Eur. Phys. J. Plus (2026) 141:694
https://doi.org/10.1140/epjp/s13360-026-07923-y

THE EUROPEAN PHYSICAL JOURNAL PLUS

Review

The seduction of correlation: on the epistemic limits of AI in materials physics

Luciano Colombo[a]

Department of Physics, University of Cagliari, Cittadella Universitaria, 09042 Monserrato (Ca), Italy

Received: 16 April 2026 / Accepted: 3 June 2026

Abstract The growing use of Artificial Intelligence in materials physics has ignited deep transformations in how knowledge is generated, interpreted, and validated. While data-driven models have demonstrated remarkable predictive capabilities, their increasing centrality raises critical questions about the status of physical understanding when theory is no longer the primary framework of inference. This Perspective explores the epistemic implications of substituting causal, mechanistic explanations with statistical correlations, highlighting the risks of conflating predictive accuracy with scientific insight. Drawing on foundational concepts in the philosophy of science, I argue that the integration of AI into materials theory must be guided not only by efficiency or performance metrics but by a commitment to interpretability, falsifiability, and conceptual coherence. Rather than rejecting AI tools, I advocate for their critical incorporation within physically grounded modeling strategies that preserve the explanatory aims of physics. Only through such a reflective synthesis can the seductive power of correlation be harnessed without compromising the epistemological integrity of the discipline.

1 Outline

One of the key challenges in theoretical materials physics is the development of reliable and predictive models that can unravel the complex, often nonlinear interactions among a multitude of particles, including the many-body effects that are most effectively described in terms of collective excitations, in order to establish accurate structure–property relationships [1–4]. This goal underpins much of modern materials research, where often the ultimate aim is to identify or design materials that exhibit optimal properties for specific technological applications. Fulfilling this agenda requires a deep theoretical understanding of how atomic-scale arrangements, bonding motifs, and electronic configurations collectively determine the materials behavior. Thus, the core theoretical effort lies in generating predictive models that can master the complexity of the fundamental interactions to reveal how structural and compositional parameters govern the emergent physical properties, ultimately enabling targeted discovery and rational design of advanced materials. Overall, this results in a huge computational workload, representing a significant bottleneck in the theoretical exploration of materials. The underlying reason is that traditional methods scale poorly with system size.

This is precisely where alternative Artificial intelligence (AI) computational strategies have begun to reshape the landscape of materials theory. These methods offer a radically different foundational and computational paradigm with respect to the standard theory-based investigation: rather than explicitly solving a governing set of constitutive equations (either classical or quantum), AI models aim at learning statistical representations of structure–property relationships directly from existing data [5–9]. In doing so, they can generalize to predict properties of new, still unpredicted materials with remarkable speed and often acceptable accuracy. Specifically, AI models are trained using datasets of materials properties to generate proxy models capable of approximating outputs obtained from either computer simulations or experiments. The present state of affairs is that AI tools have become able to infer a large number of physical features from given atomic structures, chemical compositions, or topological fingerprints [10, 11]. Beyond property prediction, AI is also increasingly being used to identify hidden patterns and correlations across vast, high-dimensional data spaces. This has proved particularly useful in uncovering descriptors (I so define any possible low-dimensional representation of materials) that correlate strongly with complex functionalities.

The growing reliance on AI-driven data-based strategies in materials research (which is fueled by their high potential and truly astonishing efficiency) raises a number of legitimate concerns that, in my view, deserve serious reflection. Foremost among these is the risk that predictive accuracy may come to overshadow physical understanding, thereby weakening the interpretative and explanatory foundations that have traditionally underpinned the theoretical study of materials. While AI models are often remarkably effective, they tend to abstract away from underlying mechanisms, producing results that may be difficult to rationalize or extrapolate beyond the specific domain on which they were trained. This is a particularly troubling perspective, as it risks undermining the very essence of doing physics.

[a] e-mail: luciano.colombo@dsf.unica.it (corresponding author)

Published online: 16 June 2026

It is within this conceptual framework that the present work fits. My intention is not to reject the use of AI techniques, as doing so would be overly naive and would overlook the many new opportunities they offer for advancing physics research. Rather, I aim at critically assessing their integration into the epistemological fabric of materials physics. In particular, I will argue for the continued centrality of theory-driven discovery: not out of methodological conservatism, but because theoretical modeling remains essential for establishing causality, identifying fundamental descriptors, and formulating generalizable principles that go beyond empirical correlations. The key scientific challenge, therefore, lies in developing a novel balanced methodology that integrates AI-driven inference with physics-based reasoning in a mutually informative way. By doing so, we can ensure that the expanding use of artificial intelligence can enhance, rather than displace, the core aims of theoretical materials science.

The paper is organized as follows. In Sect. 2, I will contrast the traditional theory-driven paradigm of discovery in materials physics with the emerging data-driven approaches introduced by AI, highlighting their differing epistemological commitments. Section 3 explores the conceptual limitations of data-centric methodologies, focusing on issues such as interpretability, falsifiability, and the risk of undermining explanatory depth. In Sect. 4, I examine recent attempts to reconcile AI tools with theoretical reasoning, emphasizing the importance of hybrid models that preserve physical meaning. Section 5 broadens the discussion by reflecting on the nature of scientific understanding itself, particularly the distinction between simulation and genuine conceptual insight. Finally, Sect. 6 draws together these threads to argue for a critical but constructive integration of AI into the fabric of theoretical materials science.

2 Theory-driven vs. data-driven discovery

As an appropriate starting point for the present discussion I recognize that the historical development of materials physics has been predominantly guided by theory-driven discovery, a process anchored in hypothesis formulation, mathematical modeling, and deductive reasoning grounded in established physical laws. Within this intellectual framework, theory has been consistently conceived not merely as just a predictive tool, but as the foundational structure that informs experimental design, supports interpretative analysis, and ultimately enables the articulation of general principles. The top-down nature of this paradigm ensures that new insights thereby generated are logically and coherently integrated into a broader theoretical structure, thus enabling both explanation and generalization. In this framework, understanding has never meant simply forecasting outcomes (such as structure–property relationships), but rather grasping the mechanisms (i.e., the causal, constitutive, and hierarchical relations) dictating the behavior of matter across scales. By contrast, the incorporation of AI-based methods into materials physics introduces a radically different epistemological stance, which I will hereafter refer to as the paradigm of data-driven discovery. Here, discovery does not emerge from physical laws, rather it is extracted from statistical patterns or empirical regularities in large datasets. The process begins with data acquisition, followed by the application of learning algorithms that extract relationships or latent structures capable of making predictions. This is indeed a bottom-up approach.

It must be recognized that AI systems have shown impressive strengths. They are remarkably capable of identifying correlations in complex, high-dimensional datasets, often revealing trends or descriptors that would likely escape conventional analysis. Their capacity to explore vast chemical and structural spaces is both fascinating and practically valuable, especially in accelerating hypothesis generation and screening candidate materials. Under this respect, AI offers tools that can expand our heuristic repertoire and open new paths for discovery. Nonetheless, concerns are raised about the limitations that come with these capabilities. Above all, the issue of interpretability is really critical, whereas theory-based models provide causal narratives grounded in physics principles and laws, AI-driven models in many respects operate through opaque mechanisms. Their internal logic is often inaccessible, making it difficult to understand why a given prediction is made. This lack of transparency undermines one of the fundamental aims of scientific practice: to explain, not merely to describe. What further complicates the situation is the dependency of AI models on the quality and structure of the data they rely on. As a matter of fact, data are never neutral; they reflect the methods, assumptions, and limitations inherent in their generation. As such, the reliability of AI outputs is tightly bound to the fidelity and representativeness of their training sets. Incomplete or biased data may propagate misleading conclusions, especially when models are applied to out-of-sample regimes where physical constraints are absent or inadequately encoded.

None of these concerns should be read as a rejection of AI. On the contrary, I am convinced that these tools have a valuable role to play, and I welcome their integration into materials theory. What I advocate, however, is a cautious and reflective integration, one that respects the distinct epistemic aims of physics. I see great promise in the hybridization of approaches, where AI augments theory rather than displaces it, and where statistical inference is brought into dialogue with physical reasoning. In this perspective, I believe that the role of the physicist (not merely a technical operator of AI models, but a human agent engaged in an intellectual pursuit) is more important than ever. We must not give up our interpretative responsibilities or reduce our role to that of algorithm supervisors. It is essential, in my view, that we remain active interpreters, critical thinkers, and epistemic agents, ensuring that the tools we use serve the goals we set. If we are to avoid a drift toward automation without understanding, we must insist that AI models remain answerable to the conceptual frameworks and explanatory standards of theoretical physics.

3 Epistemic limits of data-driven methods

By now my own mixed feelings (cautious and respectful, but certainly critical) about the massive use of AI methods in theoretical materials physics should have emerged, not so much with regard to their technical aspects (which do really bring numerous operational and computational advantages), but rather with regard to their foundation. I will, therefore, further develop my reasoning by pointing out three different topics that, in my honest opinion, should be carefully considered and placed alongside the technical fascination for AI-based methods.

The first remark lies at the boundary between what should be considered "science" and "non-science." A crucial challenge in this context is establishing the falsifiability of results produced by AI-driven methodologies. Following Karl Popper's criterion of demarcation [12], a scientific theory must be refutable in principle; it must make predictions that, if found false, would lead to its rejection. Quite the opposite, AI-generated results typically emerge from patterns that may lack explicit theoretical underpinning and, therefore, clear falsification criteria. In practice: if, say, a AI tool predicts an outcome that cannot be directly tested or is contingent upon parameter tuning, is it truly falsifiable? The difficulty in applying the Popper's falsifiability criterion still holds when the same issue is approached by the standpoint of Thomas Kuhn's paradigm-based model of scientific progress [13, 14]. Kuhn argued that scientific theories are not immediately discarded when confronted with anomalies; instead, they are evaluated within the broader structure of a scientific paradigm. As he put it, "a scientific theory is declared invalid only if an alternative candidate is available to take its place." This weaker notion of falsifiability suggests that AI-generated models might be provisionally accepted within an existing paradigm, even in the absence of direct empirical refutation, provided they offer just a computationally effective or heuristically valuable problem-solving framework. Then, the risk occurs of a relaxation of falsifiability standards. Does AI-driven physics remain aligned with Popper's vision of a critical, self-correcting scientific enterprise or could it rather drift toward a paradigm where falsifiability is secondary to computational utility?

Next, I believe it is worth remarking that while AI-generated methods typically provide highly effective answers with immediate practical applicability, their long-term validity remains uncertain. Traditional theoretical physics seeks not only to produce results that align with empirical data but also to establish robust explaining frameworks enduring over time. Contrary to this, the AI-driven approach prioritizes optimization over foundational understanding. This distinction becomes crucial when considering the longevity of AI-generated insights. Since its solutions lack the conceptual depth necessary to withstand shifts in scientific paradigms, AI reliance on data-driven heuristics implies that its results typically are context-dependent rather than universally valid. It is instructive to compare the situation under consideration with what occurs in contexts that may appear very distant, yet are subtly similar from a conceptual standpoint. N. Wiener has pointed out that political decisions tend to be shaped by consensus, and economic ones by the pursuit of profit, both governed by the imperative of short-term results [15]. Research in physics, however, should not adopt efficiency as its primary criterion without compromising its very mission. If the emphasis shifts toward rapid predictions and transient utility, theoretical research risks adopting the same short-range logic characterizing both political and economic decision-making. Casting this worry in different terms, the issue of long-term validity ties into the problem of verification: AI may produce results that appear valid within current knowledge constraints, but it fails in welcoming tests against future empirical discoveries. As highlighted by Carnap in his methodological reflections on theoretical concepts, scientific explanation requires not just empirical adequacy but also logical reconstruction through well-defined inferential rules [16]. According to Carnap, therefore, the construction of scientific knowledge is not only an empirical process, but a logically structured one. Scientific concepts and statements should, ideally, be embedded in a framework of well-defined rules and formal relations, such that the steps from data to theory are transparent and logically coherent. From this perspective, even inductive reasoning, which underlies most AI models, should be traceable and amenable to systematic reconstruction within a theoretical language. The concern, therefore, is that many AI applications in materials theory generate predictions without offering this logical transparency or the possibility of integrating their outputs into a unified theoretical scheme. This stands in tension with Carnap's epistemological ideal, where explanation is not merely functional, but logically ordered and communicable. Under this respect, the inability to rationally reconstruct the results produced by many AI models questions their status as true knowledge, notwithstanding their strong empirical success.

Eventually, I observe that AI-driven research operates predominantly as a bottom-up discovery program and, therefore, its data-centric methodology stands in contrast to the more traditional scientific approach, which, following Galileo, combines empirical investigation with a top-down theoretical framework that elevates mathematics as the fundamental language of Nature [17]. Such a Galilean synthesis allowed for the emergence of physical theories that are not merely descriptive, but do have epistemic authority, predictive power, and structural coherence. AI, by contrast, follows an inductive trajectory which aligns more closely with the scientific method proposed by Francis Bacon, who emphasized empirical accumulation of data and the systematic extraction of general laws through observation and experimentation [18]. Bacon rejected abstract reasoning and replaced it with a methodology in which scientific knowledge arises from the meticulous collection of facts, allowing Nature to "speak for itself." The contrast between Galileo and Bacon methods is thus particularly relevant in assessing the epistemic risks of AI-driven physics. Under this respect, I believe we should always keep in mind and learn from the history of science: Galileo method proved to be superior to Bacon one. All the greatest scientific revolutions of physics were achieved not merely by collecting data but by framing them within a robust mathematical framework. Bacon's vision of a purely inductive science never materialized in the same transformative way. This historical example suggests that while AI-driven methodologies can be powerful tools for discovery, their long-term impact will likely depend on their integration with the deeper theoretical structures that have traditionally defined progress in physics.

4 From disruption to dialogue: a possible reframing of AI in materials theory

Building upon the epistemological frameworks elaborated in the previous Sections, a conceptual discontinuity emerges in that AI-based approaches are not derived from, nor necessarily constrained by, fundamental physical principles; instead, they result from data-driven convergence processes that are agnostic to the ontological structure of the systems under investigation. If the success of a model is assessed solely in terms of its empirical output, independent of whether it captures causal mechanisms or respects physical constraints (like e.g., symmetry principles of conservation laws), then the status of such a model becomes ambiguous within the traditional scientific convention. In this perspective, the notion of physical understanding (as distinct from empirical adequacy) may be eroded, and with it the capacity to integrate new findings into a coherent, cumulative framework of knowledge.

Against this background, recent developments in the field have been addressed at realigning the practice of AI-enhanced modeling with the conceptual foundations of physical theory. Two particularly salient examples, namely Physics-Informed Machine Learning (PIML) [19, 20] and Lagrangian Neural Networks (LNNs)[21], can be interpreted as attempts to soften the epistemic rupture by embedding physical content into otherwise data-driven architectures. While these strategies differ in their formulation and scope, they share a common motivation: to mitigate the inherent structural opacity of AI models by introducing constraints that reflect established theoretical commitments.

PIML, in particular, aims to guide the learning process by encoding physical laws into the architecture or the loss function of a model. Constraints derived from symmetry principles, conservation laws, or constitutive equations are integrated not as corrections applied after training, but as theoretically grounded constraints integrated into the procedure from the beginning. From a methodological standpoint, this approach represents a shift from pure data-mining to a form of inductive reasoning that is bounded by domain knowledge. It attempts to reconcile statistical generalization with theoretical plausibility, thus positioning itself as a hybrid epistemology. However, the very notion of "informing" machine learning with physics raises deeper questions. Does the inclusion of a differential constraint or a variational principle genuinely endow the model with a form of physical insight? Or, rather, does it merely steer the optimization landscape toward empirically viable configurations? If the latter, then the role of theory is reduced to that of a heuristic filter rather than an explanatory framework. Moreover, it seems that these constraints work more as engineering expedients than as carriers of conceptual understanding. In other words, they represent just pragmatic strategies for regularization. In this light, the epistemological status of PIML remains conditional: it gestures toward theory but does not necessarily instantiate it.

A more structurally ambitious attempt is offered by LNNs, which aim at discovering the underlying variational structure of a system directly from data. Instead of learning an explicit mapping between inputs and outputs, LNNs are trained to infer a scalar Lagrangian function whose Euler–Lagrange equations govern the observed dynamics. This approach is addressed to bridge the gap between empirical modeling and generative explanation, potentially recovering not just predictive capacity but also the mathematical form of physical laws. In doing so, LNNs evoke the ideal of a model that both fits the data and reveals something of the ontological substrate from which the data arise. And yet, this promise must be carefully qualified. As shown in recent systematic evaluations of AI reasoning models under increasing complexity [22], even the most sophisticated architectures collapse beyond a certain threshold of compositional depth, failing to sustain consistent performance and retracting their reasoning effort when complexity grows. This behavior suggests that the capacity to "learn" physical laws is bounded not only by architectural choices but by the fundamental asymmetry between data interpolation and theory construction.

Such observations reinforce a still cautious, but ultimately more constructive, reframing of the relationship between AI and physical theory. Rather than interpreting AI as an autonomous epistemic agent (which, in fact, it is not!) it is more useful to place it within a dialogical framework: AI methods should serve as instruments that can amplify, test, or refine theoretical intuitions, but cannot autonomously generate them. The asymmetry between computation and understanding is not a flaw to be corrected, but a structural feature of scientific practice that must be respected.

This perspective has broader implications. It resists both technological determinism and methodological purism. It acknowledges that AI may detect patterns inaccessible to traditional analysis, and yet insists that these patterns acquire meaning only within a theoretical framework that specifies their relevance. It values the precision of numerical inference, while preserving the interpretive role of modeling as an act of abstraction, idealization, and integration. It welcomes AI as a co-author in the scientific process, but not as its sovereign.

5 The very essence of scientific understanding

The recent efforts (such as PIML and LNNs) to realign AI-augmented modeling with the conceptual foundations of physical theory provide compelling evidence of an ongoing concerted attempt to soften the epistemic divide between data-driven techniques and theoretical physics. This development is emblematic of a broader shift from Large Language Models toward so-called Large World Models (LWMs), which aim to embed a more comprehensive, causally structured understanding of the world [23–26]. Unlike LLMs, which primarily capture statistical correlations across massive textual or multimodal corpora, LWMs are focused to internalize a causal or probabilistic model of the external world. This capacity is guessed to enable them to simulate, predict, and even potentially approximate certain forms of reasoning about complex phenomena in ways that resemble human-like understanding. Such models,

their proponents argue, could drive the emergence of a new kind of Artificial General Intelligence, one that goes beyond mere statistical fluency and instead requires robust, context-sensitive representations of reality capable of guiding action in novel or ambiguous situations. The evolution from LLMs to LWMs echoes debates in epistemology concerning the nature of understanding: whether it resides solely in symbolic manipulation or necessitates a grounded interpretative engagement with reality. LWMs align with the latter view, positing that genuine comprehension emerges only when a system models the fundamental features of the world, e.g., physical laws or social dynamics. Nevertheless, these advances must be approached once again with caution since, as commented above, even state-of-the-art architectures tend to collapse under escalating compositional complexity, exposing inherent limitations in AI current capacity for genuine reasoning.

Within the conceptual framework developed throughout this paper, the distinction between simulation and genuine interaction proves not merely operational, but constitutive of a deeper epistemological asymmetry. Human scientific understanding arises within a dense web of ontological engagement, where knowledge is not only constructed but in fact lived, that is sedimented through embodied practices and a continuous interplay with the facts of the material world. In contrast, AI operates within a fundamentally disembedded epistemic space. Its modes of inference, however intricate (or, even, physics-based), remain structurally decoupled from anything we could claim to be close to human experience. As a matter of fact, AI (in any of its forms of implementation) processes representations, not presences. It manipulates signs, patterns, and data structures, but remains fundamentally removed from the lived tension between the unknown and its revelation, i.e., from the axis along which genuine (I mean: human) discovery unfolds. While it may excel at formal recombination or, even, reproducing outputs reminiscent of theoretical reasoning, AI performs these tasks without engaging in the interpretive processes that give rise to meaning. In other words, AI representations are not windows onto a world to be unveiled; they are merely rearranged (perhaps even brilliantly) data. In this respect, AI lacks precisely what defines the human scientific endeavor: the transformative relation to reality whereby inquiry becomes not only a method but a mode of being-in-the-world

In this light, it becomes scientifically both untenable and hazardous to consider AI as an autonomous epistemic agent. This would risk overstating what AI can do and overlooking that creativity and understanding still belong to humans. Rather, AI should be conceived as a dialogical instrument, that is a companion to the scientific process that can probe the internal coherence of theories, accelerate pattern recognition, and offer novel constraints to human reasoning, all without displacing the centrality of human judgment. When integrated with epistemic awareness and a critical approach, AI can become a stimulus that challenges the strength of our concepts and the scope of our models, enriching, rather than replacing, the dynamic interplay between imagination, formalism, and the methods that define theoretical physics.

6 Conclusions

In conclusion, what characterizes traditional theoretical physics is the presence of an underlying mechanistic (causal) model. Here, discovery is grounded in the interplay between empirical observation and theoretical abstraction and equations are not merely descriptive tools; they encode principles that aspire to universality, falsifiability, and explanatory depth. In contrast, AI systems operate without a guiding theoretical compass with results just emerging from algorithmic optimization criteria. This epistemic divergence marks a crucial boundary between data-driven and theory-driven research. Bridging these two paradigms is not simply a matter of technical integration: I believe it demands a rigorous reconsideration of what constitutes knowledge in physics. In other words, as AI becomes more deeply embedded in the practice of theoretical physics, it is imperative to engage in such a critical epistemological reflection. More specifically, we must ask whether the growing reliance on AI techniques is marking a shift in our conception of scientific explanation, understanding, and very knowledge. If AI-driven discoveries remain opaque to theoretical interpretation, we may be approaching a point where predictive accuracy replaces explanatory adequacy.

Whether this shift represents a scientific evolution or a rupture in the tradition of theoretical reasoning is a question that, in my opinion, deserves both a rigorous scrutiny and a careful approach. It would be both naive and practically counterproductive to reject AI methods outright, given their capacity. At the same time, an uncritical acceptance of these techniques risks undermining the foundational principles upon which theoretical modeling is built. This challenge requires that AI-generated models be constrained so to ensure that their outputs retain physical significance and interpretability. In this hybrid paradigm, AI should not function as an autonomous substitute for theory, but rather as a complementary instrument, enhancing and, where appropriate, challenging theoretical constructs while remaining subordinate to the explanatory logic that defines physics as a science. The successful assimilation of AI into theoretical practice therefore depends on preserving the centrality of theory as a unifying and organizing principle. Only by enforcing this conceptual coherence can we ensure that the predictive gains offered by AI do not come at the expense of physical understanding.

Acknowledgements I am grateful to prof. Silvano Tagliagambe (*Emeritus*, University of Sassari, Italy) for many illuminating discussions about the epistemological foundation of the theory-driven and data-driven paradigms and for a critical reading of the manuscript. I am as well grateful to Riccardo Dettori (University of Cagliari, Italy) and Claudio Melis (University of Cagliari, Italy) for the many insightful discussions that have significantly enriched the development of this reflection.

Funding Open access funding provided by Università degli Studi di Cagliari within the CRUI-CARE Agreement. This work was supported by the Italian Ministry of University and Research (MUR) under the Italian National Recovery and Resilience Plan (PNRR), Extended Partnership MICS—Made in Italy Circolare e Sostenibile, project SMOS-MOF (ID: D43C22003120001), funded by the European Union—NextGenerationEU.

Data availability This article does not involve the generation or analysis of any dataset. The work presents an epistemological perspective on methodological issues in materials physics, and therefore no research data are associated with this manuscript.

References

1. A. Jain, Y. Shin, K.A. Persson, Computational predictions of energy materials using density functional theory. Nat. Rev. Mater. **1**, 15004 (2016). https://doi.org/10.1038/natrevmats.2015.4
2. F. Giustino, Electron–phonon interactions from first principles. Rev. Mod. Phys. **89**, 015003 (2017). https://doi.org/10.1103/RevModPhys.89.015003
3. N. Marzari, A. Ferretti, C. Wolverton, Electronic-structure methods for materials design. Nat. Mater. **20**, 736–749 (2021). https://doi.org/10.1038/s41563-021-01013-3
4. C.J. Bartel, Review of computational approaches for thermodynamic stability and structure–property relationships in materials. J. Mater. Sci. **57**, 45 (2022). https://doi.org/10.1007/s10853-022-06915-4
5. K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science. Nature **559**, 547–555 (2018). https://doi.org/10.1038/s41586-018-0337-2
6. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová, Machine learning and the physical sciences. Rev. Mod. Phys. **91**, 045002 (2019). https://doi.org/10.1103/RevModPhys.91.045002
7. L. Jiao, X. Song, C. You, X. Liu, L. Li, P. Chen, X. Tang, Z. Feng, F. Liu, Y. Guo, S. Yang, Y. Li, X. Zhang, W. Ma, S. Wang, J. Bai, B. Hou, AI meets physics: a comprehensive survey. Artif. Intell. Rev. **57**, 256 (2024). s10462-024-10874-4
8. V. Stanev, K. Choudhary, A.G. Kusne, J. Paglione, I. Takeuchi, Artificial intelligence for search and discovery of quantum materials. Commun. Mater. **2**, 105 (2021). https://doi.org/10.1038/s43246-021-00209-z
9. X. Zhang, Y. Xiang et al., Representations of materials for machine learning. Annu. Rev. Mater. Res. **53**, 399 (2023). https://doi.org/10.1146/annurev-matsci-080921-085947
10. S. Bai, J. Zhang, X. Wang, Machine-learning for accelerated first-principles prediction of structure-property relationships in materials. Phys. Rev. Mater. **6**, 040301 (2022)
11. T.-S. Vu, M.-Q. Ha, D.-N. Nguyen, Y. Abe, T. Tran, H. Tran, H. Kino, T. Miyake, K. Tsuda, H.-C. Dam, Towards understanding structure–property relations in materials with interpretable deep learning, npj Comput. Mater. **9**, 215 (2023). https://doi.org/10.1038/s41524-023-01163-9
12. T.S. Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press, Chicago, 1962)
13. K.R. Popper, The Logic of Scientific Discovery (1959), (Routledge)
14. K.R. Popper, *Objective Knowledge: An Evolutionary Approach* (Oxford University Press, Oxford, 1972)
15. N. Wiener, *The Human Use of Human Beings: Cybernetics and Society*, rev. ed. with intro. by S. Heims (MIT Press, Cambridge, MA, 1988)
16. R. Carnap, The methodological character of theoretical concepts. J. Symbolic Log. **25**, 71–74 (1958). https://doi.org/10.2307/2964350
17. G. Galilei, *Two New Sciences, trans. with intro. and notes* in ed. by S. Drake (University of Toronto Press, 2008)
18. F. Bacon, *The New Organon*, ed. by L. Jardine and M. Silverthorne (Cambridge University Press, Cambridge, 2000)
19. M. Raissi, P. Perdkaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comp. Phys. **378**, 606 (2019). https://doi.org/10.1016/j.jcp.2018.10.045
20. G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning. Nat. Rev. Phys. **3**, 422 (2021). https://doi.org/10.1038/s42254-021-00314-5
21. M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, S. Ho, Lagrangian Neural Networks, arXiv:2003.04630 (2020). 10.48550/arXiv.2003.04630
22. P. Shojaee, I. Mirzadeh, K. Alizadeh, M. Horton, S. Bengio, M. Farajtabar, The illusion of thinking: understanding the strengths and limitations of reasoning models via the lens of problem complexity, arXiv:2506.06941 (2025). https://doi.org/10.48550/arXiv.2506.06941
23. Z. Yu, J. Ruan, D. Xing, Explainable Reinforcement Learning via a Causal World Model, arXiv:2305.02749 (2023). https://doi.org/10.48550/arXiv.2305.02749
24. J. Gkountouras, M. Lindemann, P. Lippe, E. Gavves, I. Titov, Language agents meet causality: bridging LLMs and causal world models, arXiv:2410.19923 (2024). https://doi.org/10.48550/arXiv.2410.19923
25. R. Saklad, A. Chadha, O. Pavlov, R. Moraffah, Can large language models infer causal relationships from real-world text?, arXiv:2505.18931 (2025). https://doi.org/10.48550/arXiv.2505.18931
26. I. Williams, Can structural correspondences ground real world representational content in Large Language Models?, arXiv:2506.16370 (2025). https://doi.org/10.48550/arXiv.2506.16370