ORIGINAL RESEARCH **Open access**

When Data Steers Design: Feedback Dynamics in AI-Guided Materials Exploration Pipelines

Lucas Meyer[1*], Anna Schmid[2], Stefan Braun[1]

Abstract

The integration of computational tools and data-driven methodologies has transformed materials engineering, enabling accelerated discovery through AI-assisted pipelines that link data acquisition, model training, and experimental validation. In this paradigm, materials informatics leverages vast datasets from high-throughput computations and multimodal sources to inform design decisions, yet inherent feedback dynamics often introduce biases that steer exploration trajectories in unintended ways. This conceptual manuscript identifies a critical gap in understanding how data-model-experiment loops can self-reinforce certain pathways, leading to narrowed exploration spaces and amplified discovery biases. To address this, we introduce the Feedback Steering Framework (FSF), a systems-level architecture that interprets the interplay between data representations, model inferences, and iterative design cycles. The framework elucidates mechanisms such as reinforcement discovery bias, where initial data patterns perpetuate model preferences, and exploration narrowing, wherein computational steering logics constrain the search space over successive iterations. By conceptualizing these dynamics, FSF provides insights into optimizing AI-guided materials exploration for broader epistemic coverage. Implications extend to computational materials science ecosystems, including enhanced uncertainty management in autonomous systems and more robust inverse design strategies, ultimately fostering resilient infrastructures for next-generation materials innovation. This work underscores the need for interpretive tools that balance computational efficiency with comprehensive discovery potential in data-steered environments.

Keywords Materials informatics, Uncertainty quantification, Representation learning, Machine learning in materials science, AI-guided discovery systems, Feedback loops in design

*Correspondence:
Lucas Meyer
lucas.meyer@gmail.com
[1] Department of Materials Modeling and Simulation, Faculty of Engineering, ETH Zurich, Zurich, Switzerland
[2] Department of Data-Driven Materials Science, Faculty of Engineering, University of Bern, Bern, Switzerland

Introduction

The field of computational and data-driven materials engineering has undergone a profound structural transformation over the past decade, catalyzed by converging advances in computational power, large-scale data infrastructures, and algorithmic innovation. This transformation signifies a decisive departure from historically dominant trial-and-error experimentation toward systematic, informatics-driven discovery strategies capable of navigating vast chemical and structural design spaces with unprecedented efficiency [1, 2]. Rather than relying solely on incremental empirical iteration, contemporary materials engineering increasingly operationalizes predictive modeling, automated simulation, and data integration to guide materials selection and optimization. At the core of this paradigm shift lies the integration of machine learning methodologies, which enable the extraction of latent patterns from heterogeneous datasets spanning computational simulations, laboratory measurements, and curated materials repositories [3, 4]. In this context, materials informatics has emerged as a

foundational discipline, structuring how data are aggregated, standardized, and mobilized to accelerate discovery pipelines [5]. These informatics ecosystems are deeply intertwined with high-throughput computational infrastructures that generate expansive datasets encompassing structural configurations, electronic properties, thermodynamic stabilities, and kinetic descriptors—often derived from density functional theory calculations or molecular dynamics simulations executed through automated workflows [6, 7].

The expanding role of artificial intelligence within these data ecosystems has redefined the epistemic architecture of materials engineering. Machine learning models—including graph neural networks, message-passing architectures, and deep representation learning systems—have proven instrumental in translating raw atomistic and microstructural data into actionable predictive insights [8, 9]. By encoding materials as relational graphs or hierarchical feature embeddings, these models capture complex interdependencies among compositional, structural, and environmental variables, enabling high-fidelity property prediction across diverse materials classes [10, 11]. The informational landscape has been further enriched through the proliferation of multimodal datasets that integrate experimental observations, computational outputs, imaging data, and spectroscopic signatures within unified modeling frameworks [12, 13]. Such multimodal convergence enhances predictive robustness while enabling cross-domain inference that bridges theoretical simulation with empirical validation. Within this integrated ecosystem, autonomous discovery platforms have gained increasing prominence. These systems automate iterative cycles of hypothesis generation, candidate screening, synthesis, and characterization through closed-loop experimental architectures [14, 15]. Here, artificial intelligence functions not merely as an analytical instrument but as an active orchestrator of discovery trajectories, dynamically guiding subsequent computational or laboratory actions based on real-time learning feedback [16, 17].

High-throughput infrastructures further amplify this discovery capacity by enabling the rapid screening of thousands to millions of candidate materials across expansive compositional manifolds [18]. Such pipelines depend on scalable data processing frameworks, surrogate modeling accelerators, and optimized workflow management systems capable of traversing chemical spaces that would otherwise remain computationally or experimentally intractable [19, 20]. Inverse materials design paradigms exemplify this capability by inverting conventional property-to-structure mappings: instead of predicting performance from known materials, optimization algorithms identify compositions and structures that satisfy predefined functional targets [21, 22]. This inversion is particularly consequential in chemically complex domains such as halide perovskites, catalytically active oxides, and high-entropy alloys, where combinatorial dimensionality necessitates computational steering mechanisms to guide exploration [14, 18]. Yet as these infrastructures scale, they simultaneously introduce epistemic and operational challenges concerning data quality, representational fidelity, and model generalization across heterogeneous materials regimes [23, 24].

Despite the remarkable acceleration enabled by AI-driven discovery ecosystems, contemporary computational paradigms exhibit structural limitations that constrain their exploratory completeness. A central vulnerability lies in the dependency on initial training datasets, which often encode historical research biases, experimental accessibility constraints, or simulation feasibility filters [4, 25]. Models trained within such bounded epistemic environments may systematically overlook rare, metastable, or unconventional materials candidates, thereby narrowing the effective search horizon [6, 26]. This phenomenon is further amplified within coupled simulation–experiment pipelines, where outputs from one stage recursively inform subsequent sampling decisions. Over iterative cycles, such feedback coupling can generate epistemic echo chambers in which certain material families receive disproportionate algorithmic attention while others remain underexplored [27]. Computational trade-offs—such as balancing simulation accuracy against scalability in force-field approximations or neural network architectures—further exacerbate this narrowing by privileging tractable regions of materials space over computationally demanding ones [17, 28].

Epistemic constraints also manifest through the handling of uncertainty and interpretability within AI-guided discovery systems. Robust uncertainty quantification is essential for distinguishing between confident predictions and those arising from data sparsity or model extrapolation [19]. Without explicit mechanisms to characterize epistemic uncertainty—reflecting incomplete knowledge rather than stochastic variability—discovery pipelines risk premature convergence on locally optimal but globally limited solutions [6, 23]. Interpretability challenges compound this issue, particularly in deep learning systems where predictive

rationales may remain opaque to domain experts [9, 16]. In representation learning contexts, descriptor selection and embedding architectures can inadvertently privilege certain crystallographic symmetries, bonding motifs, or compositional regimes, subtly steering exploration away from unconventional or emergent materials classes [8, 24]. Collectively, these dynamics reveal a systemic paradox: while artificial intelligence enhances discovery efficiency, it may simultaneously impose path dependencies that constrain epistemic diversity within exploration trajectories [1, 10].

Addressing these structural limitations necessitates a conceptual reexamination of the feedback architectures embedded within AI-guided materials discovery ecosystems. The interplay among data ingestion, model retraining, uncertainty evaluation, and experimental steering constitutes a densely interconnected network of iterative influences that shape how materials spaces are navigated [3, 12]. Yet these feedback dynamics remain insufficiently theorized within existing computational design literature. This manuscript introduces the Feedback Steering Framework (FSF) as a novel interpretive construct for analyzing how informational flows regulate discovery directionality. By conceptualizing materials exploration as a feedback-governed system, FSF elucidates how biases emerge, propagate, and potentially self-reinforce across iterative cycles. More importantly, the framework identifies intervention points through which adaptive recalibration—via uncertainty integration, diversity-aware sampling, and representational expansion—may broaden exploratory horizons. In doing so, FSF provides systems-level insight into the epistemic governance of AI-driven materials engineering and outlines pathways toward more reflexive, bias-aware, and discovery-expansive computational ecosystems.

Theoretical Background & Literature Synthesis

Materials data infrastructures

The foundation of contemporary computational materials engineering is anchored in increasingly sophisticated data infrastructures that enable the aggregation, standardization, and dissemination of materials knowledge across computational and experimental domains. These infrastructures encompass curated repositories containing atomic structures, thermodynamic and electronic property predictions, synthesis parameters, and experimentally validated performance metrics. Such ecosystems are often embedded within integrated materials informatics platforms designed to facilitate interoperability between simulation outputs and machine learning environments [1, 2, 5]. Through these platforms, heterogeneous datasets are rendered computationally accessible, forming the epistemic substrate upon which data-driven discovery pipelines operate.

High-throughput computational frameworks constitute a central operational engine within these infrastructures. Automated workflows orchestrate density functional theory calculations, molecular dynamics simulations, and phase stability analyses across expansive compositional spaces, generating datasets at scales unattainable through conventional experimentation alone [7, 18]. These computational pipelines not only accelerate candidate screening but also produce structured training corpora for predictive modeling architectures. Increasingly, infrastructures are evolving toward multimodal integration, wherein quantum-chemical calculations are fused with spectroscopic signatures, microstructural imaging, and process metadata to construct multidimensional representations of materials systems [12, 13]. Such multimodal assemblages enrich feature diversity and enable cross-property inference within unified learning frameworks.

Despite their transformative capacity, the expansion of materials data infrastructures introduces systemic challenges related to harmonization and epistemic reliability. Variability in file formats, metadata completeness, simulation parameters, and experimental provenance can introduce latent inconsistencies that propagate through downstream analytics [24, 26]. These heterogeneities complicate reproducibility and may distort model calibration if not systematically reconciled. Consequently, the literature underscores the necessity of standardized ontologies, interoperable data schemas, and provenance-tracking protocols capable of preserving contextual integrity across datasets [10, 11]. Standardization initiatives are thus not merely technical conveniences but foundational governance mechanisms ensuring that data flows coherently into AI pipelines while maintaining scientific traceability.

Representation learning architectures

At the algorithmic core of data-driven materials engineering lies the challenge of representation—namely, how to encode complex chemical and structural systems into machine-interpretable formats without eroding physically meaningful information. Representation learning architectures address this challenge by transforming atomic configurations, bonding environments, and crystallographic symmetries into structured numerical embeddings suitable for predictive modeling. Among these, graph neural networks (GNNs) have emerged as a dominant paradigm due to their capacity to model relational topologies intrinsic to molecules and crystalline solids. By representing atoms as nodes and interatomic interactions as edges, GNNs capture local coordination environments and long-range structural dependencies, enabling robust prediction of electronic, mechanical, and thermodynamic properties [8].

Extending this paradigm, deep learning variants such as atoms-in-molecules networks incorporate hierarchical encodings of electronic density and local chemical environments, facilitating transferability across compositional families and bonding regimes [9, 11]. These architectures enhance generalization by embedding chemically intuitive priors within latent spaces, thereby aligning statistical learning processes with domain knowledge. Parallel to these developments, surrogate modeling approaches provide computationally efficient approximations of resource-intensive simulations. By learning functional mappings between structure and property, surrogate models enable rapid screening across vast design spaces while preserving acceptable predictive fidelity [10, 17].

However, representational expressivity introduces trade-offs. Architectures with high parametric complexity demand extensive training data and computational resources, creating vulnerability to overfitting in data-scarce regimes [4, 23]. This tension between representational richness and statistical robustness remains a central design constraint in materials AI. Consequently, advances in feature engineering emphasize the construction of descriptors invariant to rotational, translational, and permutational symmetries—ensuring that learned representations reflect intrinsic material characteristics rather than coordinate artifacts [24, 25]. Such symmetry-preserving encodings are critical for enabling generalizable inference across crystallographic families and processing conditions.

AI-Guided discovery systems

Beyond static prediction, artificial intelligence increasingly functions as an active orchestrator of materials discovery processes. AI-guided discovery systems embed predictive models within iterative decision loops that dynamically refine hypotheses, prioritize candidate materials, and allocate experimental resources. Active learning frameworks exemplify this paradigm by leveraging uncertainty quantification to identify high-value data points for subsequent simulation or laboratory validation [6, 19]. Rather than exhaustively sampling materials space, these systems strategically interrogate regions where model confidence is lowest or expected information gain is highest, thereby optimizing exploration efficiency.

The operationalization of such frameworks has given rise to autonomous discovery platforms that couple machine learning algorithms with robotic synthesis and characterization infrastructure. These closed-loop laboratories execute end-to-end cycles of prediction, fabrication, measurement, and retraining with minimal human intervention, dramatically accelerating optimization timelines for functional materials [14, 15]. Within these environments, transfer learning strategies further enhance discovery capacity by enabling knowledge acquired in one compositional or structural domain to inform predictions in adjacent materials classes, mitigating data sparsity constraints [12, 16].

Yet, the recursive nature of AI-guided discovery introduces epistemic risks. Iterative retraining on model-selected data can amplify initial biases embedded within training corpora or algorithmic priors, generating self-reinforcing discovery trajectories [3]. Over successive cycles, such feedback loops may narrow exploratory diversity, privileging familiar chemistries while marginalizing unconventional candidates. Closed-loop systems, though operationally efficient, therefore require careful calibration of exploration–exploitation balances to prevent entrapment within locally optimal but globally suboptimal search spaces [27, 29]. Governance mechanisms—including diversity constraints, bias audits, and hybrid human-AI oversight—are increasingly recognized as necessary safeguards for maintaining epistemic breadth within autonomous discovery ecosystems.

Computational design paradigms

Computational design paradigms represent a fundamental epistemic shift in materials engineering, transitioning from descriptive prediction toward prescriptive generation.

Inverse design frameworks exemplify this shift by inverting the traditional forward modeling problem: rather than predicting properties from known structures, these paradigms algorithmically search compositional and structural spaces to identify candidates that satisfy predefined performance criteria [20, 21]. Optimization engines—ranging from gradient-based solvers to evolutionary algorithms—operate across high-dimensional design manifolds, iteratively refining candidate solutions based on objective functions linked to target properties such as bandgap, catalytic activity, or mechanical resilience.

Generative modeling architectures extend the exploratory capacity of inverse design by probabilistically sampling previously unobserved materials configurations. Adversarial networks, in particular, learn latent distributions of structural motifs and generate novel candidates that conform to learned chemical and physical constraints [20]. By navigating beyond enumerated databases, these models expand the explorable design domain and introduce pathways for discovering materials absent from existing repositories. Variational autoencoders and diffusion-based generators similarly contribute to this paradigm by embedding materials within continuous latent spaces amenable to interpolation and property-conditioned sampling.

To manage the computational burdens associated with design exploration, multi-fidelity modeling frameworks integrate simulations of varying accuracy and cost. Low-fidelity approximations—such as coarse-grained simulations or empirical potentials—enable rapid screening, while high-fidelity methods like density functional theory provide targeted validation for shortlisted candidates [16, 22]. This hierarchical orchestration balances computational efficiency with predictive rigor, enabling scalable design workflows without sacrificing scientific reliability.

More recently, the emergence of foundation models for scientific domains has introduced a further layer of generalization. Pre-trained on expansive, heterogeneous materials datasets, these architectures learn transferable representations that can be fine-tuned for specialized design tasks, including property prediction, synthesis planning, and stability analysis [9, 23]. Their cross-domain adaptability positions them as infrastructural models capable of supporting multiple downstream discovery objectives.

Despite these advances, computational design paradigms remain constrained by the combinatorial vastness of materials space. High-dimensional compositional and structural variables generate curse-of-dimensionality effects that render exhaustive exploration computationally intractable [18, 28]. Sampling sparsity, optimization plateaus, and latent space discontinuities can impede convergence toward globally optimal solutions. Consequently, the literature emphasizes the necessity of adaptive design logics capable of dynamically recalibrating search trajectories, incorporating feedback from uncertainty metrics and discovery diversity indicators [7, 25]. Such adaptive paradigms aim to mitigate self-reinforcing optimization tendencies that may otherwise confine exploration within narrow regions of materials space.

Uncertainty & interpretability

Uncertainty quantification constitutes a foundational pillar for epistemically robust AI deployment in materials engineering. Given the high stakes associated with materials discovery—where predictions inform costly synthesis and deployment decisions—distinguishing between different sources of predictive uncertainty is essential. The literature commonly differentiates aleatoric uncertainty, arising from intrinsic variability in measurements or stochastic processes, from epistemic uncertainty, which reflects model ignorance stemming from limited or biased training data [6, 19]. This bifurcation enables more nuanced risk assessment, guiding decisions regarding where additional data acquisition or model refinement is most warranted.

Bayesian inference frameworks provide a principled statistical foundation for uncertainty estimation, embedding probabilistic reasoning directly within model architectures. Posterior distributions over model parameters enable the derivation of predictive confidence intervals, facilitating risk-aware materials screening. Ensemble learning strategies offer a complementary approach, approximating epistemic uncertainty through variance across independently trained models [23, 29]. Such techniques are frequently integrated into active learning pipelines, where uncertainty estimates inform the selection of high-information candidates for further simulation or experimental validation.

Interpretability mechanisms operate in tandem with uncertainty quantification to enhance transparency in AI-guided inference. Feature attribution methods, saliency mapping, and attention-weight visualization enable

researchers to interrogate the structural or compositional drivers underlying model predictions [9, 16]. In materials contexts, these interpretive tools are often adapted to reflect chemically meaningful descriptors—such as coordination environments, bonding motifs, or defect distributions—thereby aligning algorithmic reasoning with domain expertise.

However, interpretability remains unevenly tractable across architectural classes. Highly parameterized deep learning systems, particularly those operating on multimodal or hierarchical representations, may exhibit substantial opacity, limiting the capacity to trace causal pathways within predictive outputs [3, 11]. This opacity introduces epistemic risk, as stakeholders may over-rely on high-confidence predictions without fully understanding their structural basis or domain validity.

Consequently, contemporary literature frames uncertainty and interpretability not as auxiliary features but as infrastructural requirements for trustworthy computational design. Integrated frameworks that embed uncertainty-aware reasoning, interpretability dashboards, and audit trails within discovery platforms are increasingly advocated as mechanisms for expanding exploration horizons while maintaining scientific accountability [4, 17]. By coupling predictive performance with epistemic transparency, such systems aim to balance acceleration in materials innovation with the reliability necessary for translational deployment.

Proposed conceptual framework

The Feedback Steering Framework (FSF) emerges as an original systems architecture designed to interpret the intricate dynamics within AI-guided materials exploration pipelines. At its core, FSF conceptualizes the discovery process as a layered network where data inflows interact with model architectures to steer experimental or computational iterations. The framework comprises three primary structural layers: the Data Assimilation Layer, which ingests and preprocesses multimodal inputs; the Inference Modulation Layer, responsible for model training and adaptation; and the Exploration Guidance Layer, which directs subsequent design actions based on inferred patterns. These layers are interconnected through bidirectional feedback channels that regulate information flow, ensuring that discoveries are not static but evolve in response to accumulating insights. The structural components, feedback pathways, and epistemic steering functions embedded within FSF are synthesized in **Table 1**.

Table 1. Structural architecture and feedback steering mechanisms within the Feedback Steering Framework (FSF).

FSF Layer	Core Function	Data / Signal Inputs	Feedba Outpu
Data Assimilation Layer	Aggregates and preprocesses multimodal materials data	High-throughput simulations; experimental datasets; literature corpora; spectroscopic and imaging data	Curate trainin dataset representa ready fea matrice
Inference Modulation Layer	Trains and updates predictive and generative AI models	Encoded structural descriptors; graph embeddings; latent representations	Proper predictic uncertai estimat candida rankin
Exploration Guidance Layer	Directs computational or experimental discovery trajectories	Model outputs; novelty metrics; optimization objectives	Candida prioritizat synthes targets simulati workflo
Feedback Coupling Channels	Transmit iterative updates across layers	Validation results; experimental feedback; retraining datasets	Mode recalibrat datase refineme steerin adjustme
Epistemic Risk Structures	Capture misalignments between representation and inference	Descriptor gaps; extrapolative predictions; uncertainty gradients	Risk sign interpreta diagnos

Central to FSF are the data-model-discovery pipelines, which form closed circuits amplifying or attenuating

exploration trajectories. In this setup, initial data representations—derived from high-throughput sources—feed into models that generate hypotheses, which in turn inform targeted explorations. Feedback loops manifest as recursive updates: model outputs refine data selection criteria, while exploration outcomes validate or challenge prior inferences. This cyclical nature introduces computational steering logics, where algorithms prioritize paths based on confidence scores or novelty metrics, potentially leading to self-reinforcing screening wherein dominant patterns overshadow alternatives.

To formalize these interactions, the reinforcement discovery bias can be conceptualized as a dynamic equilibrium between data diversity (D) and model convergence (M), expressed as

$$B = D / (M + \varepsilon) \quad (1)$$

where ε represents a regularization term accounting for external perturbations like uncertainty injections. This relation captures how diminishing data diversity accelerates model convergence, intensifying bias over iterations. Furthermore, exploration narrowing may be expressed as

$$N = \int (F(t)\ dt) \quad (2)$$

where F(t) denotes the feedback intensity at time t, integrating cumulative steering effects that contract the searchable space.

FSF also incorporates epistemic risk structures, viewing them as embedded within the pipelines. Here, risks arise from representation-inference mismatches, where incomplete encodings lead to skewed guidances. A third formula interprets this as

$$R = \Sigma (W_i * \Delta_i) \quad (3)$$

summing weighted discrepancies (Δ_i) across inference weights (W_i), highlighting how misalignments propagate through layers. These formulas underscore the framework's interpretive power, revealing trade-offs in infrastructure design—such as balancing feedback amplification for efficiency against diversification for comprehensiveness. As conceptualized in **Figure 1**, the framework provides a blueprint for analyzing discovery ecosystems and dissecting how data steers design.

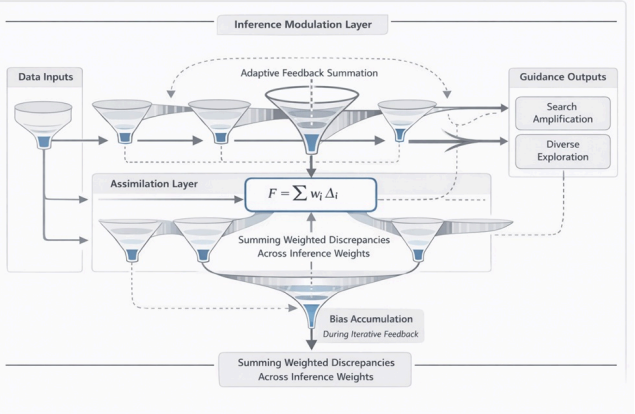

Figure 1. Layered architecture of the discovery framework, illustrating data flow from the assimilation layer through modulation and inference, with feedback arrows and narrowing funnels representing bias accumulation and steering nodes.

Analytical implications

The Feedback Steering Framework (FSF) offers interpretive lenses for dissecting the systemic behaviors in AI-guided materials exploration, revealing how feedback dynamics influence overall pipeline efficacy. At a systems level, FSF interprets the data-model-experiment loops as interconnected modules where perturbations in one layer cascade through others, affecting the trajectory of discovery. For instance, in materials informatics ecosystems, the assimilation layer's handling of multimodal datasets can either diversify or homogenize inputs, directly impacting the modulation layer's ability to generate balanced inferences [5, 13]. This interplay suggests that steering logics, if left unchecked, may prioritize computational efficiency over epistemic breadth, leading to infrastructures that favor incremental refinements rather than radical innovations [1, 6].

Pipeline optimization insights emerge from FSF's emphasis on feedback channels, which can be tuned to incorporate diversification mechanisms. By interpreting reinforcement discovery bias as a function of iterative updates, frameworks like FSF highlight opportunities for injecting variability—such as through adaptive sampling that counters model convergence [6, 19]. Consider the interaction between representation learning and inference modulation: architectures that embed graph-based encodings may inadvertently reinforce structural motifs prevalent in training data, narrowing the scope of inverse design [8, 20]. To mitigate this, computational steering

could integrate hybrid logics that blend deterministic predictions with stochastic explorations, fostering pipelines resilient to initial data skews [10, 21].

Bias mitigation strategies within FSF revolve around epistemic risk structures, where discrepancies in representation-inference alignments are monitored and adjusted. This can be conceptualized as a trade-off metric,

$$\begin{matrix} T \\ = \ \alpha \ * B \\ + \ \beta \ * N, \end{matrix} \quad (4)$$

where α and β weight the contributions of bias (B) and narrowing (N), guiding infrastructure adjustments to minimize cumulative risks [4, 25]. Such formalizations interpret how uncertainty quantification can serve as a corrective feedback, recalibrating models to explore underrepresented regions of chemical space [23, 29]. In practice, this implies designing systems that dynamically assess feedback intensity, preventing self-reinforcing screening by thresholding loop iterations or introducing external data perturbations [27].

Infrastructure resilience is another key implication, as FSF underscores the vulnerabilities in closed-loop systems to feedback amplification. Autonomous discovery platforms, for example, benefit from interpretive tools that evaluate loop stability, ensuring that high-throughput computations do not entrench biases from early cycles [14, 15]. By framing discovery broadening logics as expansive countermeasures, FSF suggests integrating multi-fidelity hierarchies that allow low-cost explorations to probe beyond converged paths [16, 22]. This resilience extends to handling computational constraints, where trade-offs in model complexity are balanced against exploration coverage, promoting ecosystems that adapt to evolving data landscapes [17, 28].

Ultimately, these analytical implications position FSF as a diagnostic framework for enhancing AI-guided pipelines. Through systems-level interpretations, it reveals how feedback dynamics can be harnessed to steer materials exploration toward more inclusive outcomes, mitigating the epistemic constraints that hinder comprehensive design [3, 12]. The formulas embedded in FSF provide conceptual anchors for these optimizations, enabling computational materials engineers to anticipate and counteract narrowing effects in real-time workflows [7, 18].

Results and Discussion

The Feedback Steering Framework (FSF) connects to broader trends in computational materials engineering, where data-driven paradigms increasingly dominate discovery ecosystems. By interpreting feedback dynamics, FSF aligns with ongoing efforts to couple simulations and experiments, offering insights into how AI can enhance rather than constrain innovation [2, 7]. In materials informatics, this framework complements representation learning by highlighting the role of feedback in refining descriptors, potentially informing the development of more adaptive architectures [8, 24]. Similarly, in high-entropy alloys or perovskite design, FSF's steering logics provide a lens for understanding how data loops influence compositional searches, bridging gaps between theoretical predictions and practical synthesis [14, 18].

Impacts on computational ecosystems are profound, as FSF encourages infrastructures that prioritize feedback-aware designs. For instance, foundation models for science could incorporate FSF principles to mitigate biases in pre-training phases, ensuring transferable knowledge across domains [9, 23]. This extends to uncertainty management, where epistemic risks interpreted through FSF foster more interpretable systems, aligning with calls for transparent AI in materials science [16, 19]. However, limitations of the framework must be acknowledged: as a conceptual tool, FSF relies on interpretive reasoning rather than quantitative validations, potentially overlooking domain-specific nuances in data handling [4, 26]. Its focus on feedback may also underemphasize external factors, such as hardware constraints or interdisciplinary integrations, that shape real-world pipelines [17, 28].

Future conceptual directions could expand FSF to encompass multi-agent systems, where collaborative AI entities negotiate steering decisions, broadening exploration in complex spaces [15, 20]. Additionally, integrating FSF with generative paradigms might yield hybrid frameworks that simulate feedback scenarios virtually, aiding in the design of resilient discovery platforms [21]. These directions underscore the need for ongoing synthesis of literature, evolving FSF into a foundational element for next-generation materials engineering ecosystems [1, 5].

Conclusion

In summary, this manuscript introduces the Feedback Steering Framework (FSF) as an original interpretive architecture for understanding feedback dynamics in AI-guided materials exploration pipelines. By conceptualizing data-model-experiment loops and their inherent biases, FSF provides systems-level insights into reinforcement discovery bias, exploration narrowing, and self-reinforcing screening, offering tools to analyze and optimize computational workflows. The implications for materials engineering are significant, emphasizing the balance between efficiency and epistemic coverage in data-steered designs.

This work calls for integrative approaches that embed feedback-aware logics into discovery infrastructures, fostering advancements in autonomous systems and inverse design. Ultimately, FSF contributes to a more nuanced comprehension of how data steers design, paving the way for robust, adaptive strategies in computational materials science.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 25 Jun 2021 Revised: 12 Aug 2021 Accepted: 18 Nov 2021
Published online: 18 March 2022

References

Schmidt J, Marques MRG, Botti S, Marques MAL. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):83. https://doi.org/10.1038/s41524-019-0221-0.

Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. Machine learning in materials informatics: Recent applications and prospects. npj Comput Mater. 2017;3(1):54. https://doi.org/10.1038/s41524-017-0056-5.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55. https://doi.org/10.1038/s41586-018-0337-2.

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.

Jablonka KM, Ongari D, Moosavi SM, Smit B. Big-data science in porous materials: Materials genomics and machine learning. Chem Rev. 2020;120(16):8066-129. https://doi.org/10.1021/acs.chemrev.0c00004.

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. https://doi.org/10.1038/s41524-019-0153-8.

Avery P, Wang X, Oses C, Gossett E, Proserpio DM, Toher C, et al. Predicting superhard materials via a machine learning informed evolutionary structure search. npj Comput Mater. 2019;5(1):89. https://doi.org/10.1038/s41524-019-0226-8.

Chen C, Ye W, Zuo Y, Zheng C. Graph networks as a universal machine learning framework for molecules and crystals. Chem Mater. 2019;31(9):3564-72. https://doi.org/10.1021/acs.chemmater.9b01294.

Deringer VL, Bartók AP, Proserpio DM, Day GM, Csányi G, Pickard CJ. Machine learning interatomic potentials as emerging tools for materials science. Adv Mater. 2019;31(46):1902765. https://doi.org/10.1002/adma.201902765.

Nyshadham C, Rupp M, Bekker B, Shapeev AV, Mueller T, Rosenbrock CW, et al. Machine-learned multi-system surrogate models for materials prediction. npj Comput Mater. 2019;5(1):51. https://doi.org/10.1038/s41524-019-0189-9.

Zubatyuk R, Smith JS, Leszczynski J, Isayev O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. Sci Adv. 2019;5(8):eaav6490. https://doi.org/10.1126/sciadv.aav6490.

Gupta V, Choudhary K, Tavazza F, Campbell C, Liao W-k, Choudhary A, et al. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. Nat Commun. 2021;12(1):6595. https://doi.org/10.1038/s41467-021-26921-5.

Rosen AS, Iyer SM, Ray D, Yao Z, Aspuru-Guzik A, Gagliardi L et al. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. Matter. 2021;4(5):1578-97. https://doi.org/10.1016/j.matt.2021.02.021.

Tao Q, Xu P, Li M, Lu W. Machine learning for perovskite materials design and discovery. npj Comput Mater. 2021;7(1):23. https://doi.org/10.1038/s41524-021-00495-8.

Hatakeyama-Sato K, Tezuka T, Ujihira W, Morita Y, Nishide H, Oyaizu K. Tackling the challenge of a huge materials science search space with quantum-inspired annealing. Adv Intell Syst. 2021;3(2):2000209. https://doi.org/10.1002/aisy.202000209.

Pilania G. Machine learning in materials science: From explainable predictions to autonomous design. Comput Mater Sci. 2021;193:110360. https://doi.org/10.1016/j.commatsci.2021.110360.

Mishin Y. Machine-learning interatomic potentials for materials science. Acta Mater. 2021;214:116980. https://doi.org/10.1016/j.actamat.2021.116980.

Dai D, Xu T, Wei X, Ding G, Xu Y, Zhang J, et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. Comput Mater Sci. 2020;175:109618. https://doi.org/10.1016/j.commatsci.2020.109618.

Keith JA, Vassilev-Galindo V, Cheng B, Chmiela S, Gastegger M, Müller KR, et al. Combining machine learning and computational chemistry for predictive insights into chemical systems. Chem Rev. 2021;121(16):9816-72. https://doi.org/10.1021/acs.chemrev.1c00107.

Kim C, Batra R, Chen L, Tran H, Ramprasad R. Polymer design using genetic algorithm and machine learning. Comput Mater Sci. 2021;186:110067. https://doi.org/10.1016/j.commatsci.2020.110067.

Glielmo A, Zeni C, De Vita A. Efficient nonparametric n-body force fields from machine learning. Phys Rev B. 2018;97(18):184307. https://doi.org/10.1103/PhysRevB.97.184307.

Chen L, Tran H, Batra R, Kim C, Ramprasad R. Machine learning models for the prediction of energy, forces, and stresses for platinum. npj Comput Mater. 2021;7(1):19. https://doi.org/10.1038/s41524-021-00490-z.

Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, Schütt KT, et al. Machine learning force fields. Chem Rev. 2021;121(16):10142-10186. https://doi.org/10.1021/acs.chemrev.0c01111.

Himanen L, Jäger MOJ, Morooka EV, Federici Canova F, Ranawat YS, Gao DZ, et al. DScribe: Library of descriptors for machine learning in materials science. Comput Phys Commun. 2020;247:106949. https://doi.org/10.1016/j.cpc.2019.106949.

Rosenbrock CW, Homer ER, Csányi G, Hart GLW. Discovering the building blocks of atomic systems via machine learning. npj Comput Mater. 2017;3(1):29. https://doi.org/10.1038/s41524-017-0034-y.

Haghighatlari M, Vishwakarma G, Altarawy D, Subramanian R, Kota BU, Sonpal A, et al. ChemML: A machine learning and informatics program package for the analysis, mining, and modeling of chemical and materials data. Wiley Interdiscip Rev

Comput Mol Sci. 2020;10(4):e1458. https://doi.org/10.1002/wcms.1458.

Veit M, Jain SK, Bonakala S, Rudra I, Hohl D, Csányi G. Equation of state of fluid methane from first principles with machine learning potentials. J Chem Theory Comput. 2019;15(4):2574-86. https://doi.org/10.1021/acs.jctc.8b01242.

Botu V, Batra R, Chapman J, Ramprasad R. Machine learning force fields: Construction, validation, and outlook. J Phys Chem C. 2017;121(1):511-22. https://doi.org/10.1021/acs.jpcc.6b10908.

Rosen AS, Fung V, Huck P, O'Donnell CT, Horton MK, Trask A, et al. High-throughput predictions of metal-organic framework electronic properties: Theoretical challenges, graph neural networks, and data exploration. npj Comput Mater. 2020;6(1):112. https://doi.org/10.1038/s41524-020-00389-1.