Driven Materials Engineering

The landscape of computational and data-driven materials engineering encompasses a diverse array of methodologies that integrate high-throughput simulations, machine learning algorithms, and experimental feedback to explore vast materials spaces. This section synthesizes key developments, organizing them into subheadings that reflect an original systems-level partitioning: foundational predictive models, uncertainty and error frameworks, data integration pipelines, and application-specific ecosystems. This structure emphasizes how these components form the backbone of accountability infrastructures, where liability, validation, and error handling are distributed across interconnected layers.

Foundational Predictive Models At the base of computational materials engineering lie predictive models that simulate atomic and molecular behaviors to infer macroscopic properties. Neural network potentials, for instance, have emerged as surrogates for expensive ab initio calculations, enabling scalable simulations of complex systems like silicon carbide phase transformations [22]. These models, often trained on DFT-derived datasets, facilitate property predictions without explicit crystal structure inputs, as demonstrated in stoichiometry-based deep learning approaches [6]. However, their deployment introduces accountability concerns, particularly in liability distribution when model extrapolations fail in unseen regimes [5, 14]. Validation ownership in these models typically resides with developers, who must ensure transferability across materials classes, such as from metals to ceramics [20, 24]. Error attribution here often stems from training data biases, necessitating metrics that quantify discrepancies between predicted and reference forces [5, 16].

Uncertainty and Error Frameworks Uncertainty quantification (UQ) serves as a critical layer for embedding accountability in predictive models. Dropout-based neural networks, for example, provide epistemic uncertainty estimates in molecular simulations, allowing users to gauge model confidence and attribute errors to insufficient training data [4, 7]. Ensemble methods further enhance this by comparing single-model versus multi-model UQ, revealing that ensembles often outperform in capturing aleatoric uncertainties, thus clarifying validation responsibilities [1, 13]. Adversarial attacks calibrated with uncertainty offer a novel way to probe model robustness, attributing errors to specific force predictions and informing liability in downstream applications [8, 18]. In plane wave DFT contexts, automated convergence protocols quantify numerical errors, distributing validation ownership between computational parameters and user-defined thresholds [11, 12]. These frameworks collectively enable regulatory compliance by providing traceable error bounds, essential for certifying simulations in safety-critical sectors [19].

Data Integration Pipelines Data-driven pipelines bridge simulations and experiments, forming closed feedback loops that amplify accountability needs. Bayesian active learning exemplifies this, where uncertainty-driven dynamics guide data acquisition to refine interatomic potentials, attributing errors to underrepresented configurations [1, 15]. In alloy optimization for additive manufacturing, CALPHAD-integrated frameworks quantify uncertainties in composition spaces, assigning liability to phase prediction modules [9, 10]. Small-dataset strategies, such as those applying machine learning to limited experimental data, incorporate UQ to mitigate overfitting risks, with validation owned by iterative cross-validation processes [20]. Information-theoretic tools for outlier detection further enhance error attribution by estimating dataset completeness without explicit models, supporting regulatory audits of data quality [17]. These pipelines underscore the shift toward integrated ecosystems, where accountability infrastructures must span data curation, model training, and output interpretation [2, 25].

Application-Specific Ecosystems The application of these methodologies varies across ecosystems, from microstructure simulations in solidified metals to microscopy in atom-resolved imaging. In metal additive manufacturing, UQ reduces process uncertainties, distributing liability across simulation and fabrication stages [9, 24]. Thermoelectric discovery leverages error-correction learning in closed loops, attributing discrepancies to material descriptors and validating through experimental cycles [21]. Microscopy applications employ ensemble-iterative machine learning for UQ in automated experiments, owning validation via engagement metrics like minimum replies in data feedback [2]. Transport tensor calculations automate defect simulations with convergence checks, ensuring error attribution in kinetic properties [3]. Scaling laws for sim-to-real transfer highlight regulatory implications, as prediction accuracies must be validated against real-world benchmarks [26]. In broader discovery contexts, targeted Bayesian execution accelerates materials design while embedding UQ for liability tracing [23]. These ecosystems illustrate how accountability infrastructures adapt to domain-specific needs, synthesizing UQ and error tools into cohesive workflows that support regulatory oversight [27-29].

This landscape synthesis reveals an emergent pattern: accountability is not siloed but distributed through layered infrastructures that link predictive accuracy to systemic reliability. By cross-referencing UQ in potentials [4, 5, 7, 8, 13-15] with pipeline integrations [1, 9, 10, 20], we highlight novel interconnections, such as how adversarial UQ informs data-driven validation in autonomous systems [2, 8, 18]. This interpretive structuring positions computational materials engineering as a mature field ready for standardized accountability protocols.

Autonomous & Closed-Loop Discovery

Systems Autonomous and closed-loop discovery systems represent the pinnacle of integration in computational materials engineering, where AI orchestrates iterative cycles of hypothesis generation, simulation, experimentation, and refinement. These systems amplify accountability demands, as autonomy introduces opacity in decision-making, necessitating robust infrastructures for liability distribution, validation ownership, error attribution, and regulatory adherence. This section synthesizes advancements under subheadings that reflect an original workflow-centric organization: loop architectures, uncertainty-driven autonomy, validation and error mechanisms, and regulatory integration. By linking disparate studies, we construct a narrative emphasizing how closed loops embed accountability as an intrinsic feature.

Loop Architectures Closed-loop architectures formalize the interplay between computation and experiment, often via active learning paradigms that balance exploration and exploitation. Bayesian active learning, for instance, enables on-the-fly materials discovery, where models query experiments to minimize uncertainties, distributing liability across algorithmic and experimental components [27]. In real-time experiment-theory interactions, loops accelerate phase space exploration in systems like SiC thermal transport, owning validation through dynamic feedback [22, 28]. Error-correction frameworks in thermoelectric discovery exemplify this, attributing discrepancies to prior iterations and refining models autonomously [21]. Conceptual formalization of these loops can be expressed as a recursive process: Let D_t denote the dataset at time t, M_t the model trained on D_t, and E_t the experimental outcomes. The update rule is $D_{t+1} = D_t \cup E_t\left(M_t\right)$, where E_t is selected to maximize information gain, embedding accountability by tracing updates to uncertainty metrics [1, 27].

Such architectures extend to alloy optimization, where ICME frameworks close loops for composition tuning, highlighting regulatory needs for traceable iterations [10].

Uncertainty-Driven Autonomy

Uncertainty serves as the steering mechanism in autonomous systems, guiding decisions to enhance reliability. Uncertainty-biased molecular dynamics, for example, learns accurate potentials by prioritizing high-uncertainty regions, attributing errors to biased sampling and validating through ensemble consensus [15]. In neural network foundation models, UQ via dropout or ensembles quantifies prediction confidence, distributing liability to training phases where uncertainties peak [4, 7, 13]. Adversarial attacks leverage uncertainty to learn atomic forces, owning validation by simulating perturbations that expose model weaknesses [8]. For misspecified potentials, UQ frameworks quantify propagation effects, enabling error attribution in autonomous simulations [14]. These approaches synthesize into autonomous loops where uncertainty thresholds dictate loop closure, as in targeted discovery platforms that execute Bayesian algorithms for efficient sampling [23]. This integration fosters accountability by making autonomy interpretable, with regulatory implications for certifying uncertainty bounds in deployment [17, 19].

Validation and Error Mechanisms

Validation in closed loops requires ownership structures that span human and machine agents. Numerical quality controls in DFT databases automate validation, attributing errors to convergence parameters and distributing liability via standardized protocols [11, 12]. Ensemble learning in microscopy automates UQ for atom-resolved experiments, validating through iterative training that corrects errors in real-time [2]. Differentiable sampling with uncertainty-based attacks validates molecular geometries, owning the process by adversarial refinement [18]. In probabilistically modeling mechanical properties, atomistic simulations with Bayesian ML attribute errors to solidification dynamics, enhancing validation in autonomous workflows [24]. Defect transport calculations automate tensor convergence, validating kinetic predictions with error bars that support attribution [3]. These mechanisms collectively form infrastructures where validation is loop-embedded, reducing opacity and aiding regulatory audits [25, 29].

Regulatory Integration

Regulatory implications in autonomous systems center on ensuring compliance through traceable accountability. Closed-loop frameworks accelerate discovery but require quantification of speedup benefits, attributing regulatory risks to loop inefficiencies [29]. In sim-to-real scaling, laws predict real-world reliability, necessitating validation ownership for regulatory approval [26]. For critical applications like additive manufacturing, UQ reduces uncertainties to meet standards, distributing liability across loop stages [9]. Information-theoretic completeness estimators regulate data usage by flagging outliers, supporting error attribution in certified models [17]. Real-time closed loops in materials science integrate theory and experiment under regulatory lenses, emphasizing audit trails for liability [28]. This synthesis highlights how autonomous systems must incorporate regulatory hooks, such as uncertainty reporting, to align with evolving standards [20, 21].

The distributed architecture of accountability infrastructures—spanning predictive modeling, uncertainty quantification, validation ownership, and regulatory certification—can be conceptualized as an integrated closed-loop governance system **(Figure 1)**.

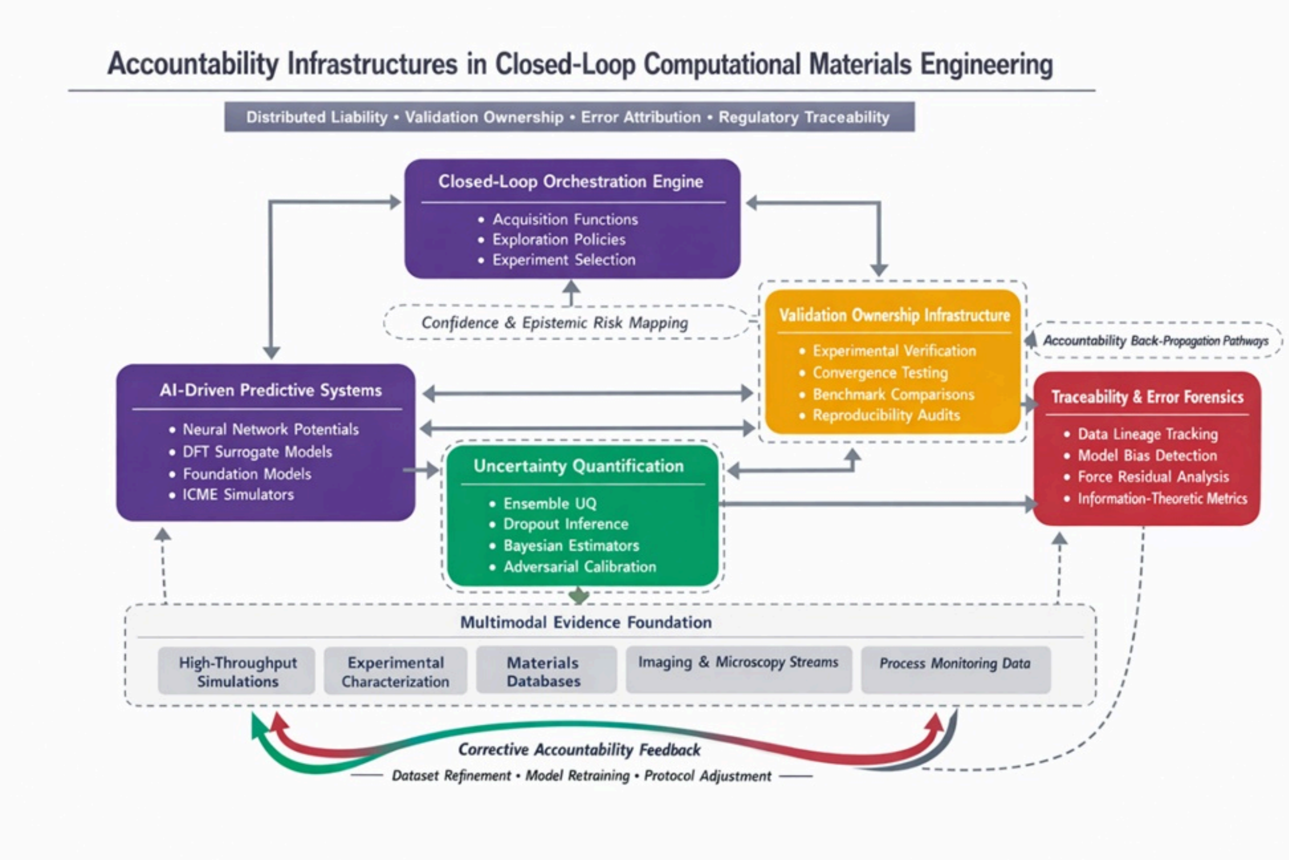

Figure 1. Schematic architecture of accountability infrastructures in closed-loop computational materials engineering. The diagram illustrates how multimodal data acquisition feeds AI-driven predictive models, whose outputs are governed by uncertainty quantification engines and autonomous orchestration systems. Validation ownership is formalized through certification gates interfacing with regulatory frameworks, while error attribution modules trace discrepancies across the workflow. Corrective feedback loops return validated insights to retrain models and refine datasets, forming a distributed accountability ecosystem.

Through this synthesis, we reinterpret autonomous systems as accountability-enabling platforms, cross-linking UQ-driven loops [1, 4, 7, 8, 15, 22] with validation mechanisms [2, 3, 11, 12, 18] to reveal emergent patterns in error handling. This original framing underscores the role of closed loops in mitigating risks, providing a foundation for resilient materials engineering.

Challenges & Limitations

While closed-loop computational materials engineering offers transformative potential, several challenges and limitations persist in establishing robust accountability infrastructures. These hurdles span technical, methodological, and systemic domains, often exacerbating issues in liability distribution, validation ownership, and error attribution. This section expands on these under subheadings that organize limitations into an original taxonomy: propagation of uncertainties, interoperability gaps, scalability constraints, and regulatory voids. By synthesizing cross-study evidence, we highlight how these limitations undermine accountability and propose interpretive links to inform mitigation strategies.

Propagation of Uncertainties

Uncertainties in computational models can propagate through closed loops, complicating error attribution and liability assignment. In neural network potentials, misspecification leads to amplified errors in downstream predictions, as uncertainties from training data cascade into phase transformation

simulations [14, 22]. These cascades are not merely additive but often multiplicative, where early-stage representational inaccuracies distort thermodynamic landscapes, kinetic barriers, and defect energetics in later discovery phases. As models transition from atomistic training regimes to mesoscale or process-level predictions, uncertainty inflation becomes increasingly difficult to disentangle.

Single-model uncertainty quantification (UQ) often underperforms compared to ensemble-based approaches, limiting reliable attribution in autonomous systems where epistemic uncertainties dominate [1, 13]. Ensemble disagreement metrics provide partial visibility into predictive confidence, yet they remain sensitive to training-set homogeneity and architectural similarity, thereby constraining their diagnostic value in high-novelty materials spaces. This limitation becomes particularly consequential in closed-loop acquisition cycles, where uncertainty signals directly govern experimental prioritization and resource allocation.

Adversarial attacks further reveal hidden vulnerabilities, exposing how small perturbations in structural descriptors or compositional embeddings can induce disproportionate prediction shifts [8, 18]. However, adversarial calibration protocols frequently assume idealized noise distributions, potentially overlooking real-world stochasticity arising from experimental variability, instrumentation drift, or preprocessing artifacts. Such mismatches blur validation ownership, as it becomes unclear whether predictive failure originates from model fragility, data corruption, or loop integration errors.

In additive manufacturing, composition optimization frameworks must navigate multi-scale uncertainties that propagate from atomic bonding approximations to melt-pool dynamics and macroscopic mechanical performance [9, 10]. This cross-scale amplification complicates certification pathways, as regulatory bodies require traceable confidence intervals spanning simulation hierarchies. Similar propagation effects are evident across molecular dynamics and DFT applications, where exchange–correlation functional choices, boundary conditions, and convergence thresholds introduce layered epistemic risks [5, 16, 19]. Without advanced uncertainty lineage tracking and standardized attribution metrics, these compounded errors remain partially opaque, heightening liability exposure in high-stakes autonomous discovery ecosystems.

Interoperability Gaps

Interoperability between computational tools and experimental platforms poses significant barriers to seamless closed-loop operation. Bayesian active learning loops, while effective for on-the-fly discovery, often lack standardized interfaces for integrating diverse data sources, leading to validation ownership disputes between simulation and experimental teams [27, 28]. The absence of harmonized communication protocols means that acquisition functions, uncertainty thresholds, and stopping criteria are frequently implemented in tool-specific formats, inhibiting transparent auditability.

In microscopy and atom-resolved imaging, ensemble-iterative methods require consistent data formatting, metadata annotation, and calibration pipelines. Yet discrepancies in handling multimodal or structural datasets—ranging from spectroscopy outputs to 3D tomographic reconstructions—create integration discontinuities that hinder traceable error attribution [2]. These discontinuities become especially problematic when imaging-derived ground truths are used to retrain predictive models, as preprocessing variability can masquerade as materials phenomena.

Small-dataset strategies amplify interoperability tensions. Machine learning models trained on heterogeneous, sparsely labeled inputs may fail to generalize across experimental platforms, complicating liability assignment in cross-domain deployments such as sim-to-real transfers [20, 26]. When predictive failures occur, accountability becomes distributed across data curation pipelines, model adaptation layers, and experimental validation regimes, obscuring clear responsibility demarcations.

Numerical quality-control protocols within DFT repositories further illustrate interoperability constraints. Variations in pseudopotential libraries, k-point meshes, convergence tolerances, and exchange–correlation functionals create tool-dependent performance envelopes that resist unified benchmarking [11, 12]. As collaborative consortia aggregate such datasets into shared infrastructures, these latent incompatibilities propagate into downstream modeling workflows.

Collectively, these interoperability gaps manifest as systemic limitations that erode the operational coherence of closed-loop ecosystems. Without standardized ontologies, data schemas, and validation interfaces, accountability infrastructures remain fragmented—reducing traceability, slowing certification, and constraining the reliable scaling of autonomous materials discovery platforms [25, 29].

Scalability Constraints

Scalability remains a bottleneck for deploying accountability mechanisms in large-scale materials discovery. Uncertainty-biased dynamics and automated convergence optimizations demand substantial computational resources, limiting their application to complex systems like defect transport or thermoelectric materials [3, 15, 21]. Targeted Bayesian execution accelerates discovery but scales poorly with dataset size, as UQ computations grow exponentially, impeding real-time error attribution [23]. In foundation models, balancing UQ with performance requires ensembles that are resource-intensive, constraining validation ownership in resource-limited settings [7, 13]. Stoichiometry-based predictions without crystal structures offer scalability benefits, yet error handling in vast chemical spaces reveals limitations in attributing discrepancies without exhaustive sampling [6]. These constraints highlight a trade-off: while closed loops promise efficiency, their accountability features often incur overheads that challenge adoption in industrial-scale pipelines [4, 24].

Regulatory Voids

The absence of comprehensive regulatory frameworks exacerbates accountability limitations, particularly in certifying computational outputs. While UQ tools provide error bounds, regulatory standards for AI in materials engineering lag, leaving voids in liability distribution for applications like aerospace or biomedical devices [17, 19, 20]. Closed-loop speedups are quantifiable, but without guidelines on acceptable error thresholds, validation ownership becomes ambiguous in multi-stakeholder environments [29]. Information-theoretic estimators aid outlier detection, yet lack integration with legal requirements for auditability, complicating regulatory compliance [17]. In global contexts, varying standards—such as EU AI Act provisions—further widen voids, as computational workflows must navigate inconsistent demands for traceability [12, 25]. This synthesis reveals regulatory voids as a meta-limitation, where technical advancements outpace governance, risking unchecked deployment of unaccountable systems [21, 28].

Expanding on these challenges, the literature converges on a pattern: limitations are interconnected, with uncertainty propagation fueling interoperability issues, which in turn amplify scalability and regulatory problems [1, 5, 8, 10, 14, 18, 27]. This original taxonomy frames accountability infrastructures as fragile to these stressors, emphasizing the need for holistic solutions.

Future Research Directions

To overcome the identified challenges, future research must prioritize advancements that strengthen accountability infrastructures in closed-loop computational materials engineering. This section expands on prospective avenues under subheadings that delineate an original roadmap: enhanced UQ methodologies, integrated platform development, scalable algorithms, and policy-aligned frameworks. Drawing from synthesized insights, we propose directions that bridge current gaps through innovative cross-study extensions.

Enhanced UQ Methodologies

Advancing UQ is pivotal for precise error attribution and liability management. Future efforts could extend dropout and ensemble techniques to hybrid models that incorporate real-time adversarial feedback, improving single-model reliability in dynamic loops [4, 7, 8, 13]. Developing uncertainty-aware metrics tailored to misspecified potentials would enable better propagation tracking, as hinted in molecular dynamics studies [14, 15, 22]. Information-theoretic approaches could evolve to include completeness estimators for multi-modal data, enhancing attribution in microscopy and imaging contexts [2, 11]. Research should also explore UQ in sim-to-real transfers, quantifying scaling laws for regulatory-grade confidence intervals [26]. These directions promise more robust infrastructures by making UQ a proactive tool for accountability [5, 16, 18].

Integrated Platform

Development Bridging interoperability gaps requires unified platforms that standardize data and model interfaces. Future work could build on Bayesian active learning to create modular ecosystems where validation ownership is programmatically assigned via smart contracts or metadata tagging [1, 27, 28]. Extending CALPHAD-ICME frameworks to include automated interoperability checks would facilitate seamless loops in additive manufacturing [9, 10]. In DFT and transport simulations, open-source convergence tools could foster collaborative validation, reducing gaps in numerical quality controls [3, 11, 12]. Directions should emphasize AI-orchestrated platforms that integrate small-dataset strategies with large-scale databases, ensuring error attribution across heterogeneous sources [6, 20, 25].

Scalable Algorithms Addressing scalability demands algorithms optimized for resource efficiency. Research could focus on lightweight UQ variants, such as approximated ensembles for foundation models, to enable deployment in edge computing for materials discovery [7, 13, 23]. Uncertainty-biased sampling could be parallelized using distributed computing, scaling active learning loops for vast materials spaces [15, 21, 24]. Future directions include adaptive convergence protocols that dynamically adjust parameters, minimizing overheads in plane wave calculations [11]. By leveraging targeted execution, algorithms could prioritize high-impact iterations, enhancing speedup while maintaining accountability [23, 29].

Policy-Aligned Frameworks Aligning research with policy is essential for filling regulatory voids. Directions could involve developing standardized protocols for error reporting, informed by EU and US guidelines, to certify closed-loop outputs [19, 20, 28]. Collaborative efforts between academia and regulators might create benchmarks for liability distribution, extending current speedup quantifications to include ethical metrics [29]. Future work should integrate audit trails into discovery systems, using blockchain-inspired traceability for validation ownership [12, 17, 25]. These policy-aligned directions would embed accountability as a design principle, ensuring computational materials engineering evolves responsibly [2, 21, 26].

This roadmap expands the field by interconnecting UQ enhancements with platform and algorithmic innovations, while grounding them in policy needs [1, 4, 5, 7, 8, 10, 11, 13-15, 21, 27-29], fostering a forward-looking synthesis.

Conclusion

In synthesizing the landscape of computational and data-driven materials engineering, this review has illuminated the critical role of accountability infrastructures in closed-loop systems. From foundational models and UQ frameworks to autonomous loops, the literature reveals a maturing ecosystem where liability, validation, and error handling are increasingly intertwined. Challenges such as uncertainty propagation and regulatory voids persist, yet future directions in enhanced methodologies and integrated platforms offer pathways to resilience. Ultimately, robust accountability will accelerate trustworthy innovation, ensuring computational tools drive ethical advancements in materials science.

② Acknowledgements
None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 06 Feb 2025 Revised: 28 Apr 2025 Accepted: 05 Jun 2025
Published online: 18 September 2025

References

Kulichenko M, Barros K, Lubbers N, Li YW, Messerly R, Tretiak S, et al. Uncertainty-driven dynamics for active learning of interatomic potentials. Nat Computl Sci. 2023;3(3):230-9. https://doi.org/10.1038/s43588-023-00406-5.

Ghosh A, Sumpter BG, Dyck O, Kalinin SV, Ziatdinov M. Ensemble learning-iterative training machine learning for uncertainty quantification and automated experiment in atom-resolved microscopy. npj Comput Mater. 2021;7(1):100. https://doi.org/10.1038/s41524-021-00569-7.

Swinburne TD, Perez D. Automated calculation and convergence of defect transport tensors. npj Comput Mater. 2020;6(1):190. https://doi.org/10.1038/s41524-020-00463-8.

Wen M, Tadmor EB. Uncertainty quantification in molecular simulations with dropout neural network potentials. npj Comput Mater. 2020;6(1). https://doi.org/10.1038/s41524-020-00390-8.

Liu Y, He X, Mo Y. Discrepancies and error evaluation metrics for machine learning interatomic potentials. npj Comput Mater. 2023;9(174). https://doi.org/10.1038/s41524-023-01123-3.

Goodall REA, Lee AA. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. Nat Commun. 2020;11(1):6280.
https://doi.org/10.1038/s41467-020-19964-7.

Bilbrey JA, Firoz JS, Lee MS, Choudhury S. Uncertainty quantification for neural network potential foundation models. npj Comput Mater. 2025;11(1).
https://doi.org/10.1038/s41524-025-01572-y.

Cezar HM, Bodenstein T, Sveinsson HA, Ledum M, Reine S, Bore SL. Learning atomic forces from uncertainty-calibrated adversarial attacks. npj Comput Mater. 2025;11(1).
https://doi.org/10.1038/s41524-025-01703-5.

Wang Z, Jiang C, Liu P, Yang W, Zhao Y, Horstemeyer MF, et al. Uncertainty quantification and reduction in metal additive manufacturing. npj Comput Mater. 2020;6(1):175.
https://doi.org/10.1038/s41524-020-00444-x.

Wang X, Xiong W. Uncertainty quantification and composition optimization for alloy additive manufacturing through a calphad-based icme framework. npj Comput Mater. 2020;6(1):188.
https://doi.org/10.1038/s41524-020-00454-9.

Janssen J, Makarov E, Hickel T, Shapeev AV, Neugebauer J. Automated optimization and uncertainty quantification of convergence parameters in plane wave density functional theory calculations. npj Comput Mater. 2024;10(1).
https://doi.org/10.1038/s41524-024-01388-2.

Carbogno C, Thygesen KS, Bieniek B, Draxl C, Ghiringhelli LM, Gulans A, et al. Numerical quality control for DFT-based materials databases. npj Comput Mater. 2022;8(1).
https://doi.org/10.1038/s41524-022-00744-4.

Tan AR, Urata S, Goldman S, Dietschreit JCB, Gómez-Bombarelli R. Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles. npj Comput Mater. 2023;9(1).
https://doi.org/10.1038/s41524-023-01180-8.

Perez D, Subramanyam APA, Maliyov I, Swinburne TD. Uncertainty quantification for misspecified machine learned interatomic potentials. npj Comput Mater. 2025;11(1).
https://doi.org/10.1038/s41524-025-01758-4.

Zaverkin V, Holzmüller D, Christiansen H, Errica F, Alesiani F, Takamoto M, et al. Uncertainty-biased molecular dynamics for learning uniformly accurate interatomic potentials. npj Comput Mater. 2024;10(1).
https://doi.org/10.1038/s41524-024-01254-1.

Zhang H, Chen W, Iyer A, Apley DW, Chen W. Uncertainty-aware mixed-variable machine learning for materials design. Sci Rep. 2022;12(1):19760.
https://doi.org/10.1038/s41598-022-23431-2.

Schwalbe-Koda D, Hamel S, Sadigh B, Zhou F, Lordi V. Model-free estimation of completeness, uncertainties, and outliers in atomistic machine learning using information theory. Nat Commun. 2025;16(1):4014.
https://doi.org/10.1038/s41467-025-59232-0.

Schwalbe-Koda D, Tan AR, Gómez-Bombarelli R. Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. Nat Commun. 2021;12(1).
https://doi.org/10.1038/s41467-021-25342-8.

Yuk SF, Sargin I, Meyer N, Krogel JT, Beckman SP, Cooper VR. Putting error bars on density functional theory. Sci Rep. 2024;14(1):20219.
https://doi.org/10.1038/s41598-024-69194-w.

Zhang Y, Ling C. A strategy to apply machine learning to small datasets in materials science. npj Comput Mater. 2018;4(1):25.
https://doi.org/10.1038/s41524-018-0081-z.

Choubisa H, Haque MA, Zhu T, Zeng L, Vafaie M, Baran D,et al. Closed-Loop error-correction learning accelerates experimental discovery of thermoelectric materials. Adv Mater. 2023;35(40).
https://doi.org/10.1002/adma.202302575.

Xie Y, Vandermause J, Ramakers S, Protik NH, Johansson A, Kozinsky B. Uncertainty-aware molecular dynamics from Bayesian active learning for phase transformations and thermal transport in SiC. npj Comput Mater. 2023;9(1).
https://doi.org/10.1038/s41524-023-00988-8.

Chitturi SR, Ramdas A, Wu Y, Rohr B, Ermon S, Dionne J, et al. Targeted materials discovery using Bayesian algorithm execution. npj Comput Mater. 2024;10(1).
https://doi.org/10.1038/s41524-024-01326-2.

Mahata A, Mukhopadhyay T, Chakraborty S, Asle Zaeem M. Atomistic simulation assisted error-inclusive Bayesian machine learning for probabilistically unraveling the mechanical properties of solidified metals. npj Comput Mater. 2024;10.
https://doi.org/10.1038/s41524-024-01200-1.

Kailkhura B, Gallagher B, Kim S, Hiszpanski A, Han TY. Reliable and explainable machine-learning methods for accelerated material discovery. npj Comput Mater. 2019;5(1).
https://doi.org/10.1038/s41524-019-0248-2.

Minami S, Hayashi Y, Wu S, Fukumizu K, Sugisawa H, Ishii M, et al. Scaling law of Sim2Real transfer learning in expanding computational materials databases for real-world predictions. npj Comput Mater. 2025;11(1):146.
https://doi.org/10.1038/s41524-025-01606-5.

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.
https://doi.org/10.1038/s41467-020-19597-w.

Wang C, Takeuchi I, Liu H, Yu H, Zhang J, Kusne AG,et al. Real-time experiment-theory closed-loop interaction for autonomous materials science. Science Advances. 2025;11(27).
https://doi.org/10.1126/sciadv.adu7426.

Mukherjee B, Kavalsky L, Hansen V, Viswanathan V. By how much can closed-loop frameworks accelerate computational materials discovery? Digit Discov. 2023;2:1112-25.
https://doi.org/10.1039/D2DD00133K.