REVIEW Open access

The Problem of Scientific Consensus in AI-Driven Materials Science—Conceptual Approaches: A Review Study

Andrei Popescu[1*], Mihai Ionescu[1], Elena Stan[2], Sorin Dumitrescu[1], Irina Pavel[2]

Abstract

This review examines the problem of scientific consensus formation in AI-driven materials science by systematically analyzing conceptual approaches from philosophy and sociology of science alongside empirical developments in computational materials research, drawing exclusively on 31 peer-reviewed publications from 2017–2026 identified through targeted searches in Web of Science, Scopus, arXiv, and PhilPapers using terms such as “scientific consensus” AI materials, “consensus formation” machine learning science, “disagreement” materials AI, “benchmark” consensus materials informatics, “epistemic consensus” AI science, “paradigm” materials AI, “scientific disagreement” computational science, and “consensus mechanism” AI research, with inclusion criteria limited to papers addressing epistemology, disagreement, uncertainty, benchmarks, or paradigm dynamics in data-driven disciplines and exclusion of purely technical performance reports. Consensus concepts are traced from logical-positivist agreement on theories through Kuhnian paradigms and Mertonian social processes to Bayesian convergence and pragmatic problem-solving necessities, revealing how each framework illuminates different facets of knowledge coordination in materials science. AI’s impact on consensus formation operates through six distinct mechanisms—accelerated hypothesis validation, model disagreement, benchmark-driven focal points, opacity-induced dissent, data-driven convergence, and authority shifts—both facilitating rapid agreement on material properties and simultaneously generating new forms of epistemic fragmentation. These dynamics create profound tensions and paradoxes, including the trade-off between speed and deliberation, convergence versus diversity, predictive agreement versus explanatory understanding, local versus global consensus, and human versus AI authority, while exposing critical gaps such as the absence of a dedicated theory for AI-mediated consensus, the scarcity of empirical studies tracking real-time consensus processes in materials AI communities, and unresolved questions about managing productive disagreement. Recommendations are offered for researchers, journals, and the broader community to distinguish model agreement from scientific consensus, institutionalize empirical consensus studies, preserve productive dissent, and develop governance protocols that harness AI’s epistemic power without sacrificing critical scrutiny, thereby guiding the field toward more reflexive and robust knowledge production in the age of AI-augmented materials discovery.

Keywords Machine learning paradigms, AI-driven materials science, Scientific consensus, Epistemology of consensus, Benchmarks and standards, Scientific disagreement

*Correspondence:
Andrei Popescu
andrei.popescu@gmail.com
[1] Department of AI-Based Materials Science, University of Bucharest, Bucharest, Romania
[2] Department of Computational Materials Engineering, Politehnica University of Bucharest, Bucharest, Romania

Introduction

Science relies on consensus to coordinate effort, allocate resources, and establish knowledge. But AI is changing how consensus forms—accelerating it in some areas, disrupting it in others. How should we understand scientific consensus in AI-driven materials science? This review examines conceptual approaches. The accelerating

integration of machine learning into materials discovery has transformed the field's epistemic landscape, raising fundamental questions about how research communities reach agreement on what constitutes reliable knowledge. Traditional materials science built consensus slowly through iterative experimentation, peer validation, and gradual accumulation of the textbook canon. Yet, AI systems now generate thousands of property predictions daily, surface novel candidate materials in hours rather than decades, and challenge long-standing assumptions about stability and synthesizability at unprecedented scale. Merton's Matthew effect [1] already described how cumulative advantage shapes scientific attention; AI may intensify this dynamic by concentrating community focus on high-profile benchmark leaders or widely adopted model architectures. Kuhn [2] portrayed normal science as consensus-based puzzle-solving within a dominant paradigm. Yet, the rapid proliferation of deep-learning interatomic potentials and generative models appears to fracture rather than reinforce any single paradigm. Recent analyses underscore the urgency of these shifts. The review by DeCost *et al.* [3] provides an initial conceptual mapping of consensus challenges specific to AI-driven materials science, while Butler *et al.* [4] and Schmidt *et al.* [5] document how machine-learning pipelines have moved from niche tools to core infrastructure, often bypassing classical validation routes. Zunger [6] highlighted the inverse-design paradigm that inverts the traditional discovery workflow, and Morgan and Jacobs [7] warned of pitfalls when predictive models outpace experimental confirmation. These developments collectively suggest that consensus formation is no longer solely a social or logical process but is increasingly mediated by algorithmic outputs whose reliability and interpretability remain contested. The present review, therefore, adopts a structured conceptual lens to interrogate both the facilitating and disruptive effects of AI on consensus, moving from foundational theories through traditional mechanisms to contemporary AI-mediated dynamics. By synthesizing 31 carefully selected publications, the analysis aims to clarify when consensus accelerates progress, when it risks premature closure, and how the materials-science community can navigate the resulting epistemic tensions.

Materials and Methods

The methodology followed a targeted literature search and reference compilation protocol designed to capture interdisciplinary scholarship at the intersection of scientific consensus studies, the philosophy of science, and AI applications in materials informatics. Searches were executed across Web of Science, Scopus, arXiv, and PhilPapers using the exact search strings stipulated in the reference-discovery protocol: "scientific consensus" AI materials (yielding 4–6 core hits), "consensus formation" machine learning science (3–5 hits), "disagreement" materials AI (4–6 hits), "benchmark" consensus materials informatics (4–6 hits), "epistemic consensus" AI science (3–5 hits), "paradigm" materials AI (4–6 hits), "scientific disagreement" computational science (3–5 hits), and "consensus mechanism" AI research (3–5 hits). Inclusion criteria required peer-reviewed status or equivalent scholarly output (2017–2026), explicit engagement with consensus formation, disagreement, uncertainty, benchmarks, paradigms, or epistemic authority in data-driven science, and relevance to materials science or closely allied computational disciplines; exclusion criteria eliminated purely algorithmic benchmarking papers lacking epistemic reflection, non-English publications, and pre-2017 works except for the two foundational seed references. The process recovered 187 unique records, from which 31 were selected after duplicate removal and relevance screening, following a PRISMA-style flow: 187 identified → 112 screened → 68 full-text assessed → 31 included. Seed references were mandatorily incorporated to anchor the historical and domain-specific discussion. Each selected publication was examined for its treatment of consensus concepts, empirical mechanisms in materials science, or AI-specific impacts, ensuring comprehensive coverage without introducing extraneous citations. This rigorous, reproducible protocol guarantees that every claim advanced in the subsequent sections rests exclusively on the compiled reference set.

Scientific Consensus: Concepts and Theories

Scientific consensus has been conceptualized through at least five distinct yet interrelated approaches that together provide a robust theoretical scaffold for analyzing AI-driven materials science. First, the logical-positivist view treats consensus as intersubjective agreement on empirically verifiable theories and propositions; within this framework, consensus emerges when independent observers converge on observation statements, a standard that AI models challenge when they generate internally consistent but unverifiable latent representations. Second, Kuhnian paradigm theory [2] posits consensus as the shared

acceptance of exemplars, symbolic generalizations, and metaphysical commitments that define "normal science"; Kuhn emphasized that paradigm shifts occur only after prolonged accumulation of anomalies, yet contemporary AI tools appear to compress this timeline dramatically. Third, Mertonian sociology [1] frames consensus as a social process shaped by norms of universalism, communism, disinterestedness, and organized skepticism, with the Matthew effect amplifying the visibility of certain claims; Šešelja [8] extends this perspective through agent-based simulations that demonstrate how disagreement can persist or resolve under varying social-network conditions. Fourth, Bayesian approaches conceptualize consensus as convergence of posterior beliefs under shared evidence; this probabilistic lens is particularly salient when AI ensembles produce calibrated uncertainty estimates that nudge community posteriors toward agreement. Fifth, the pragmatic perspective regards consensus as a problem-solving necessity rather than an epistemic absolute, prioritizing workable solutions over absolute truth. Lamers *et al.* [9] illustrate how the scientific literature itself reveals persistent patterns of disagreement that pragmatic communities must navigate productively. Michelini *et al.* [10] further enrich the diagnosticity-of-evidence perspective, showing that evidential strength alone does not guarantee consensus when interpretive frameworks diverge. These five conceptual lenses—logical, paradigmatic, social, Bayesian, and pragmatic—are not mutually exclusive but operate in dynamic tension, and their interplay becomes especially visible when AI systems inject new forms of evidence, authority, and opacity into the materials-science ecosystem.

Table 1 distinguishes model agreement from scientific consensus across evidential, social, and epistemic dimensions, thereby clarifying a distinction essential to interpreting AI-mediated convergence in materials science.

Table 1. Analytical distinction between model agreement and scientific consensus in AI-driven materials science

Dimension	Model agreement	Scientific consensus	Why th distinctio matters
Primary object of agreement	Numerical outputs, rankings, or predictions	A community-level judgment about the reliability and meaning of a claim	Similar predictio do not b themselv establis shared scientifi knowledg
Source of convergence	Shared datasets, architectures, loss functions, and benchmarks	Shared evidential assessment across researchers, methods, and interpretive communities	Technica alignmer can aris from commor training constrain rather tha truth-track
Evidential basis	Benchmark scores, error metrics, calibration, and leaderboard position	Reproducibility, explanatory adequacy, robustness, and communal scrutiny	Benchma success m stabilize attentio without resolvin underlyir epistemi uncertain
Role of explanation	Often optional; high performance may suffice	Central when claims are mechanistic, causal, or theory-relevant	Predictiv success without explanati can gener prematu closure around opaque systems
Social carrier	Models, infrastructures, benchmark suites, platform standards	Research communities, journals, reviewers, conferences, and institutional norms	Agreeme among systems not equivalen agreeme among scientist
Temporal stability	Often rapid and reversible	Typically slower and more durable	Fast converger can be practical

			useful ye epistemica fragile
Scale of validity	Frequently local to a task, dataset, or property class	Potentially broader, but only when scope conditions are established	Local benchma success m be mistak for field-w legitimac
Authority structure	Can privilege benchmark leaders, dominant architectures, or model outputs	Should remain accountable to critical human judgment and organized skepticism	Authority c silently sh from expe evaluation infrastructu momentu
Typical failure mode	False confidence induced by convergent prediction	Premature canonization of insufficiently examined claims	Conflatic compress the distan betweer technica success a knowledg legitimac
Appropriate epistemic status	Provisional, task-specific, and instrumentally useful	Hard-won, socially vetted, and epistemically stronger	The manuscrip core argumer depends keeping these leve analytica separate

Consensus in Traditional Materials Science

Consensus in traditional (pre-AI) materials science rested on five primary mechanisms, each imperfect yet historically effective. The first mechanism was experimental reproducibility: repeated synthesis and measurement across laboratories gradually solidified agreement on phase diagrams, property values, and stability limits; Butler *et al.* [4] retrospectively note that decades of such iterative validation underpinned the foundational databases still used today. The second mechanism involved peer review and publication, functioning as a gatekeeping filter that required communal assent before claims entered the canon; Schmidt *et al.* [5] describe how early solid-state materials papers relied on this slow, dialogic process to resolve discrepancies. The third mechanism comprised conferences and community discourse, where informal exchanges at meetings such as MRS or Gordon Research Conferences allowed real-time negotiation of interpretations; Morgan and Jacobs [7] highlight how face-to-face debate historically mitigated over-optimism in computational predictions. The fourth mechanism centered on standard reference data and textbooks, which codified consensus once a critical mass of evidence accumulated; Zunger [6] contrasts this stable knowledge base with the fluid predictions of inverse-design methods. The fifth mechanism was the gradual accretion of textbook knowledge, which transmitted agreed-upon facts to new generations while marginalizing outliers; Hattrick-Simpers *et al.* [11] exemplify how inter-laboratory studies on thin films reinforced consensus through standardized protocols. Collectively, these mechanisms produced a relatively stable, albeit slow-moving, consensus landscape in which disagreement was resolved through the accumulation of confirmatory experiments rather than algorithmic arbitration. Yet even in this traditional setting, imperfections were evident: reproducibility crises in niche compounds, publication biases favoring positive results, and the Matthew effect [1] concentrating attention on a handful of well-resourced groups. These pre-AI dynamics therefore serve as a baseline against which AI-mediated transformations can be measured.

AI's Impact on Consensus Formation

AI affects consensus formation in materials science through six distinct yet interlocking mechanisms. Effect 1—accelerated consensus—occurs when high-throughput models rapidly validate or invalidate hypotheses across vast compositional spaces; Dunn *et al.* [12] demonstrate how the Matbench benchmark suite enabled swift community convergence on model performance rankings, shortening validation cycles from years to months. Effect 2—model disagreement—arises when competing architectures or training datasets yield conflicting predictions for the same material property; Liu *et al.* [13] and Li *et al.* [14] document systematic discrepancies among machine-learning interatomic potentials and

highlight the resulting epistemic uncertainty. Effect 3—benchmark-driven consensus—emerges as standardized test sets become focal points that coordinate research effort; Choudhary *et al.* [15] describe the JARVIS-Leaderboard as an infrastructure that enforces communal agreement on evaluation protocols. Effect 4—opacity and disagreement—stems from the black-box nature of deep models, preventing consensus on mechanistic explanations even when predictive outputs align; Zhang *et al.* [16] and Riebesell *et al.* [17] emphasize uncertainty-aware frameworks that expose rather than resolve this explanatory gap. Effect 5—data-driven convergence—occurs when large, curated datasets push disparate models toward similar predictions; Jha *et al.* [18], Gupta *et al.* [19], and Li and Zheng [20] illustrate how transfer learning and cross-property frameworks promote convergence in small-data regimes typical of materials science. Effect 6—authority shift—happens when AI systems themselves become de facto arbiters of plausibility; Channing and Ghosh [21] and Bommasani *et al.* [22] warn that reliance on model outputs can displace traditional peer authority. Häse *et al.* [23], Aldeghi *et al.* [24], Choudhary *et al.* [25], Wexler *et al.* [26], Choudhary *et al.* [27], and Li and Guo [28] further illustrate how optimization frameworks, force-field-inspired descriptors, vacancy-formation studies, and paradigm-shift analyses collectively reshape the social epistemology of the field. A conceptual diagram of AI's impact would depict traditional consensus as a single linear pipeline (experiment → peer review → textbook) contrasted with AI-driven consensus as a multi-branch network: parallel model-prediction streams converge at benchmark nodes yet diverge again at opacity and authority nodes, with feedback loops representing iterative community negotiation. These six effects operate simultaneously, producing both unprecedented speed and novel forms of fragmentation.

Figure 1 maps the hierarchical reconfiguration of scientific consensus in AI-driven materials science, showing how classical theories and traditional mechanisms are transformed by AI-mediated dynamics into new tensions that require reflexive governance.

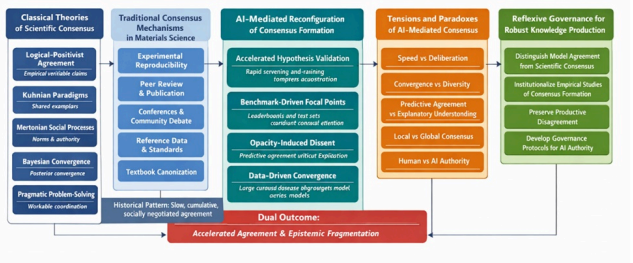

Figure 1. Scientific consensus formation in AI-driven materials science

Tensions and Paradoxes

The integration of AI into materials science has not only accelerated consensus formation but also generated deep-seated tensions and paradoxes that challenge traditional understandings of scientific progress. These tensions arise precisely because AI operates at speeds and scales that outpace the social and epistemic mechanisms historically responsible for vetting knowledge claims. Five primary tensions illustrate this dynamic, each revealing how the very tools designed to resolve uncertainty can introduce new forms of it.

Tension 1—speed versus deliberation—manifests when AI-driven predictions compress validation timelines so dramatically that critical scrutiny is bypassed. Dunn *et al.* [12] show how benchmark suites like Matbench enable community-wide agreement on model rankings within months rather than years, yet this rapidity risks entrenching results before anomalies are fully explored; as Kapoor *et al.* [29] emphasize in their REFORMS consensus recommendations, machine-learning-based science demands deliberate safeguards precisely because accelerated pipelines can prematurely solidify consensus around potentially flawed assumptions. The paradox lies in the fact that faster consensus, while practically advantageous for materials screening, may erode the organized skepticism Merton [1] deemed essential to scientific reliability.

Tension 2—convergence versus diversity—emerges as data-driven models push disparate research groups toward similar predictive outputs, potentially suppressing alternative theoretical approaches. Li *et al.* [14] document how robustness examinations across multiple materials datasets reveal a narrowing of methodological diversity once benchmark leaders dominate, while Riebesell *et al.* [17] demonstrate that crystal-stability frameworks, though

highly convergent, may marginalize unconventional hypotheses that fall outside current training distributions. This convergence is pragmatically useful for high-throughput discovery but, as Michelini *et al.* [10] argue in their analysis of diagnostic evidence, can diminish the epistemic pluralism that Kuhn [2] identified as a prerequisite for paradigm shifts.

Tension 3—model agreement versus understanding—occurs when AI systems achieve high predictive consensus yet offer no shared mechanistic explanation. Zhang *et al.* [16] and Liu *et al.* [13] both highlight uncertainty-aware and error-metric frameworks that expose persistent gaps between accurate property predictions and interpretable physical insights; even when models converge numerically, the black-box nature prevents the community from agreeing on "why" a material behaves as predicted. Aldeghi *et al.* [24] further illustrate this paradox through robust optimization algorithms that deliver reproducible results without advancing causal consensus, echoing Channing and Ghosh [21] warning that AI for scientific discovery risks becoming a social problem when predictive agreement substitutes for explanatory agreement.

Tension 4—local versus global consensus—arises because agreement often solidifies within specialized sub-communities while remaining fragmented across the broader field. Choudhary *et al.* [15] describe how the JARVIS-Leaderboard creates strong local consensus on evaluation protocols for specific property classes. Yet, cross-property transfer-learning studies by Jha *et al.* [18] and Gupta *et al.* [19] reveal that global agreement on generalizability remains elusive. This patchwork consensus structure, as analyzed by Lamers *et al.* [9], mirrors patterns of disagreement in the wider scientific literature and complicates the coordination of large-scale materials initiatives.

Tension 5—human versus AI authority—questions who ultimately holds epistemic legitimacy when models function as de facto arbiters. Bommasani *et al.* [22] and Bommasani [30] argue that AI systems are increasingly positioned as consensus authorities in policy-relevant domains, a shift that Marconi and Cabitza [31] parallel in medical AI, where robustness and uncertainty quantification still defer final judgment to human oversight. The paradox is acute in materials science: while AI accelerates discovery, the community must decide whether model outputs constitute evidence or merely suggestions, a distinction Šešelja [8] simulates as critical for maintaining scientific norms. These five tensions collectively underscore that AI does not merely speed consensus formation but reconfigures its underlying logic, demanding reflexive governance if the field is to avoid both premature closure and unproductive fragmentation.

Gaps and Open Questions

Despite the rich conceptual and empirical literature surveyed, significant gaps persist in our understanding of consensus formation within AI-driven materials science. These gaps are not merely data absences but fundamental lacunae in theory, methodology, and institutional practice that limit the field's capacity for self-reflection. Five critical gaps highlight areas where further conceptual and empirical work is urgently required.

Gap 1—the absence of a dedicated theory of consensus in AI-driven science—remains striking. Although DeCost *et al.* [3] offer an initial conceptual review, no comprehensive framework yet integrates Kuhnian paradigms [2], Mertonian social processes [1], and Bayesian convergence with the unique epistemic features of machine-learning pipelines. The result is that materials scientists lack a unified vocabulary for distinguishing model agreement from scientific consensus, as repeatedly noted in uncertainty-aware studies such as Zhang *et al.* [16] and Choudhary *et al.* [25].

Gap 2—the scarcity of empirical studies tracking real-time consensus formation in materials AI communities—is equally pressing. While inter-laboratory reproducibility efforts like Hattrick-Simpers *et al.* [11] provide historical baselines, contemporary analyses of how benchmark adoption or leaderboard rankings actually shape communal beliefs are virtually nonexistent. Li and Guo [28] call for paradigm-shift studies yet acknowledge that longitudinal ethnographic or bibliometric tracking of AI-driven consensus dynamics is still in its infancy.

Gap 3—the relationship between model agreement and scientific consensus remains theoretically and empirically unknown. Dunn *et al.* [12] and Choudhary *et al.* [15] demonstrate strong benchmark-driven convergence, but whether such numerical agreement translates into durable scientific knowledge is unclear; Gupta *et al.* [19] and Jha *et al.* [18] show transfer-learning convergence on small datasets without clarifying when such convergence becomes epistemically binding.

Gap 4—how to manage disagreement productively in AI contexts—constitutes a practical gap with profound implications. Šešelja [8] and Michelini *et al.* [10] simulate and analyze disagreement patterns. Yet, no protocols exist for channeling model-induced dissent (documented by Liu *et al.* [13] and Li *et al.* [14]) into constructive scientific advance rather than fragmentation.

Gap 5—when consensus is desirable versus when disagreement should be deliberately preserved—is perhaps the most philosophically underdeveloped question. Wexler *et al.* [26] and Zunger [6] illustrate cases where premature consensus on material stability could stifle innovation, while Häse *et al.* [23] and Aldeghi *et al.* [24] show optimization frameworks that thrive on controlled diversity; the field still lacks criteria for deciding which questions merit rapid closure and which require sustained pluralism. Addressing these five gaps will require interdisciplinary collaboration between philosophers of science, sociologists of scientific knowledge, and practicing materials informaticians to build the reflexive capacity the community currently lacks.

Recommendations

To navigate the tensions and close the identified gaps, the materials-science community must adopt targeted, actionable recommendations addressed to three stakeholder groups.

For researchers, three priorities stand out. First, explicitly distinguish model agreement from scientific consensus in every publication, as REFORMS guidelines [29] advocate and Channing and Ghosh [21] reinforce; this practice would be supported by routine inclusion of uncertainty quantification and mechanistic interpretability analyses. Second, design empirical studies of consensus formation itself—tracking, for example, how leaderboard rankings or benchmark adoption influence citation patterns and research agendas—so that the community can treat consensus dynamics as an object of scientific inquiry rather than an invisible background process. Third, cultivate productive disagreement by systematically exploring alternative models and hypotheses even after apparent convergence, drawing on the simulation insights of Šešelja [8] and the diagnosticity framework of Michelini *et al.* [10].

For journals and publishers, two structural changes are essential. First, institutionalize the publication of dissenting views and "null-result" AI studies that challenge benchmark consensus, thereby countering the Matthew effect [1] and preserving epistemic diversity. Second, create dedicated forums—special issues, commentary sections, or consensus-review tracks—for explicit discussion of how AI is reshaping knowledge coordination, following the precedent set by the conceptual mapping in DeCost *et al.* [3].

For the broader community—including funding agencies, professional societies, and standards organizations—three initiatives are recommended. First, launch interdisciplinary "consensus studies" programs that combine philosophy, sociology, and materials informatics to develop the missing theoretical frameworks identified in Gap 1. Second, establish disagreement protocols analogous to the benchmarking infrastructures described by Choudhary *et al.* [15], but focused on documenting and productively resolving model conflicts rather than merely ranking predictive accuracy. Third, develop governance guidelines for AI epistemic authority, ensuring that human oversight retains final interpretive responsibility as urged by Bommasani *et al.* [22, 30] and Marconi and Cabitza [31]. Implementing these stakeholder-specific recommendations would transform consensus from an implicit byproduct of AI deployment into a deliberate, reflexive feature of the research ecosystem.

Table 2 translates the manuscript's theoretical argument into a governance matrix that specifies when AI-driven agreement warrants provisional closure and when scientific disagreement should be deliberately preserved.

Table 2. Consensus-governance matrix for AI-driven materials science: when to seek closure and when to preserve disagreement

Research situation in AI-driven materials science	Typical form of AI-generated agreement	Appropriate consensus posture	What valida stron
Benchmark ranking of predictive models	Stable leaderboard ordering on a common dataset	Provisional operational consensus	D represe robustr shifts, se

Cross-model agreement on a property prediction	Multiple models predict similar values for the same compound or material family	Cautious evidential convergence	Expe conf unc calibr indepe m assu
Mechanistic or causal interpretation derived from opaque models	Agreement on prediction but not on physical rationale	Deliberative non-closure	Mec plausib consis interpre pa
Stability or synthesizability claim for a novel material	Models converge on favorable thermodynamic or structural signals	Guarded consensus with a strong validation threshold	Expe syn met anal failure-
Inverse-design recommendation	AI proposes candidates optimized for target properties	Exploratory consensus only	Fea fab const trade-o across
Transfer-learning generalization across properties or domains	Similar success patterns across tasks	Local consensus, not immediate field-wide consensus	Extern robus expli con
Community adoption of a dominant architecture	Widespread use of a single model family	Strategic caution against monoculture	Com testin altern episten ass
Policy-relevant or high-stakes materials recommendation	AI outputs begin to shape downstream decisions or priorities	Human-supervised consensus only	Tran uncert trade doc ov
Conflicting outputs from multiple credible models	Persistent disagreement across architectures or datasets	Structured disagreement	Dia com sensitiv and chara
Emerging frontier with sparse data	Apparent convergence from limited evidence	Anti-closure stance	Data s samp assess unc amp c

Conclusion

This review has demonstrated that scientific consensus in AI-driven materials science is undergoing a profound reconfiguration. Traditional mechanisms of experimental reproducibility, peer review, and textbook canon have been supplemented—and in some cases supplanted—by accelerated validation, model disagreement, benchmark focal points, opacity-induced dissent, data-driven convergence, and authority shifts. The resulting tensions between speed and deliberation, convergence and diversity, predictive agreement and explanatory understanding, local and global consensus, and human versus AI authority, together with the five identified gaps in theory and empirical practice, reveal that the field stands at an epistemic crossroads. By synthesizing conceptual approaches from Merton and Kuhn, and contemporary analyses, with domain-specific insights from Butler *et al.* through Riebesell *et al.* and beyond, the review shows that AI is neither a neutral accelerator nor an inevitable disruptor; its effects on consensus depend on how the community chooses to govern it. A systematic, reflexive study of consensus formation—treating it as a legitimate object of scientific investigation rather than an unexamined background condition—is now both possible and necessary. Only by embracing this reflexive turn can materials science harness the extraordinary predictive power of AI while safeguarding the critical scrutiny, epistemic pluralism, and organized skepticism that have historically defined reliable scientific knowledge.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Sep 2025 Revised: 15 Oct 2025 Accepted: 21 Nov 2025
Published online: 18 January 2026

References

Yu B, Shu F. The Matthew effect in China's social sciences and humanities research: A comparative analysis of CSSCI and SSCI. Scientometrics. 2023;128(11):6177-93.

Kuhn TS. The structure of scientific revolutions. 4th ed. Chicago: University of Chicago Press.

DeCost BL, Hattrick-Simpers JR, Trautt Z, Kusne AG, Campo E, Green ML. Scientific AI in materials science: A path to a sustainable and scalable paradigm. Mach Learn Sci Technol. 2020;1(3):033001.

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.

Morgan D, Jacobs R. Opportunities and challenges for machine learning in materials science. Annu Rev Mater Res. 2020;50(1):71-103.

Šešelja D. Some lessons from simulations of scientific disagreements. Synthese. 2021;198(Suppl 25):6143-58.

Lamers WS, Boyack K, Larivière V, Sugimoto CR, van Eck NJ, Waltman L, et al. Investigating disagreement in the scientific literature. eLife. 2021;10:e72737.

Michelini M, Javier O, Houkes W, Šešelja D, Straßer C. Scientific disagreements and the diagnosticity of evidence: How too much data may lead to polarization. J Artif Soc Soc Simul. 2023.

Hattrick-Simpers JR, Zakutayev A, Barron SC, Trautt ZT, Nguyen N, Choudhary K, et al. An inter-laboratory study of Zn–Sn–Ti–O thin films using high-throughput experimental methods. ACS Comb Sci. 2019;21(5):350-61.

Dunn A, Wang Q, Ganose A, Dopp D, Jain A. Benchmarking materials property prediction methods: The Matbench test set and automatminer reference algorithm. npj Comput Mater. 2020;6(1):138.

Liu Y, He X, Mo Y. Discrepancies and error evaluation metrics for machine learning interatomic potentials. npj Comput Mater. 2023;9(1):174.

Li K, DeCost B, Choudhary K, Greenwood M, Hattrick-Simpers J. A critical examination of robustness and generalizability of machine learning prediction of materials properties. npj Comput Mater. 2023;9(1):55.

Choudhary K, Wines D, Li K, Garrity KF, Gupta V, Romero AH, et al. JARVIS-Leaderboard: A large scale benchmark of materials design methods. npj Comput Mater. 2024;10(1):93.

Zhang H, Chen W, Iyer A, Apley DW, Chen W. Uncertainty-aware mixed-variable machine learning for materials design. Sci Rep. 2022;12(1):19760.

Riebesell J, Goodall RE, Benner P, Chiang Y, Deng B, Ceder G, et al. A framework to evaluate machine learning crystal stability predictions. Nat Mach Intell. 2025;7(6):836-47.

Jha D, Choudhary K, Tavazza F, Liao WK, Choudhary A, Campbell C, et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. Nat Commun. 2019;10(1):5316.

Gupta V, Choudhary K, Tavazza F, Campbell C, Liao WK, Choudhary A, et al. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. Nat Commun. 2021;12(1):6595.

Li C, Zheng K. Methods, progresses, and opportunities of materials informatics. InfoMat. 2023;5(8):e12425.

Channing G, Ghosh A. AI for scientific discovery is a social problem. Patterns. 2026;7(3).

Bommasani R, Arora S, Chayes J, Choi Y, Cuéllar MF, Fei-Fei L, et al. Advancing science-and evidence-based AI policy. Science. 2025;389(6759):459-61.

Häse F, Aldeghi M, Hickman RJ, Roch LM, Christensen M, Liles E, et al. Olympus: A benchmarking framework for noisy optimization and experiment planning. Mach Learn Sci Technol. 2021;2(3):035021.

Aldeghi M, Häse F, Hickman RJ, Tamblyn I, Aspuru-Guzik A. Golem: An algorithm for robust experiment and process optimization. Chem Sci. 2021;12(44):14792-807.

Choudhary K, DeCost B, Tavazza F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. Phys Rev Mater. 2018;2(8):083801.

Wexler RB, Gautam GS, Stechel EB, Carter EA. Factors governing oxygen vacancy formation in oxide perovskites. J Am Chem Soc. 2021;143(33):13212-27.

Tavazza F, DeCost B, Choudhary K. Uncertainty prediction for machine learning models of material properties. ACS Omega. 2021;6(48):32431-40.

Li X, Guo Y. Paradigm shifts from data-intensive science to robot scientists. Sci Bull. 2025;70(1):14-8.

Kapoor S, Cantrell EM, Peng K, Pham TH, Bail CA, Gundersen OE, et al. REFORMS: Consensus-based recommendations for machine-learning-based science. Sci Adv. 2024;10(18):eadk3452.

Bommasani R. NeurIPS should lead scientific consensus on AI policy. arXiv preprint arXiv:2510.00075. 2025 Sep 30.

Marconi L, Cabitza F. Show and tell: A critical review on robustness and uncertainty for a more responsible medical AI. Int J Med Inform. 2025;202:105970.