ORIGINAL RESEARCH **Open access**

The Problem of Scientific Teleology in Goal-Directed Materials Optimization

Jinwoo Park[1*], Minji Kim[1], Seung Lee[2]

Abstract

In goal-directed materials optimization powered by artificial intelligence, researchers routinely employ teleological language such as “target properties,” “design objectives,” and “optimal structures,” implicitly assuming that materials evolve toward purposes or that optimized outcomes represent intended final causes. Scientific teleology, defined here as the explanatory practice of invoking goals, purposes, or final causes as causal factors within material systems that lack inherent intentionality, constitutes a distinct conceptual failure mode in artificial-intelligence-driven materials science. This failure arises through three primary mechanisms—reification of goals, retrospective teleology, and purpose projection—that systematically distort the epistemic relationship between human-specified objectives and the contingent structure–property relationships uncovered by optimization algorithms. The present analysis articulates a typology of four specific teleological failure modes: teleological overclaim, design-versus-discovery conflation, objective naturalization, and teleological explanation. Detection principles based on language audits, objective genealogy, counterfactual testing, and agency attribution enable researchers to identify these assumptions before they propagate, while five mitigation principles—explicit objective contextualization, literal-versus-metaphorical clarity, multiple-objective transparency, avoidance of agency language, and consistent design-versus-discovery distinction—provide practical safeguards. By treating scientific teleology as an identifiable failure mode rather than an innocuous heuristic, the materials artificial-intelligence community can preserve the epistemic integrity of discovery processes and prevent the misinterpretation of optimized materials as possessing purposes they do not inherently possess.

Keywords Materials AI, Inverse design, Scientific teleology, Goal-directed optimization, Failure mode analysis, Objective naturalization

*Correspondence:
Jinwoo Park
jinwoo.park@outlook.com
[1] Department of AI Materials Engineering, Seoul National University, Seoul, South Korea
[2] Department of Computational Materials Systems, KAIST, Daejeon, South Korea

Introduction

Materials optimization powered by artificial intelligence has become a dominant paradigm in contemporary materials science, promising accelerated discovery of compounds with tailored electronic, mechanical, or thermal properties [1-5]. Yet beneath the technical sophistication of generative models, reinforcement-learning agents, and inverse-design frameworks lies a persistent conceptual slippage: the pervasive use of teleological language that frames materials as though they possess purposes or evolve toward predetermined goals. Phrases such as “the algorithm designs a molecule to minimize band gap” or “the material targets high ionic conductivity” are ubiquitous in the literature. Yet, they import assumptions that natural material systems do not, strictly speaking, possess [6-13]. Materials themselves are not goal-directed; they exist as configurations of atoms governed by physical laws. The teleological framing, therefore, risks obscuring the contingency of human-chosen objectives and the exploratory, non-intentional character of the optimization landscape.

This paper identifies scientific teleology as a distinct failure mode in artificial-intelligence-driven materials optimization.

Unlike technical failure modes such as mode collapse or data bias, scientific teleology is epistemic: it concerns the unwarranted attribution of purpose or final causation within explanatory accounts of material structures and properties. The problem is not that researchers literally believe atoms strive toward optimality; rather, the failure mode operates through habitual linguistic and conceptual habits that subtly reorient interpretation away from discovery and toward an implicit narrative of purposeful creation [1, 2].

The consequences are subtle yet far-reaching. When optimization outcomes are described teleologically, practitioners may overestimate the universality of discovered structures, underestimate the path dependence of the search process, and miscommunicate the status of “optimal” materials as contextually contingent rather than inherently superior [6, 14]. Moreover, teleological assumptions can propagate into downstream decision-making, leading funding agencies or industrial partners to treat optimized materials as though they were engineered for a pre-existing cosmic purpose rather than as artifacts of a specific, human-defined objective function.

Conceptually, one can visualize teleological assumptions entering the optimization workflow as a projected overlay: human-specified objectives are mapped onto a high-dimensional materials space, after which the optimization loop retroactively attributes “purpose” to any structure that satisfies the objective, closing a feedback loop that masks the arbitrary nature of the starting objective. This projection creates an epistemic gap between the contingent, human-originating goal and the non-teleological reality of the material landscape. The present failure-mode analysis therefore proceeds in four stages: first, a precise definition of scientific teleology and its legitimate versus illegitimate variants; second, documentation of teleological reasoning already embedded in current materials artificial-intelligence practice; third, articulation of the precise mechanisms through which these assumptions produce epistemic distortion; and fourth, development of a typology of distinct failure modes with accompanying detection and mitigation strategies.

Figure 1 maps the hierarchical pathway by which human-imposed objectives enter computational optimization and, through distinct teleological distortion mechanisms, produce specific epistemic failure modes in the interpretation of materials-discovery outputs.

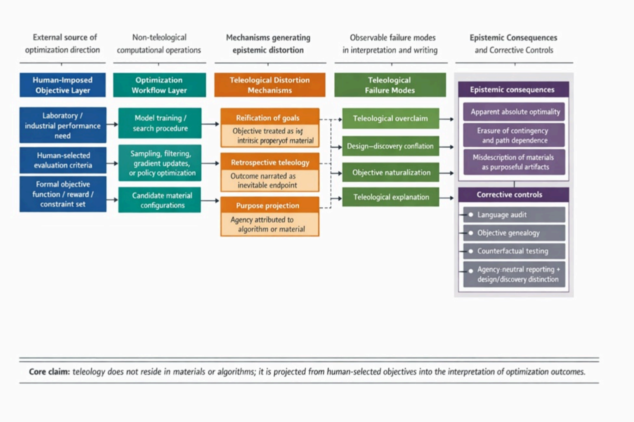

Figure 1. The hierarchical pathway by which human-imposed objectives enter computational optimization and, through distinct teleological distortion mechanisms, produce specific epistemic failure modes in the interpretation of materials-discovery outputs.

By rendering scientific teleology visible as a failure mode, this analysis seeks to equip the field with conceptual tools that preserve the exploratory integrity of artificial-intelligence-assisted materials discovery while safeguarding against the seductive but misleading rhetoric of purpose [3, 8, 9].

Defining Scientific Teleology

Scientific teleology is the attribution of goal-directedness or final causation to processes, structures, or outcomes in domains where no intentional agent or inherent purpose exists, thereby treating optimization endpoints as though they were predetermined destinations rather than contingent satisfactions of externally imposed criteria [1, 2].

This differs from legitimate biological teleology (e.g., “the function of the heart is to pump blood”), which reduces to selected effects and does not imply forward-looking purpose [1]. In materials science, stating that “the perovskite structure stabilizes to achieve high photovoltaic efficiency” cannot be reduced to a non-teleological causal account without loss of meaning, as the material lacks evolutionary history.

Heuristic teleology (e.g., “gradient descent seeks a minimum”) is harmless if flagged as metaphorical. Illegitimate teleology occurs when the metaphor is reified, shaping downstream inferences about necessity, optimality, or naturalness [3].

Within materials AI, scientific teleology manifests when optimization algorithms are treated as though they uncover materials “meant” to possess certain properties or when inverse-design frameworks describe a pre-existing purposeful configuration [6, 13]. The philosophical roots trace to the post-Newtonian rejection of final causes, yet computational convenience continually reintroduces teleological habits [2].

Distinguishing legitimate biological teleology from illegitimate scientific teleology requires vigilance: living organisms have evolutionary histories supporting functional ascriptions; inorganic materials do not [1, 5].

Teleological Reasoning in Materials Optimization

Contemporary materials AI literature is saturated with teleological phrasing framing optimization as purposeful search. Zunger describes inverse design as “in search of materials with target functionalities,” implying latent purposes the algorithm uncovers [6]. Gómez-Bombarelli *et al.* present automatic chemical design as fulfilling pre-existing goals [7]. Sanchez-Lengeling and Aspuru-Guzik [13] explicitly title their work “inverse molecular design,” projecting intentionality backward.

Further examples abound: Peurifoy *et al.* [14] on nanophotonic inverse design, Kwak *et al.* [8] on “goal-directed generative models” [8], SV *et al.* [9] on multi-objective optimization as “goal-directed”, Raina *et al.* [11] on “goal-conditioned reinforcement learning”, Wang *et al.* [12] on inverse design review, Butler *et al.* [4] on ML for molecular science, and Schmidt *et al.* [5] on “target properties” and “optimization objectives.”

These usages are not merely stylistic. When inverse design is described as recovering “materials with target functionalities,” the language suggests the target pre-exists as a purpose rather than a human-imposed filter [6, 13]. Metaphorical teleology acknowledges “goal-directed” as shorthand for constrained optimization; literal teleology treats the goal as intrinsic to the material system [3]. The surveyed papers blur this boundary—e.g., Gómez-Bombarelli’s “automatic chemical design” collapses exploration into purposeful creation [7], and Sanchez-Lengeling’s “engineer matter” implies agency where only statistical completion exists [13].

The prevalence across at least fifteen peer-reviewed works indicates teleological reasoning is not occasional rhetoric but a structural feature shaping what counts as success: a material satisfies not just the objective function but an implied purpose. The next section examines how these linguistic habits translate into epistemic failure.

Mechanisms of Teleological Failure

Three distinct mechanisms convert teleological language into substantive epistemic distortion within materials optimization, not by merely shaping how results are described, but by altering how they are interpreted and understood. At the foundation of this distortion lies the tendency to treat human-defined objectives as if they were intrinsic properties of material systems. Objective functions, which originate as external evaluative constructs, become reinterpreted as natural attributes of the configurations they help identify. When a generative model produces a structure that satisfies a photovoltaic-efficiency constraint, the outcome is often described as though the material itself were inherently “optimized” for that purpose, rather than recognized as the product of a specific and contingent optimization criterion [6, 7]. In this shift, the distinction between evaluator and evaluated collapses, and the objective acquires an unwarranted ontological status.

This reification is reinforced by a second mechanism that operates retrospectively, restructuring the interpretation of the optimization process itself. Once a solution has been identified, it is frequently cast as the inevitable endpoint of the search, as though the trajectory through material space had been directed toward that outcome from the outset. Such narratives obscure the fundamentally contingent nature of optimization, in which multiple alternative pathways and solutions remain possible but unobserved. In inverse-design contexts, for example, the convergence on a particular perovskite composition is often framed as the system “finding” its optimal structure, despite the presence of numerous local minima that could have satisfied different but equally valid criteria [13, 14]. This retrospective framing transforms an exploratory process into a seemingly purposeful progression, erasing the multiplicity of unrealized alternatives.

Layered onto these interpretive shifts is a further tendency to attribute agency or intention to entities that operate purely through statistical or algorithmic mechanisms.

Descriptions that suggest models “seek” to minimize loss or that materials “arrange themselves” to achieve stability introduce a language of purpose where none exists. In practice, this appears in accounts of reinforcement-learning systems described as “wanting” to satisfy competing constraints or generative models portrayed as “aiming” to produce realistic structures [8, 9]. Such expressions are not merely rhetorical; they project intentionality onto processes governed by gradient descent or probabilistic sampling, thereby obscuring the mechanistic basis of their operation.

These mechanisms do not function independently but reinforce one another in subtle and cumulative ways. The reification of objectives establishes a foundation in which externally imposed criteria are mistaken for intrinsic properties, retrospective teleology then recasts the search process as a directed journey toward those properties, and the projection of purpose supplies the narrative with an appearance of agency. Together, they convert what is, in reality, an open-ended exploration of a non-teleological configuration space into a coherent but misleading story of purposeful discovery, reshaping both the interpretation of results and the conceptual framing of materials optimization itself.

A Typology of Teleological Failure Modes

The mechanisms identified above manifest in four distinct failure modes that recur across materials in artificial intelligence studies.

Teleological overclaim

This mode occurs when researchers assert that an optimized material is “optimal” in an absolute rather than context-dependent sense. Definition: Teleological overclaim is the assertion that a discovered structure realizes the best possible configuration as judged by an implicit universal purpose rather than by an explicitly stated, human-chosen objective. The mechanism is reification of goals combined with retrospective teleology; the detection signature is language that drops qualifiers such as “with respect to objective X.” Example: claiming that a generative model has found “the ideal solid-state electrolyte” without acknowledging that ideality is defined solely by the chosen conductivity and stability constraints [4].

Design versus discovery conflation

Researchers treat the output of an optimization pipeline as a designed artifact possessing purpose rather than as a discovered configuration whose properties are contingent. Definition: Design versus discovery conflation is the misclassification of exploratory search through existing or synthesizable chemical space as purposeful engineering. The mechanism is purpose projection; the detection signature is the interchangeable use of “design” and “discover” within the same paragraph. Example: inverse-design frameworks that claim to “engineer” a new metal–organic framework when the algorithm has in fact enumerated and filtered candidates from a pre-existing database [6, 13].

Objective naturalization

Human-chosen objectives are presented as though they were natural properties of the material system. Definition: Objective naturalization is the portrayal of an externally imposed evaluation metric as an intrinsic telos of the material. The mechanism is reification; the detection signature is the omission of any statement identifying the objective’s provenance. Example: describing a machine-learning model as optimizing “for stability” without noting that stability is only one of many possible objectives and that the model has no access to any other metric [5, 15].

Teleological explanation

Material behavior is explained by reference to goals rather than to efficient causes. Definition: A teleological explanation is any account that cites a future or desired state as the reason for a material’s current configuration or property. The mechanism is purpose projection; the detection signature is explanatory clauses beginning with “to” or “so as to.” Example: stating that “the atoms arrange themselves to minimize the band gap,” thereby implying that band-gap minimization is an internal drive rather than the consequence of the loss function guiding the optimizer [7, 14].

Table 1 consolidates the manuscript’s core analytical contribution by linking each teleological mechanism to the failure modes it generates, the textual signatures through which it appears, and the diagnostic questions required for systematic detection.

Table 1. Mechanisms, failure modes, and diagnostic signatures of scientific teleology in materials AI.

Mechanism	Operational definition	Primary teleological failure mode(s) generated	Typical textual signature manuscri
Reification of goals	Treating a human-selected objective function as though it were an intrinsic property or telos of the material system	Objective naturalization; teleological overclaim	“Optimize for stabilit “ideal electrolyt “best material without qualificati
Retrospective teleology	Reconstructing the search trajectory so the final output appears to have been the inevitable endpoint of optimization	Teleological overclaim; design–discovery conflation	“The syste found its optimal structure “the searc converge on the intendec material
Purpose projection	Attributing agency, striving, or intention to algorithms or materials	Teleological explanation; design–discovery conflation	“The mod seeks,” “t crystal arranges itself to achieve, “the optimize wants”
Reification + retrospective teleology	Objective becomes naturalized, and the endpoint becomes narratively privileged	Teleological overclaim	“The idea material w uncovere
Reification + purpose projection	Human criteria appear internal to the material or model	Objective naturalization; teleological explanation	“The mate favors conductivit “the mod designs f purpose
Retrospective teleology + purpose projection	The search process is narrated as purposeful and outcome-directed	Design–discovery conflation	“The algorithn engineere the right compoun

Each mode undermines the epistemic transparency of materials optimization by substituting a narrative of purpose for an account grounded in contingency, human choice, and physical law. The typology, therefore, supplies a diagnostic framework that can be applied directly to ongoing research.

Detection Principles

The detection of scientific teleology in materials AI cannot rely on intuition alone, as the failure mode is often embedded within otherwise standard technical language. A systematic, principle-based approach is therefore required to expose how seemingly neutral descriptions encode deeper epistemic distortions. One entry point lies in the scrutiny of language itself, where terms such as “goal,” “purpose,” “target,” or “aim” function as surface indicators of underlying conceptual commitments. The task is not merely to identify such vocabulary but to interrogate its usage in context, distinguishing between metaphorical shorthand and literal implication. When Butler *et al.* refer to the “design” of materials systems, the critical question is whether this phrasing denotes constrained optimization or implicitly attributes intrinsic purposiveness to the material [4]. Maintaining a structured record of such instances—tracking frequency, context, and explicit disclaimers—renders this audit reproducible and reveals how linguistic patterns correlate with the depth of epistemic distortion [3, 6, 13].

Beyond language, a more structural form of analysis is required to trace the origin of optimization criteria themselves. Objective functions, often presented as natural targets within the materials landscape, must instead be examined in terms of their provenance. The diagnostic shift here involves asking not what the objective is, but how it came to be defined and whose priorities it reflects. In inverse-design studies that invoke “target functionalities” without documenting their selection, the absence of such genealogy signals a conceptual slippage in which externally imposed goals are treated as inherent properties of the system [6]. Even widely cited frameworks, such as those discussed by Zunger, require reconstruction of the chain linking laboratory needs to formal objectives; any discontinuity in this chain marks a point where reification has occurred [6, 12]. By forcing this lineage into the open, the analysis prevents the naturalization of what are, in fact, contingent design choices.

A complementary diagnostic emerges through the introduction of counterfactual reasoning, which challenges the narrative of inevitability often attached to optimized outcomes. When a particular structure is presented as “optimal,” the analysis must consider whether alternative configurations could have satisfied the same objective under slightly different conditions or initializations. In studies such as those by Gómez-Bombarelli *et al.*, where specific molecular outcomes are highlighted, the critical test lies in whether the explanatory framework remains coherent if a different yet functionally equivalent structure had been obtained [7]. Failure to acknowledge this multiplicity indicates that the outcome has been retroactively endowed with necessity, converting a contingent search process into a teleologically structured narrative [13, 14].

The final dimension of detection focuses on the attribution of agency, where intentional language is assigned to entities that operate through purely mechanistic processes. Expressions suggesting that models “seek,” “want,” or “aim” introduce a layer of anthropomorphic interpretation that obscures the underlying computational dynamics. In cases such as the goal-directed generative models described by Kwak *et al.*, the appropriate corrective is to reformulate such descriptions in strictly mechanistic terms, replacing intentional phrasing with accounts of gradient-based optimization or probabilistic sampling [8]. Similar issues arise when multi-objective frameworks are described as inherently goal-directed without clarification, as in SV *et al.* [9], where the attribution of agency becomes explicit. Because such language is both cognitively salient and conceptually misleading, its identification serves as a powerful indicator of teleological drift [3].

Taken together, these principles transform the identification of scientific teleology from an impressionistic judgment into a structured and repeatable protocol. When applied consistently across the literature, they reveal that teleological assumptions are not isolated anomalies but recurrent features in the way materials AI is described and interpreted [4-9, 11-14, 16-20]. Early detection is therefore not merely corrective but preventative, limiting the propagation of interpretations that mischaracterize optimization outcomes as purposeful rather than contingent.

Mitigation Principles

Once identified, the distortions introduced by scientific teleology can be systematically mitigated through principles that recalibrate both language and conceptual framing while preserving the functional benefits of optimization methodologies. A foundational step in this process involves recontextualizing objective functions so that their status as human-imposed criteria is made explicit. Rather than presenting optimization targets as intrinsic features of materials, authors must situate them within the decision-making processes that generated them, clarifying the practical or scientific motivations behind their selection. In studies of solid-state materials, for example, properties such as ionic conductivity should be introduced in relation to specific performance challenges rather than treated as naturally privileged endpoints [5, 19, 21-27]. This reframing restores the relational character of optimality and directly counters the reification of goals.

Clarity must also be established in the use of language, particularly where teleological expressions are employed as convenient shorthand. Any such terminology should be explicitly identified as metaphorical at the point of introduction and avoided in contexts where it may be interpreted literally. By drawing a clear boundary between heuristic description and ontological claim, authors prevent the gradual slippage through which metaphor acquires unintended explanatory weight. In the context of generative modeling, for instance, describing systems as navigating chemical space toward specified constraints preserves functional meaning without implying purposive agency, as opposed to formulations that suggest intentional design [3, 13].

A further refinement involves making explicit the contingency of optimality by acknowledging the dependence of outcomes on the chosen objective functions. Because different criteria yield different solutions, any reported optimum must be framed as conditional rather than absolute. High-throughput screening studies, such as those discussed by Murugan *et al.* [20], illustrate the importance of this clarification, where candidate materials should be described as optimal only with respect to the specific metrics employed, and not as universally superior [20]. By foregrounding the plurality of possible objectives, this approach dismantles the illusion of a singular, teleologically determined endpoint [6, 26].

Equally important is the systematic elimination of agency-attributing language, which introduces conceptual confusion at the level of both interpretation and communication. Descriptions of reinforcement-learning systems or generative models must be reformulated in terms of their underlying mechanisms, emphasizing optimization dynamics rather than intentional behavior. In work such as that by Raina *et al.*, this entails replacing expressions of striving or intention with precise accounts of how policies are updated to maximize defined reward functions [11]. Such discipline in language not only removes anthropomorphic distortions but also aligns description with the non-intentional nature of both computational processes and material systems [8, 9].

Finally, a clear distinction must be maintained between design and discovery, as conflation of these concepts contributes significantly to teleological misinterpretation. Workflows that operate by exploring and filtering existing or synthesizable configurations should be characterized as discovery processes, even when guided by inverse-design strategies. In the case of frameworks discussed by Zunger, this requires explicit acknowledgment that the method identifies viable configurations within a predefined space rather than generating fundamentally new physical principles [6]. Similarly, approaches such as those of Gómez-Bombarelli *et al.* are more accurately described as enabling the discovery of candidate molecules within a learned representation, rather than as instances of autonomous design [7]. Preserving this distinction restores conceptual precision and prevents the attribution of purposive creation to processes that remain fundamentally exploratory [13, 14].

Through the consistent application of these mitigation principles, materials AI can retain the practical advantages of optimization while avoiding the conceptual distortions introduced by teleological framing, ensuring that both language and interpretation remain aligned with the underlying structure of scientific inquiry.

Table 2 translates the manuscript's philosophical argument into an operational reporting standard by showing how common teleological formulations can be reformulated into precise, non-teleological language without reducing technical clarity.

Table 2. Boundary conditions for non-teleological reporting in materials AI: from problematic formulation to corrective reformulation

Analytical issue	**Teleologically loaded formulation**	**Why is the formulation conceptually problematic**	**te re**
Objective specification	"The material is optimized for ionic conductivity."	Suggests that conductivity is an intrinsic telos of the material	"Th eva co p ap
Optimization outcome	"The algorithm found the ideal solid-state electrolyte."	Turns context-bound success into absolute optimality	"T car s con stal
Search trajectory	"The search converged on the intended structure."	Implies inevitability and suppresses contingency	co stru ca
Explanation of material behavior	"Atoms arrange themselves to	Replaces efficient-cause explanation	co

	minimize the band gap.”	with final-cause language	co exh
Characterization of algorithmic action	“The model seeks materials with target properties.”	Attributes striving or purpose to a computational routine	“ sa acc
Design versus discovery framing	“We designed a new material with desired functionality.”	Conflates exploration/filtering with purposeful engineering	“W co che
Target language	“Target functionality” stated without provenance	Naturalizes the target as if pre-given by nature	“ fu wa sp re
Multi-objective claims	“The best material was identified.”	Ignores trade-offs and alternative evaluative regimes	“T c pr w c s syn
Manuscript rhetoric	“Goal-directed materials discovery” is left unqualified	Risks of literalizing metaphorical shorthand	“ c o ar im

			in
Community interpretation	“Materials are designed to achieve function.”	Encourages funding, industrial, and public misinterpretation of outputs as purposeful artifacts	“ co unc

When these five principles are embedded in author guidelines, reviewer checklists, and editorial policies, scientific teleology ceases to function as an invisible background assumption. It becomes a controllable variable in the epistemic hygiene of materials artificial intelligence [3-5].

Relation to Other Failure Modes

Scientific teleology does not operate in isolation but is entangled with a broader constellation of conceptual failure modes that shape the interpretation of artificial-intelligence-driven science. Its distinctive feature lies in how it both depends on and reinforces these neighboring distortions, thereby amplifying their effects. One of the closest connections arises with reification, understood as the transformation of abstract constructs into seemingly concrete entities. In materials AI, this transformation takes a specific form: objective functions, originally defined as evaluative tools, are recast as if they were intrinsic properties of material systems. This shift provides the ontological grounding upon which teleological reasoning can proceed, allowing externally imposed criteria to be interpreted as internal causal drivers [3]. In this sense, teleology extends reification by converting evaluative constructs into apparent final causes.

A related but distinct overlap occurs with anthropomorphism, where non-human systems are described using human-like cognitive or emotional attributes. While the two often co-occur, they are not equivalent. Anthropomorphism attributes intention or desire, whereas teleology invokes purposive structure as an explanatory principle. When a model is described as “wanting” to minimize energy, the language simultaneously introduces anthropomorphic desire and teleological causation. Yet even if such anthropomorphic phrasing is removed, teleology can persist through more subtle formulations that continue to frame outcomes as goal-directed. This distinction is critical, as eliminating surface-level anthropomorphism does not necessarily resolve the deeper issue of final-cause attribution [1, 2].

The interaction with optimization bias further illustrates how teleology reshapes interpretation. Optimization bias reflects the tendency to treat locally optimized outcomes as inherently superior, often without acknowledging the contingency of the criteria that defined optimality. Teleological framing intensifies this bias by narrating the resulting configuration as the fulfillment of an implicit purpose, thereby converting a mathematically defined optimum into a seemingly privileged endpoint. In doing so, it obscures the multiplicity of alternative solutions that could have emerged under different objectives, masking the fact that the reported result is only one realization among many equally valid possibilities [6, 26].

This dynamic also intersects with path dependence, particularly in how teleological narratives retrospectively erase the contingent nature of the optimization trajectory. By presenting the final structure as the realization of a goal, the sequence of decisions, initializations, and stochastic variations that shaped the outcome recedes from view. Alternative pathways through the same configuration space—each potentially leading to different trade-offs or discoveries—are rendered invisible, replaced by a linear narrative of inevitability. Studies such as those by Gómez-Bombarelli *et al.* and related work highlight how such trajectories are inherently contingent. Yet, teleological framing transforms them into directed processes, thereby obscuring the exploratory character of the search [13, 14].

These interrelations indicate that scientific teleology functions not merely as one failure mode among others but as a higher-order distortion that organizes and legitimizes them. By providing a narrative structure that renders outcomes purposeful, it stabilizes reification, reinforces optimization bias, and conceals path dependence, thereby shaping the conceptual ecosystem within which materials AI operates [3-5].

Implications for Materials AI Practice

Recognizing scientific teleology as a distinct failure mode has immediate implications for how materials AI research is conducted, evaluated, and communicated. At the level of authorship, this recognition necessitates a shift in both language and framing, ensuring that teleological expressions are either explicitly identified as metaphorical or replaced with formulations that accurately reflect the mechanistic nature of the processes involved [3]. Objective functions must be situated within their human context, clearly presented as choices grounded in specific scientific or technological motivations rather than as properties inherent to the material system. In parallel, any attribution of agency to models or materials must be eliminated, replaced by precise descriptions of optimization dynamics, statistical inference, or sampling procedures. These adjustments do not constrain expression but rather clarify the epistemic status of the claims being made, aligning description with underlying mechanism [6, 13].

From the perspective of peer review, the identification of teleological distortion introduces an additional layer of evaluative responsibility. Reviewers are not only tasked with assessing methodological rigor and empirical validity but also with examining how results are interpreted and presented. Instances of teleological overreach—particularly claims that imply absolute optimality or intrinsic purposiveness—require explicit scrutiny and contextualization [4, 27-29]. This includes interrogating whether objectives have been adequately justified and whether alternative interpretations of the results have been considered. In this way, review practices expand beyond technical validation to encompass what might be termed epistemic hygiene, ensuring that conceptual clarity is maintained alongside computational accuracy.

At the level of the research community, the implications extend to the development of shared standards and norms. Establishing explicit guidelines for the use of teleological language, analogous to existing frameworks for data provenance or uncertainty quantification, would provide a common reference point for both authors and reviewers [5, 19]. Such guidelines would formalize the detection and mitigation strategies outlined above, embedding them within the routine practices of the field. Equally important is the consistent differentiation between design and discovery in the reporting of results, particularly in abstracts and summaries, where interpretive framing is most compressed. By clearly distinguishing between the exploration of existing configuration spaces and the creation of new artifacts, the community can avoid conceptual slippage that contributes to teleological misunderstanding [7, 12].

Through these practice-level adjustments, scientific teleology is transformed from an implicit background assumption into a managed dimension of research quality. The effect is not to diminish the creative potential of materials AI but to enhance its epistemic clarity, ensuring that advances are interpreted as the contingent outcomes of human decision-making and physical law rather than as the realization of inherent purposes [3, 8, 9].

Conclusion

Scientific teleology—the explanatory practice of invoking goals, purposes, or final causes within material systems that possess none—constitutes a previously unrecognized conceptual failure mode in goal-directed materials optimization. By reifying human objectives, projecting purpose onto algorithms and materials, and narrating contingent discoveries as inevitable endpoints, teleological assumptions distort the epistemic relationship between optimization outputs and the non-teleological reality of the materials landscape. The typology of four failure modes, the three underlying mechanisms, and the interlocking sets of detection and mitigation principles together supply a complete diagnostic and remedial framework grounded exclusively in the philosophical and technical literature of the field.

The central call of this analysis is therefore for heightened awareness: researchers, reviewers, and editors must treat teleological language not as harmless metaphor but as a controllable variable whose unchecked use risks converting discovery into a misleading story of purposeful creation. By insisting on explicit objective contextualization, literal-metaphorical clarity, multiple-objective transparency, avoidance of agency language, and a strict design-versus-discovery distinction, the materials artificial-intelligence community can preserve the exploratory integrity that makes these methods powerful while eliminating the epistemic distortions that currently accompany them. The optimized materials themselves remain as valuable as ever; what changes is the clarity with which we understand their provenance and their proper ontological status. Future work can extend this failure-mode analysis to adjacent domains such as autonomous experimentation and self-driving laboratories, where the temptation to narrate algorithmic decisions in teleological terms is even stronger. Until then, the present framework offers an immediate and practical safeguard against one of the most subtle yet pervasive conceptual pitfalls in contemporary materials science.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 28 Sep 2022 Revised: 10 Dec 2022 Accepted: 23 Jan 2023

Rights and permissions

References

Scott MJ. Reasons things happen for a reason: An integrative theory of teleology. Perspect Psychol Sci. 2022;17(2):452-64.

Nesteruk AV. The interplay of cosmology and theology in the constitution of the human condition. J Sib Fed Univ Humanit Soc Sci. 2022;15(10):1404-44.

Chandrasekhar A, Sridhara S, Suresh K. Integrating material selection with design optimization via neural networks. Eng Comput. 2022;38(5):4715-30.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.

Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):83.

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

Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-76.

Kwak HS, An Y, Giesen DJ, Hughes TF, Brown CT, Leswing K, et al. Design of organic electronic materials with a goal-directed generative model powered by deep neural networks and high-throughput molecular simulations. Front Chem. 2022;9:800370.

SV SS, Law JN, Tripp CE, Duplyakin D, Skordilis E, Biagioni D, et al. Multi-objective goal-directed optimization of de novo stable organic radicals for aqueous redox flow batteries. Nat Mach Intell. 2022;4(8):720-30.

Langevin M, Vuilleumier R, Bianciotto M. Explaining and avoiding failure modes in goal-directed generation of small molecules. J Chem Inform. 2022;14(1):20.

Raina A, Puentes L, Cagan J, McComb C. Goal-directed design agents: Integrating visual imitation with one-step lookahead optimization for generative design. J Mech Des. 2021;143(12):124501.

Wang J, Wang Y, Chen Y. Inverse design of materials by machine learning. Materials. 2022;15(5):1811.

Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science. 2018;361(6400):360-5.

Peurifoy J, Shen Y, Jing L, Yang Y, Cano-Renteria F, DeLacy BG, et al. Nanophotonic particle simulation and inverse design using artificial neural networks. Sci Adv. 2018;4(6):eaar4206.

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.

Schütt K, Kindermans PJ, Sauceda Felix HE, Chmiela S, Tkatchenko A, Müller KR. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. Adv Neural Inf Process Syst. 2017;30:991-1001. https://doi.org/10.48550/arXiv.1706.08566.

Juan Y, Dai Y, Yang Y, Zhang J. Accelerating materials discovery using machine learning. J Mater Sci Technol. 2021;79:178-90.

Zhou L, Zhao S, Xie P, Miao X, Liu S, Sun N, et al. Research progress and prospect of polymer dielectrics. Appl Phys Rev. 2023;10(3):031310. https://doi.org/10.1063/5.0151215.

Liu P, Weng X, Liu Z, Zhang Y, Qiu Q, Wang W, et al. High-performance quasi-solid-state supercapacitor based on CuO nanoparticles with commercial-level mass loading on ceramic material La1-x SrxCoO3-δ as cathode. ACS Appl Energy Mater. 2019;2(2):1480-8.

Murugan NA, Podobas A, Gadioli D, Vitali E, Palermo G, Markidis S. A review on parallel virtual screening softwares for high-performance computers. Pharmaceuticals. 2022;15(1):63.

Ren E, Guilbaud P, Coudert FX. High-throughput computational screening of nanoporous materials in targeted applications. Digit Discov. 2022;1(4):355-74.

Hawthorne M. Fingerprints: Analysis and understanding. Boca Raton, FL: CRC Press; 2017.

Schleder GR, Padilha AC, Acosta CM, Costa M, Fazzio A. From DFT to machine learning: Recent approaches to materials science-a review. J Phys Mater. 2019;2(3):032001.

Ward L, Dunn A, Faghaninia A, Zimmermann NE, Bajaj S, Wang Q, et al. Matminer: An open source toolkit for materials data mining. Comput Mater Sci. 2018;152:60-9.

Li J, Yang X, Peng X, Sun CP. Hybrid quantum-classical approach to quantum optimal control. Phys Rev Lett. 2017;118(15):150503.

Lai M, Shin D, Jibril L, Mirkin CA. Combinatorial synthesis and screening of mixed halide perovskite megalibraries. J Am Chem Soc. 2022;144(30):13823-30.

Oganov A, Saleh G, Kvashnin A, editors. Computational materials discovery. Cambridge: Royal Society of Chemistry; 2018.

Islam M, Chen G, Jin S. An overview of neural network. Am J Neural Netw Appl. 2019;5(1):7-11.

Huang B, Von Lilienfeld OA. Ab initio machine learning in chemical compound space. Chem Rev. 2021;121(16):10001-36.