ORIGINAL RESEARCH **Open access**

Algorithmic Path Dependence in AI-Driven Materials Design

Maria Gonzalez[1*], Javier Ruiz[1], Lucia Torres[2]

Abstract

The integration of artificial intelligence into materials design processes introduces complex dynamics where initial algorithmic choices shape subsequent trajectories, often embedding persistent dependencies that influence innovation pathways. This manuscript explores the conceptual underpinnings of path dependence, examining how data selection, model architectures, and iterative learning mechanisms interweave to form self-reinforcing structures in AI-assisted materials discovery. Through a synthesis of recent literature, it examines the interpretive implications of bias propagation, feedback loops, and epistemic constraints in computational materials science. The proposed framework conceptualizes these elements as interconnected layers, in which early decisions cascade through design cycles, shaping the exploration of material spaces and the emergence of novel properties. By focusing on systems-level insights, the analysis highlights trade-offs between efficiency and diversity in algorithmic guidance, as well as ethical considerations in steering material innovation. This interpretive approach underscores the need for reflective practices in AI-driven workflows, emphasizing how path-dependent logics can both constrain and enable creative outcomes in materials engineering. Ultimately, the discussion integrates these dynamics to reveal broader implications for sustainable and equitable advancements in the field, without positing empirical directives.

Keywords Feedback loops, Systems-level insights, Algorithmic path dependence, AI-driven materials design, Bias propagation, Epistemic constraints

*Correspondence:
Maria Gonzalez
maria.gonzalez@gmail.com
[1] Department of Materials Informatics and Data Modeling, Faculty of Engineering, University of Granada, Granada, Spain
[2] Department of AI Materials Simulation, Faculty of Engineering, University of Seville, Seville, Spain

Introduction

The advent of artificial intelligence (AI) in materials design represents a transformative shift in how scientists and engineers create and optimize new materials. Traditionally, materials discovery relied on empirical experimentation and theoretical modeling, often constrained by time, resources, and human intuition. With AI, particularly machine learning algorithms, the process accelerates, enabling the navigation of vast chemical and structural spaces that were previously inaccessible [1, 2]. Yet, this acceleration introduces subtle yet profound dependencies, where the algorithms' initial configurations and training paradigms imprint lasting influences on the design outcomes. These dependencies, akin to path-dependent phenomena observed in other technological domains, manifest as algorithmic path dependence, in which early choices in data curation, model selection, and optimization strategies shape the evolutionary paths of material innovation.

At its core, algorithmic path dependence in AI-driven materials design arises from the interplay between computational tools and the inherent complexities of material systems. Materials, by nature, encompass multifaceted properties—mechanical, electronic, thermal—that emerge from atomic and molecular interactions [3, 4]. AI models, trained on historical datasets, interpret these interactions through pattern recognition, but the datasets themselves carry legacies of

prior research priorities, experimental biases, and methodological preferences [5, 6]. Consequently, the algorithms may reinforce certain material classes or property profiles, potentially overlooking alternative pathways that could lead to breakthrough discoveries. This interpretive lens reveals how path dependence operates not merely as an artifact but as a fundamental dynamic shaping the epistemic landscape of materials science.

Consider the broader context: the push for sustainable materials, such as those for energy storage or environmental remediation, demands innovative designs that AI promises to deliver [7, 8]. However, if algorithmic frameworks inherit path-dependent constraints, the pursuit of sustainability might inadvertently favor incremental improvements over radical reinventions. For instance, models optimized for high-throughput screening may prioritize familiar crystal structures, thereby embedding a conservatism stemming from training data imbalances [9, 10]. This raises conceptual questions about the balance between exploiting known knowledge and exploring uncharted territories, where path dependence serves as a steering mechanism that influences resource allocation and creative potential.

Furthermore, integrating AI into materials design amplifies feedback loops. Iterative learning processes, where models refine predictions based on simulated or experimental feedback, create loops that can entrench initial assumptions [11, 12]. These loops, while enhancing predictive accuracy in bounded domains, may propagate epistemic blind spots, such as underrepresentation of rare events or emergent behaviors in complex materials [13, 14]. Interpreting these dynamics requires acknowledging the trade-offs: enhanced efficiency in design cycles comes at the cost of potential lock-ins, where divergent paths become increasingly costly to pursue.

Ethical dimensions also permeate this discourse. As AI-driven tools democratize access to advanced design capabilities, path dependence could exacerbate inequalities, favoring institutions with superior data resources or computational power [15, 16]. This interpretive insight suggests that the design process is not neutral; it reflects societal values embedded in algorithmic architectures. For example, biases in data sourcing might skew innovation toward applications in affluent sectors, sidelining needs in global challenges such as climate adaptation [17, 18].

In synthesizing these elements, the introduction sets the stage for a deeper exploration. The subsequent sections delve into the theoretical foundations, drawing on recent advances to examine how path dependence manifests in AI contexts. By examining literature on machine learning applications in materials, the analysis uncovers interaction dynamics between algorithmic elements and material realities [19, 20]. This leads to a proposed framework that integrates these insights, offering a conceptual map for understanding path-dependent logics without prescriptive mandates.

Ultimately, this manuscript contributes to the scholarly conversation by interpreting algorithmic path dependence as a lens for reflective practice in materials design. It encourages a nuanced appreciation of how, while powerful, AI embeds historical contingencies that shape future possibilities. Through this interpretive approach, the field can foster more resilient and inclusive innovation ecosystems, attuned to the intricate dance between computation and creation.

Theoretical Background and Literature Synthesis

Foundations of path dependence in technological systems

Path dependence, originating from economic and historical analyses, describes how initial conditions and contingent events lead to outcomes that become increasingly difficult to alter over time. In technological contexts, this concept illuminates how early design choices create lock-in effects that influence evolutionary trajectories [21, 22]. Applied to AI-driven materials design, path dependence extends beyond mere historical inertia to encompass algorithmic structures that perpetuate specific interpretive paradigms in material exploration.

Recent scholarship interprets these foundations through the lens of computational systems, where algorithms serve as mediators between human intent and material outcomes [23, 24]. For instance, the selection of initial training datasets imprints preferences that cascade through model iterations, akin to how early standardization in technologies like

QWERTY keyboards entrenches suboptimal configurations. In materials science, this manifests in the prioritization of certain chemical compositions or structural motifs, driven by data availability rather than intrinsic merit [25, 26].

Interpretive implications arise from the interaction between path-dependent algorithms and the stochastic nature of material properties. Feedback mechanisms, where model outputs inform subsequent data collection, reinforce dominant paths, potentially marginalizing alternative material spaces [27, 28]. This dynamic underscores trade-offs in design efficiency: while path dependence streamlines convergence toward viable solutions, it may constrain the epistemic breadth necessary for disruptive innovations.

Bias and data imbalances in AI models for materials

Bias propagation constitutes a central mechanism through which algorithmic path dependence manifests in AI-driven materials science. Initial data imbalances—often inherited from historically contingent research priorities—become embedded within model representations, shaping not only predictive outcomes but also the interpretive horizons of design processes. The literature consistently indicates that dominant datasets disproportionately reflect well-established material classes such as crystalline metals, semiconductors, and oxides. At the same time, comparatively sparse coverage exists for biomaterials, amorphous systems, complex nanostructures, or materials relevant to low-resource contexts [29, 30]. This asymmetry establishes an epistemic gradient in which certain material families are rendered algorithmically legible, while others remain peripheral or effectively invisible.

Such imbalances do not remain static; rather, they propagate through recursive feedback mechanisms. Models trained on skewed datasets generate predictions that preferentially reinforce familiar regions of material space, which in turn guide subsequent data acquisition, validation efforts, and research attention. Over iterative cycles, these dynamics solidify path-dependent trajectories in which algorithmic confidence grows around historically dominant materials, while exploratory breadth contracts. From a systems perspective, bias thus operates not merely as a flaw in representation but as a self-stabilizing feature of learning ecosystems, reinforcing continuity at the expense of epistemic plurality.

Active learning strategies, often proposed as corrective mechanisms, introduce additional layers of complexity. While designed to adaptively query underexplored regions, such strategies remain sensitive to initial query formulations and uncertainty estimations. Systems-level analyses suggest that if early selection criteria privilege specific property regimes or performance metrics, active learning can inadvertently amplify path dependence by repeatedly sampling within narrowly defined zones of interest [31, 32]. This dynamic illustrates a broader epistemic constraint: attempts to algorithmically “correct” imbalance remain bounded by the interpretive assumptions embedded at initialization. Ethical reasoning becomes salient here, as such constrained trajectories risk overlooking materials that align with societal priorities—such as affordability, resilience, or accessibility—particularly in regions where data scarcity intersects with structural inequities.

Interaction dynamics between data structures and algorithmic frameworks further complicate bias propagation. Multi-objective optimization paradigms, common in materials design, explicitly navigate trade-offs among competing objectives such as stability, performance, and cost. However, these frameworks inherit path-dependent logics from initial parameterizations, objective weightings, and dataset compositions [1, 3]. As a result, optimization landscapes are shaped less by abstract optimality than by historically sedimented preferences, guiding exploration along trajectories that reflect prior value commitments. Conceptually, this suggests that bias mitigation cannot be fully disentangled from path dependence itself; rather, biases are constitutive elements of how algorithmic systems learn, prioritize, and stabilize meaning within material spaces.

Feedback loops and iterative learning in design processes

Feedback loops form a second foundational pillar in the emergence of algorithmic path dependence, particularly within iterative design workflows. Adaptive learning mechanisms—central to contemporary AI systems—operate by recursively refining models based on prior outputs, evaluations, and selections. While such loops enable responsiveness and

efficiency, they also amplify early signals, embedding initial exploratory conditions into long-term convergence behaviors [4, 5]. In materials design, this implies that early-stage decisions regarding search direction, evaluation criteria, or surrogate fidelity exert disproportionate influence over eventual discovery outcomes.

At a conceptual level, feedback loops transform exploratory processes into trajectory-dependent systems, where the balance between exploration and exploitation becomes progressively constrained. Steering logics embedded within algorithms govern how this balance is negotiated, often privileging exploitation as confidence accumulates [6, 7]. While such convergence enhances predictive stability, it simultaneously narrows the interpretive bandwidth of design spaces, limiting the likelihood of encountering structurally novel phases, unconventional alloys, or atypical property regimes. Epistemic reasoning thus frames feedback loops as integrative yet delimiting mechanisms, synthesizing heterogeneous knowledge streams—ranging from quantum-scale simulations to macroscopic performance indicators—within bounded interpretive frames defined by early conditions.

Trade-offs between computational cost and representational accuracy further illuminate these dynamics. High-fidelity models, such as those grounded in detailed quantum mechanical approximations, offer precision but are tightly coupled to specific training paradigms, datasets, and assumptions [8, 9]. Their computational expense incentivizes reuse and refinement over reconfiguration, reinforcing dependence on established paths. Lower-fidelity models, while more flexible, often inherit constraints from their role as surrogates calibrated against high-fidelity outputs. Collectively, these trade-offs underscore how feedback loops shape innovation horizons by stabilizing certain forms of knowledge while marginalizing others, not through explicit exclusion but through cumulative reinforcement.

Epistemic and ethical dimensions of algorithmic guidance

Algorithmic guidance in AI-driven materials design is inherently epistemic, structuring how knowledge is represented, prioritized, and interpreted. Path dependence intensifies these effects by locking in specific representational schemas that persist across design cycles [10, 11]. Model architectures—particularly those encoding relational or topological assumptions, such as graph-based frameworks—instantiate structural priors that shape how materials are conceptualized and compared [12, 13]. Over time, these priors become normalized, guiding interpretation even as models evolve, thereby constraining the range of material narratives that are rendered plausible.

Ethical considerations are deeply entwined with these epistemic constraints. Path-dependent designs may inadvertently perpetuate societal biases by aligning material innovation with historically dominant industrial priorities, thereby influencing applications across sectors such as energy, infrastructure, and healthcare [14, 15]. Systems-level insights reveal a tension between algorithmic autonomy and human interpretive agency: as reliance on AI-guided pathways increases, the capacity for critical reflection on underlying assumptions may diminish. This raises ethical questions about responsibility, accountability, and the distribution of decision-making authority within human–AI collaborations.

Emerging technologies, particularly generative models, introduce additional layers to this analysis. While generative approaches promise expansive exploration of material possibilities, they remain grounded in foundational datasets and learned representations that encode prior dependencies [16, 17]. As such, generative creativity is tempered by inherited biases, producing outputs that often recombine familiar motifs rather than truly transcending established paradigms. Conceptually, this underscores the need for reflective interpretive practices that acknowledge the ethical–epistemic entanglements of algorithmic guidance without assuming that technical expansion alone resolves path-dependent constraints.

Integration of multi-scale perspectives in materials innovation

Multi-scale integration is frequently presented as a hallmark of AI-enabled materials innovation, linking atomic-level interactions with mesoscopic structures and macroscopic properties. However, path dependence plays a decisive role in shaping how these scales are connected and prioritized [18, 19]. Synthesis of recent work suggests that AI models often

privilege specific scales based on initial problem formulations, data availability, and representational convenience, creating dependencies that influence the coherence of cross-scale understanding [20, 22].

Feedback structures operating across scales further compound these effects. Micro-level biases—such as the overrepresentation of certain bonding environments or crystal symmetries—can propagate upward, shaping macro-level predictions of performance, durability, or sustainability [23, 24]. These dynamics reveal how local epistemic choices generate global interpretive consequences, reinforcing particular scale linkages while obscuring alternative couplings. Ethical reasoning highlights the implications of such asymmetries, as inclusive multi-scale frameworks are challenged by dependencies that favor established hierarchies of representation and interpretation.

Taken together, this expanded synthesis interprets the literature as converging on the recognition that bias, feedback, and scale integration are not separable concerns but interdependent expressions of algorithmic path dependence. Rather than framing these dynamics as correctable anomalies, the analysis positions them as constitutive features of AI-driven materials design, inviting nuanced awareness of how dependencies shape the evolution of material knowledge, innovation trajectories, and ethical commitments over time.

Proposed conceptual framework

The conceptual framework interprets algorithmic path dependence in AI-driven materials design as an interwoven system of layers in which initial conditions interact with iterative processes to shape design trajectories. At the foundational layer, data curation and model initialization form the bedrock, embedding interpretive biases that cascade upward. These elements interact dynamically with feedback mechanisms in the intermediary layer, creating steering logics that balance efficiency against exploratory breadth. The uppermost layer encompasses emergent outcomes, in which path-dependent constraints manifest as the diversity of material innovations.

Central to this framework is the recognition of trade-offs: enhanced predictive power through reinforced paths comes at the expense of epistemic flexibility, potentially narrowing the material space explored. Systems-level insights reveal how these layers integrate, with early dependencies amplifying through loops, influencing ethical considerations in design equity.

Interaction dynamics are key, as algorithmic choices at one layer reverberate across others, fostering self-sustaining structures. For instance, biased data inputs may steer models toward conservative optimizations, while reflective interventions could modulate these paths, though always within the framework's interpretive bounds.

Epistemic reasoning underscores the framework's utility for understanding how path dependence serves as a lens for innovation, highlighting feedback structures that both constrain and enable creative potential. The systemic risk of algorithmic path dependence is conceptualized in **Figure 1** as a hierarchical pyramid. This model illustrates how initial choices in data and model setup at the base cascade upward through self-reinforcing feedback loops, ultimately constraining the space of possible material innovation trajectories at the apex.

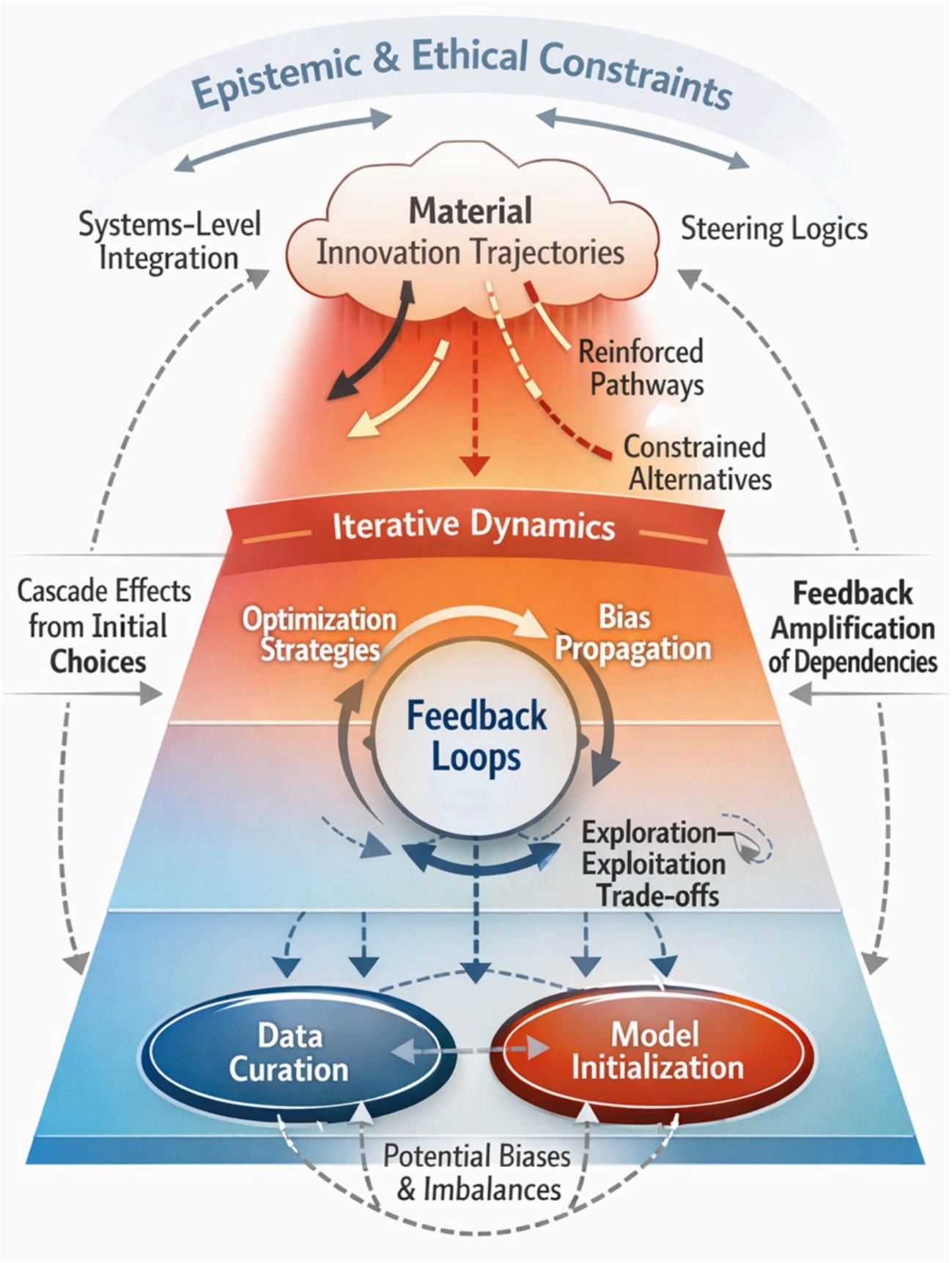
Epistemic & Ethical Constraints
Systems-Level Integration
Material Innovation Trajectories
Steering Logics
Reinforced Pathways
Constrained Alternatives
Iterative Dynamics
Cascade Effects from Initial Choices
Optimization Strategies
Bias Propagation
Feedback Loops
Feedback Amplification of Dependencies
Exploration–Exploitation Trade-offs
Data Curation
Model Initialization
Potential Biases & Imbalances

Figure 1. A multi-layered framework for algorithmic path dependence in AI-driven materials discovery. The schematic uses a pyramid structure to model the progression from foundational inputs and self-reinforcing iterative dynamics to emergent innovation trajectories, all shaped by overarching epistemic and ethical constraints.

These dynamics are synthesized in **Table 1**, which maps the systemic dimensions of algorithmic path dependence across foundational, iterative, epistemic, and ethical layers of AI-driven materials design.

Table 1. Algorithmic path dependence in AI-driven materials design: systemic dimensions, dynamics, and trade-offs

Dimension of path dependence	Core mechanism	Dominant interaction dynamics	Primary trade-offs	Epistemic and ethical implications
Foundational initialization	Early data curation, model architecture selection, and objective framing embed historical contingencies	Initial conditions cascade through iterative learning cycles, stabilizing interpretive regimes	Efficiency of convergence vs. openness of exploration	Early choices silently delimit material spaces, shaping what is considered *plausible* or *valuable* knowledge
Bias and data imbalance	Skewed datasets reflecting legacy research priorities	Recursive reinforcement through prediction-guided data acquisition	Predictive confidence vs. epistemic plurality	Marginalization of underrepresented material classes and socio-economic applications
Feedback loops and iteration	Adaptive refinement based on prior outputs and evaluations	Self-reinforcing loops amplify early signals and suppress deviation	Stability and accuracy vs. novelty and diversity	Innovation trajectories become resilient but resistant to epistemic disruption
Algorithmic steering logics	Exploration–exploitation balances encoded in learning strategies	Progressive tilt toward exploitation as confidence accumulates	Short-term optimization vs. long-term adaptability	Design spaces narrow over time, privileging familiar performance narratives
Model architecture effects	Structural priors embedded in representations (e.g., relational assumptions)	Normalization of specific material interpretations across cycles	Interpretive coherence vs. representational flexibility	Certain material narratives become dominant, shaping the ethical valuation of applications
Multi-scale integration	Selective linkage of atomic, mesoscopic, and macroscopic scales	Micro-level biases propagate upward through scale coupling	Computational tractability vs. holistic understanding	Local epistemic choices yield global interpretive consequences
Generative expansion	Creative recombination constrained by learned distributions	Apparent exploration remains tethered to foundational paths	Breadth of generation vs. depth of transformation	Generative novelty reproduces inherited assumptions
Interdisciplinary convergence	Early epistemic alignments across domains	Transfer of constraints between disciplinary frameworks	Integration efficiency vs. conceptual pluralism	Risk of reinforcing silos and unequal innovation capacity

Temporal structuring	Accumulation of dependencies over extended horizons	Past decisions scaffold future design narratives	Immediate performance vs. future flexibility	Long-term rigidity emerges from short-term optimization
System-level emergence	Interaction of all layers in self-organizing ecosystems	Path dependence becomes normalized as infrastructure	Robustness vs. epistemic reflexivity	Ethical responsibility diffuses as dependencies become invisible

Analytical implications

The interpretive analysis of algorithmic path dependence reveals a set of layered implications that extend beyond technical optimization, reshaping how AI-driven materials design is conceptually framed, ethically interpreted, and temporally governed. At a systems level, path dependence operates as a structuring condition rather than a residual artifact, delimiting the contours of explorable material spaces through the cumulative effects of early algorithmic decisions. Initial configurations—spanning data selection, representation choices, and objective prioritization—interact with iterative learning processes to generate emergent innovation patterns that stabilize over time. Consequently, design trajectories unfold not as neutral search processes but as interwoven networks of dependency, where early epistemic commitments cascade into enduring interpretive regimes [1, 2]. This implication foregrounds a fundamental trade-off in algorithmic steering: the efficiency gained through cumulative refinement versus the risk of epistemic lock-in that constrains interpretive diversity.

From an architectural perspective, the interaction between model formalisms and material complexity amplifies these implications. Graph-based neural networks, for example, encode relational assumptions that privilege certain structural motifs and interaction pathways, thereby reinforcing specific interpretive lenses while potentially marginalizing alternative material organizations [3, 4]. Over successive design cycles, such embedded assumptions are recursively validated by model outputs, producing feedback structures that normalize particular material narratives as dominant. This amplification effect extends into ethical domains, as reinforced paradigms implicitly guide judgments about material relevance, feasibility, and societal value. In application areas such as energy materials, path-dependent dynamics may favor incremental performance gains as measured by established efficiency metrics. At the same time, more radical sustainability transformations remain underexplored due to misalignment with inherited representational priors [5, 6].

Epistemically, path dependence functions as a mediator of knowledge integration across scales, shaping how insights from atomic-level interactions are aggregated into macroscopic design decisions. While algorithmic systems facilitate rapid synthesis across heterogeneous data sources, they simultaneously impose interpretive filters that prioritize historically salient correlations over emergent or weakly represented phenomena [7, 8]. The resulting trade-off lies between computational tractability and epistemic inclusivity: accelerated design cycles reduce exploratory friction but embed dependencies that influence which material futures are rendered legible. This has ethical implications for the allocation of research attention, as biases toward well-characterized, high-throughput datasets may narrow interpretations of material viability, sidelining considerations such as long-term accessibility, environmental justice, or region-specific material needs [9, 10].

At the level of innovation ecosystems, path dependence emerges as a catalyst for self-organizing behavior, in which feedback loops interact with initial conditions to produce resilient yet robust structures. Once stabilized, these structures shape how novelty is recognized and valued, creating adaptive but bounded innovation pathways [11, 12]. Such dynamics imply that deviation from dominant trajectories requires not merely technical intervention but epistemic disruption, as entrenched dependencies define the conditions under which alternative designs are deemed plausible. In generative modeling contexts, this manifests as an expansion of design spaces that remains tethered to foundational paths, trading apparent creative breadth for the persistence of embedded assumptions that guide generative outputs toward familiar motifs [13, 14].

The implications of path dependence also extend into interdisciplinary convergence, where AI-driven materials design intersects with domains such as biomimetics, quantum materials, or soft matter physics. Here, early integrations between algorithmic tools and domain-specific epistemologies shape how knowledge is translated across disciplinary boundaries [15, 16]. If initial alignments favor particular interpretive frameworks, subsequent collaborations may inherit these constraints, reinforcing silos rather than fostering genuinely integrative perspectives. Ethical reasoning underscores that such dynamics risk exacerbating disparities in innovation capacity, as path-dependent infrastructures may privilege well-resourced research communities while limiting the interpretive agency of emerging or underrepresented groups [17, 18].

Temporal considerations further enrich these analytical implications by framing path dependence as a scaffolding that structures the unfolding of design narratives over extended horizons. Early algorithmic efficiencies, when amplified through feedback loops, can yield long-term interpretive rigidities that reduce adaptability to evolving challenges such as climate resilience or resource scarcity [19, 20]. This temporal asymmetry highlights a trade-off between short-term performance optimization and long-term epistemic flexibility, emphasizing how present design decisions shape future horizons of possibility. Feedback structures thus function as interpretive bridges, linking past commitments to future constraints, while simultaneously revealing the cumulative ethical weight of early choices.

Collectively, these analytical implications position algorithmic path dependence as an intrinsic feature of AI-driven materials design rather than an incidental limitation. By elucidating how dependencies shape interaction dynamics, knowledge integration, and ethical interpretation across systems and timescales, the analysis deepens the conceptual understanding of how AI mediates material evolution. Rather than prescribing corrective strategies, this perspective invites reflective engagement with the conditions under which algorithmic trajectories are formed, sustained, and normalized within contemporary materials science.

Epistemic integration also features prominently, interpreting path dependence as a lens for understanding knowledge evolution in the field. By weaving together disparate strands—from bias mitigation strategies to multi-scale modeling—the discussion reveals how dependencies foster resilient yet adaptable interpretive structures [31, 32]. Trade-offs in this context involve balancing algorithmic autonomy with human interpretive oversight, where excessive reliance on paths could erode the richness of conceptual exploration.

In reflecting on these elements, the discussion posits path dependence not as a hindrance but as an interpretive opportunity, enabling a deeper grasp of how AI interlaces with materials design. Through systems-level and ethical reasoning, it illuminates the nuanced trade-offs and dynamics at play, contributing to a scholarly dialogue on fostering reflective and equitable innovation pathways.

Conclusion

In synthesizing the conceptual exploration of algorithmic path dependence in AI-driven materials design, the manuscript interprets this phenomenon as a fundamental dynamic intertwining computational logics with material innovation. Through analytical implications and integrative discussions, path dependence emerges as a structuring element, shaping interpretive and ethical landscapes as initial algorithmic choices interact with feedback mechanisms. Systems-level insights reveal trade-offs that balance efficiency and diversity, highlighting how dependencies both constrain and expand the field’s epistemic horizons.

This interpretive approach underscores the importance of reflective practices, acknowledging that path-dependent structures influence the steering of design trajectories toward sustainable and inclusive outcomes. Ultimately, by integrating these dynamics, the analysis enriches understanding of AI’s role in materials science, fostering a nuanced appreciation of its conceptual intricacies without empirical assertions.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 06 May 2022 Revised: 10 Jun 2022 Accepted: 15 Jul 2022
Published online: 18 January 2023

References

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. Annu Rev Mater Res. 2020;50:71-103.
https://doi.org/10.1146/annurev-matsci-070218-010015.

Vlassis NN, Flaschel M, Carrara P, De Lorenzis L. A review on data-driven constitutive laws for solids. arXiv. 2024.
https://doi.org/10.48550/arXiv.2405.03658.

Jones R, Papeleux L, Duan C, Vanaerschot A, Daehn G, Duchene L, et al. Machine learning models of plastic flow based on representation theory. npj Comput Mater. 2022;8:172.
https://doi.org/10.1038/s41524-022-00849-8.

Lin P, Ratnam R, Sankari H, Sarker MK, Kumar S, Looper RE, et al. Machine learning-based multi-objective optimization for efficient identification of crystal plasticity model parameters. Comput Methods Appl Mech Eng. 2023;406:115899.
https://doi.org/10.1016/j.cma.2023.115899.

Okabe R, Cheng M, Chotrattanapituk A, Mandal M, Mak K, Buehler MJ. Structural constraint integration in a generative model for the discovery of quantum materials. Nat Mater. 2025;24(3):1-8.
https://doi.org/10.1038/s41563-024-02049-4.

Cheng M, Fu CL, Okabe R, Chotrattanapituk A, Boonkird A, Hung NT, et al. AI-driven materials design: A mini-review. arXiv. 2025. https://doi.org/10.48550/arXiv.2502.02905.

Gao W, Coley CW, Mo Y, Nobuta K, Teo WX, Merchant A, et al. Closing the execution gap in generative AI for chemicals and materials: Freeways or safeguards. ACS Cent Sci. 2024;10(6):1114-28. https://doi.org/10.1021/acscentsci.3c01243.

De Breuck PP, Evans ML, Rignanese GM. Robust model benchmarking and bias-imbalance in data-driven materials science: A case study on MODNet. J Phys Condens Matter. 2021;33(40):404002. https://doi.org/10.1088/1361-648X/ac1c95.

Zhang H, Chen W, Rondinelli JM, Chen W. Mitigating bias in scientific data: A materials science case study. arXiv. 2023. https://doi.org/10.48550/arXiv.2311.01481.

Chotrattanapituk A, Cheng M, Fu CL, Okabe R, Boonkird A, Hung NT, et al. Artificial intelligence-driven approaches for materials design and discovery. arXiv. 2022. https://doi.org/10.48550/arXiv.2211.08934.

Badini S, Regondi S, Pugliese R. Unleashing the power of artificial intelligence in materials design. Materials. 2023;16(17):5927. https://doi.org/10.3390/ma16175927.

Ando Y, Tanaka A, Tsuda K, Katsura Y, Kurosaki K. Effects of data bias on machine-learning-based material discovery using experimental property data. Sci Technol Adv Mater Methods. 2022;2(1):302-9. https://doi.org/10.1080/27660400.2022.2109447.

Zhang Q, Hu Y, Yan J, Xu S, Chen Y, Zhang X. Impact of data bias on machine learning for crystal compound synthesizability predictions. Mach Learn Sci Technol. 2025;6(1):015001. https://doi.org/10.1088/2632-2153/ad9378.

De Breuck PP, Hautier G, Rignanese GM. Materials property prediction for limited datasets enabled by feature selection and joint learning with MODNet. npj Comput Mater. 2021;7:83. https://doi.org/10.1038/s41524-021-00552-2.

Ando Y, Tanaka A, Tsuda K. ET-AL: Entropy-targeted active learning for bias mitigation in materials data. Appl Phys Rev. 2023;10(2):021403. https://doi.org/10.1063/5.0138913.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624:80-5. https://doi.org/10.1038/s41586-023-06735-9.

Chen C, Nguyen DT, Lee SJ, Baker NA, Karakoti AS, Lauw L, et al. Accelerating computational materials discovery with machine learning and cloud high-performance computing: From large-scale screening to experimental validation. J Am Chem Soc. 2024;146(29):20009-18. https://doi.org/10.1021/jacs.4c03849.

Hu M, Tan Q, Knibbe R, Xu M, Jiang B, Wang S, et al. Designing unique and high-performance Al alloys via machine learning: Mitigating data bias through active learning. J Mater Sci Technol. 2023;161:1-12. https://doi.org/10.1016/j.jmst.2023.03.038.

Lin ZJ, Buehler MJ. Simultaneously improving accuracy and computational cost under parametric constraints in materials property prediction tasks. J Cheminform. 2024;16:15. https://doi.org/10.1186/s13321-024-00811-6.

Miret S, McNeely-White D, Blumenstiel J, Lu AX. Generative models for extrapolation prediction in materials informatics. ACS Omega. 2021;6(31):20109-17.
https://doi.org/10.1021/acsomega.1c01716.

Buehler MJ. Generative retrieval-augmented ontologic graph and multiagent strategies for interpretive large language model-based materials design. ACS Eng Au. 2024;4(2):150-63.
https://doi.org/10.1021/acsengineeringau.3c00055.

Zhang Q, Hu Y, Yan J, Xu S, Chen Y, Zhang X. Machine learning-driven material intelligence research and development. Nano Res. 2025;18(1):1-20.
https://doi.org/10.1007/s12274-024-6080-9.

Lin ZJ, Buehler MJ. Unsupervised machine learning prediction of novel 1:3 intermetallic phase with synthesis of TbIr3 (PuNi3-type) as experimental validation. ChemRxiv. 2025.
https://doi.org/10.26434/chemrxiv-2025-cc0dq.

Lin ZJ, Buehler MJ. Tackling data scarcity with transfer learning: A case study of thickness characterization from optical spectra of perovskite thin films. Digit Discov. 2023;2(1):142-54.
https://doi.org/10.1039/D2DD00149G.

Noda K, Wakiuchi A, Yoshida R. Extrapolative prediction of small-data molecular property using quantum mechanics-assisted machine learning. npj Comput Mater. 2023;9:221.
https://doi.org/10.1038/s41524-023-01194-2.

Lin ZJ, Buehler MJ. AI design for high entropy alloys: Enabling discovery, design, and automation. Metals. 2025;15(9):1012.
https://doi.org/10.3390/met15091012.

Lin ZJ, Buehler MJ. Machine learning-driven nanoscale synthesis for electrocatalytic performance: Enabling discovery, design, and closed-loop optimization. Adv Mater. 2025;37(12):e2508263.
https://doi.org/10.1002/adma.202508263.

Lin ZJ, Buehler MJ. Machine learning for accelerating energy materials discovery: Enabling discovery, design, and computational efficiency. Adv Energy Mater. 2025;15(12):e2503356.
https://doi.org/10.1002/aenm.202503356.

Zhang Y, Ling S, Chen W, Buehler MJ, Kaplan DL. Exploring nature’s toolbox: The role of biopolymers in sustainable materials science. Adv Mater. 2025;37(12):e2507822.
https://doi.org/10.1002/adma.202507822.

Zhang H, Song Y, Hou Z, Miret S, Liu B. HoneyComb: A flexible LLM-based agent system for materials science. arXiv. 2024.
https://doi.org/10.48550/arXiv.2409.00135.

Yu S, Ran N, Liu J. Large-language models: The game-changers for materials science research. Artif Intell Chem. 2024;2:100076.
https://doi.org/10.1016/j.aichem.2024.100076.

Buehler MJ. Agentic material science: Enabling discovery, design, and automation. Artif Intell Agents. 2025;1(1):87.
https://doi.org/10.20517/aiagent.2025.87.