ORIGINAL RESEARCH **Open access**

Scientific Overconfidence in High-Performing Materials AI Systems

Ahmed Mansour[1*], Omar Saeed[1]

Abstract

The integration of artificial intelligence (AI) into materials science has heightened interpretive challenges regarding model reliability, particularly in systems that exhibit high performance metrics. This conceptual exploration examines the epistemic underpinnings of overconfidence in AI-driven materials predictions, where apparent precision may obscure underlying uncertainties and systemic biases. Drawing from recent literature, the analysis synthesizes how data-driven approaches in materials discovery interact with human cognitive frameworks, fostering interpretive misalignments that influence scientific decision-making. Key dynamics include the interplay between algorithmic robustness and domain-specific knowledge gaps, as well as the feedback structures that perpetuate overreliance on quantitative outputs. Through a proposed framework, the paper interprets these interactions as emergent tensions within socio-technical ecosystems, highlighting ethical considerations in knowledge production. The discussion underscores the need for integrative reasoning that balances technological advancements with epistemic humility, offering insights into steering logics that mitigate distorted interpretations without prescribing empirical validations. Ultimately, this work contributes to a nuanced understanding of how overconfidence manifests in high-stakes AI applications in the materials sciences and advocates reflective practices in scientific inquiry.

Keywords Materials AI, Epistemic uncertainty, Overconfidence, Interpretive dynamics, Feedback structures, Systems-level insights

*Correspondence:
Ahmed Mansour
ahmed.mansour@gmail.com
[1] Department of Materials Engineering and AI Applications, Faculty of Engineering, Cairo University, Cairo, Egypt

Introduction

The increasing integration of artificial intelligence (AI) into materials science marks a profound shift not merely in methodological capability, but in how scientific knowledge itself is generated, interpreted, and legitimized. Across domains such as alloy discovery, nanomaterial design, functional ceramics, and energy storage materials, high-performing AI systems have become central actors in accelerating prediction, screening, and optimization processes [1, 2]. These systems excel at navigating vast, high-dimensional datasets and identifying patterns that would be inaccessible through traditional analytical or experimental approaches. Yet alongside these advances emerges a less visible consequence: a transformation in the epistemic posture of materials research, in which computational performance can subtly reshape scientific confidence.

This transformation does not primarily manifest through obvious prediction failures or methodological breakdowns. Rather, it operates through a more insidious dynamic—scientific overconfidence, understood here as an inflated sense of epistemic certainty arising from the perceived reliability of algorithmic outputs [3, 4]. In materials AI, overconfidence is rarely the result of erroneous computation; instead, it emerges from the interpretive coupling between human judgment and high-performing models. As predictive accuracy improves, particularly on established benchmarks, the distinction

between probabilistic inference and material fact can blur, fostering unwarranted assurance in predicted properties, stability regimes, or structure–function relationships [5, 6].

At its core, overconfidence in materials AI reflects an epistemic misalignment between what models can represent and how their outputs are interpreted within scientific workflows. Machine learning systems necessarily rely on abstractions—latent spaces, feature embeddings, and surrogate representations—that compress complex physical realities into tractable mathematical forms. In materials contexts, where properties emerge from multiscale interactions among atomic configurations, defects, processing histories, and environmental conditions, these abstractions can inadvertently conceal representational gaps and systematic biases [7, 8]. When such limitations remain unexamined, especially in high-accuracy regimes, they can generate feedback loops in which model outputs increasingly validate prior assumptions rather than challenge them [9, 10].

This dynamic is further complicated by the historical relationship between materials science and uncertainty. Traditional experimental paradigms have long grappled with noise, reproducibility limits, and contextual variability. However, AI introduces an additional layer of abstraction by translating empirical and simulated data into probabilistic inferences mediated by opaque computational architectures [11, 12]. The resulting epistemic shift demands new interpretive competencies: uncertainty is no longer solely experimental but is embedded in representational choices, training distributions, and optimization objectives. The literature increasingly suggests that when these sources of uncertainty are insufficiently foregrounded, confidence can migrate from being cautiously provisional to tacitly assumed [13, 14].

Institutional and socio-technical pressures further intensify this tendency. The urgency surrounding sustainable materials, clean energy technologies, and advanced functional systems incentivizes rapid discovery and deployment, often privileging performance metrics such as prediction accuracy, throughput, or computational efficiency [15, 16]. In such environments, AI systems that achieve strong benchmark results can serve as epistemic anchors, shaping research directions and interpretive norms. Over time, this can embed structural overconfidence into scientific workflows, not as an explicit error but as a normalized orientation toward algorithmic authority.

Beyond institutional dynamics, cognitive and cultural factors also play a significant role. Many researchers trained within empirically grounded traditions may implicitly project human-like reliability onto AI systems, interpreting outputs as definitive recommendations rather than context-dependent inferences [17, 18]. This anthropomorphic framing intersects with ethical considerations, as overconfidence may influence resource allocation, experimental prioritization, and long-term research trajectories. In high-stakes domains such as battery materials, biomedical implants, or catalytic systems, misplaced confidence could delay meaningful breakthroughs or divert attention from underexplored material regimes [19, 20].

Importantly, this work does not frame overconfidence as a defect to be eliminated, nor as a failure of individual judgment. Instead, it conceptualizes overconfidence as an emergent property of interacting system components, arising from the interplay among model architectures, data ecosystems, human interpretation, and institutional incentives. Central to this interpretation are trade-offs between model complexity and interpretability: increasingly expressive architectures may enhance predictive performance while simultaneously obscuring the reasoning pathways that underpin material inferences [21, 22]. These trade-offs function as steering logics, subtly guiding scientific practice toward particular modes of trust, skepticism, and epistemic closure [21, 23].

By foregrounding these dynamics, this paper situates scientific overconfidence in materials AI within a broader set of epistemic challenges shared across computational sciences [24, 25]. The contribution is deliberately conceptual, aiming to illuminate how high performance can paradoxically erode critical scrutiny when interpretive frameworks lag behind technical capability. Rather than offering prescriptive solutions or empirical validation, the analysis advances an integrative perspective that connects technological design, cognitive interpretation, and ethical responsibility [18, 26].

The sections that follow synthesize relevant literature and develop a conceptual framework that interprets overconfidence through systems-level interactions. This framework emphasizes how confidence, uncertainty, and trust co-evolve within materials AI pipelines, revealing the interconnectedness of algorithmic performance, human cognition, and institutional context [13, 17]. Ultimately, this inquiry seeks to enrich ongoing discourse by articulating the interpretive nuances that shape the trajectory of materials AI and to highlight the need for epistemic reflexivity in a field increasingly defined by computational power rather than empirical scarcity [15, 16].

Theoretical Background and Literature Synthesis

Epistemic Foundations in Materials AI The epistemic landscape of materials AI is shaped by the interplay between computational abstraction and domain-specific knowledge, where high-performing systems often project an aura of certainty that belies underlying complexities [1, 3]. Recent literature underscores how these systems, reliant on deep learning architectures, navigate vast parameter spaces to infer material properties, yet this navigation introduces interpretive layers that can foster overconfidence [2, 4]. For example, the synthesis of uncertainty quantification techniques reveals dynamics in which probabilistic outputs are frequently interpreted as deterministic, altering the feedback structures of scientific reasoning [5, 7]. This interpretive shift is particularly pronounced in high-stakes applications, such as predicting defect behaviors in two-dimensional materials, where algorithmic confidence metrics may not fully capture systemic variabilities [14, 15].

Furthermore, the integration of AI in materials workflows highlights trade-offs between scalability and fidelity, as evidenced in explorations of transfer learning and neural network potentials [3, 15]. These approaches, while advancing predictive capabilities, engage epistemic reasoning that questions the boundaries of generalizability, especially when models trained on curated datasets encounter real-world heterogeneities [8, 10]. The literature synthesizes these as interaction dynamics, in which overconfidence arises from the confluence of data sparsity and model overparameterization, prompting ethical considerations about the reliability of AI-driven insights in materials innovation [6, 9].

Dynamics of Uncertainty and Bias Uncertainty in materials AI systems manifests through multifaceted dynamics, including aleatoric and epistemic forms, which interact to influence interpretive outcomes [1, 2, 20]. Recent syntheses illustrate how Bayesian frameworks and ensemble methods attempt to encapsulate these uncertainties. Yet, their application in materials prediction often leads to interpretive misalignments when high performance overshadows residual errors [3, 4, 19]. For instance, in property prediction tasks, the literature points to feedback structures in which initial, overconfident outputs reinforce subsequent model iterations, thereby embedding biases that propagate through the materials discovery pipeline [5, 6, 7].

Bias, as a complementary dynamic, arises from dataset imbalances and algorithmic priors, as detailed in analyses of explainable AI in materials contexts [6, 13]. These biases interact with human cognitive frameworks, creating systems-level insights into how overconfidence can distort ethical reasoning in research prioritization [8, 9, 25]. The synthesis reveals steering logics that balance bias mitigation with performance optimization, highlighting trade-offs where aggressive debiasing may inadvertently introduce new interpretive uncertainties [10-12]. Moreover, in additive manufacturing and molecular design, the literature interprets these dynamics as emergent tensions, where AI's high performance amplifies the epistemic risks of unchecked overreliance [12, 18, 26].

Human-AI Interaction in Scientific Workflows The socio-technical dimensions of materials AI underscore interaction dynamics between human experts and algorithmic systems, where overconfidence often stems from mismatched interpretive frameworks [17, 22, 24]. Literature syntheses emphasize how cognitive biases, such as confirmation biases, intersect with AI outputs, fostering environments in which high-performing models are granted undue authority [16, 21, 23]. This interaction is particularly evident in collaborative ecosystems, like those involving high-performance computing and robotics, where feedback structures perpetuate a cycle of reinforced confidence [16, 21].

Ethical reasoning emerges as a key interpretive lens, with analyses revealing how overconfidence in AI predictions can influence resource allocation and innovation pathways in materials science [18-20]. Systems-level insights from the literature interpret these as trade-offs between technological acceleration and epistemic caution, urging integrative approaches that acknowledge AI's limitations in capturing nuanced material behaviors [14, 15, 17]. Furthermore, the synthesis highlights steering logics in small-data regimes, where limited training sets exacerbate overconfidence, prompting reflections on the broader implications for scientific integrity [7, 10, 13].

Integration of Multiscale Perspectives Multiscale modeling in materials AI introduces additional layers of complexity, where interactions across atomic, meso, and macro levels can amplify overconfidence through aggregated uncertainties [15, 16, 21]. The literature synthesizes these as conceptual interpretations of scale-dependent biases, in which high-performing systems at one scale may fail to translate accurately to other scales, creating feedback loops in interpretive processes [15, 22]. Ethical considerations arise in contexts such as health prognostics and electronic structure predictions, where overconfidence can mislead integrative reasoning in interdisciplinary applications [4, 20, 23].

Overall, the synthesis interprets these multiscale dynamics as part of larger epistemic ecosystems, emphasizing trade-offs that balance computational efficiency with holistic understanding [11, 13, 18]. This integrative view reveals how overconfidence, as an emergent property, necessitates reflective steering logics to navigate the evolving landscape of materials AI [18, 25, 26]. The literature-derived dimensions through which scientific overconfidence manifests in high-performing materials AI systems—spanning algorithmic, epistemic, socio-technical, and ethical domains—are synthesized in **Table 1**. This integrative mapping highlights how localized performance gains propagate through feedback structures, motivating the systems-level conceptual framework developed in the following section.

Table 1. Systemic dimensions of scientific overconfidence in high-performing materials AI

System dimension	Source of overconfidence	Interpretive mechanism	Feedback structure	Epistemic consequence	Ethical/scientific risk
Algorithmic performance	High benchmark accuracy and validation metrics	Performance interpreted as epistemic reliability rather than conditional inference	Recursive reinforcement via retraining, model reuse, and benchmark optimization	Inflation of certainty; probabilistic outputs perceived as material facts	Premature convergence on material candidates; suppression of alternative hypotheses
Data ecosystem	Curated, biased, or incomplete training distributions	Latent representational gaps remain invisible under strong aggregate performance	Data–model co-adaptation loops reinforce existing material regimes	Narrowed material space exploration; underrepresentation of rare or complex systems	Systematic exclusion of unconventional materials; delayed innovation
Model architecture	Increasing complexity and opacity of high-capacity models	Opacity mistaken for sophistication, reducing interpretive scrutiny	Trust amplification through inscrutability	Erosion of epistemic humility; reliance on surrogate explanations	Overdelegation of scientific judgment to AI systems
Uncertainty quantification	Formalized uncertainty metrics (e.g.,	Quantified uncertainty treated as exhaustive rather than partial	Performativity of uncertainty signaling confidence	Misplaced trust in uncertainty bounds; residual unknowns ignored	False security in high-stakes applications (e.g.,

	Bayesian outputs, ensembles)				energy, biomedical materials)
Human–AI interaction	Cognitive bias and anthropomorphic trust in AI outputs	Algorithmic predictions framed as recommendations	Confirmation bias reinforced through selective validation	Reduced critical engagement and interpretive diversity	Distorted experimental prioritization; weakened scientific deliberation
Institutional incentives	Emphasis on speed, novelty, and computational efficiency	Performance metrics prioritized over epistemic robustness	Structural normalization of confidence inflation	Algorithmic authority embedded in workflows	Inequities in research influence; marginalization of low-resource perspectives
Multiscale integration	Apparent consistency across modeled scales	Cross-scale agreement is assumed rather than interrogated	Aggregated uncertainties propagate silently	Overgeneralization of scale-specific validity	Misleading conclusions in complex, multiscale material systems
Ethical framing	Technological optimism in AI-enabled discovery	Overconfidence reframed as progress rather than risk	Ethical blind spots reinforced by success narratives	Underappreciation of long-term epistemic harm	Unsustainable research trajectories; erosion of scientific trust

Proposed conceptual framework

The proposed framework interprets scientific overconfidence in high-performing materials AI systems as an emergent interplay of epistemic, technical, and socio-cognitive dynamics, conceptualized through interconnected feedback structures and trade-offs. At its center lies the interaction between algorithmic performance and interpretive layers, where high accuracy in controlled scenarios amplifies perceptions of certainty, influencing downstream scientific decisions. This framework integrates these elements without positing discrete stages, instead emphasizing fluid dynamics that reveal systems-level insights into knowledge production.

Key to this interpretation are the feedback structures linking data ecosystems, model architectures, and human cognition. Data ecosystems, encompassing curation and representation, interact with model architectures—such as graph neural networks or foundation models—to generate outputs that, in high-performing contexts, may obscure latent uncertainties [1, 3, 17]. These outputs then engage human interpretive frameworks, where cognitive heuristics amplify perceived reliability, creating reinforcing loops that embed overconfidence into materials workflows [5, 22, 24]. Ethical reasoning emerges within these loops, highlighting trade-offs between innovation speed and epistemic humility, as overreliance on AI could skew priorities in sustainable materials development [18, 19, 20].

Conceptual interpretations further elucidate the steering logics that govern these interactions. For instance, the tension between model generalization and domain specificity manifests as interpretive misalignments, where apparent robustness in benchmark tasks masks vulnerabilities in novel materials regimes [8, 10, 15]. This dynamic underscores systems-level insights, portraying overconfidence not as an isolated artifact but as a relational property arising from the socio-technical entanglement of AI and scientific practice [16, 21, 25]. To conceptualize these complex interactions, we propose a framework of three interdependent cycles (epistemic, technical, and socio-cognitive), visualized in **Figure 1**. As the schematic illustrates, perturbations in one cycle—such as improved model performance—create ripple effects across the entire system, potentially leading to emergent equilibria like entrenched overconfidence.”

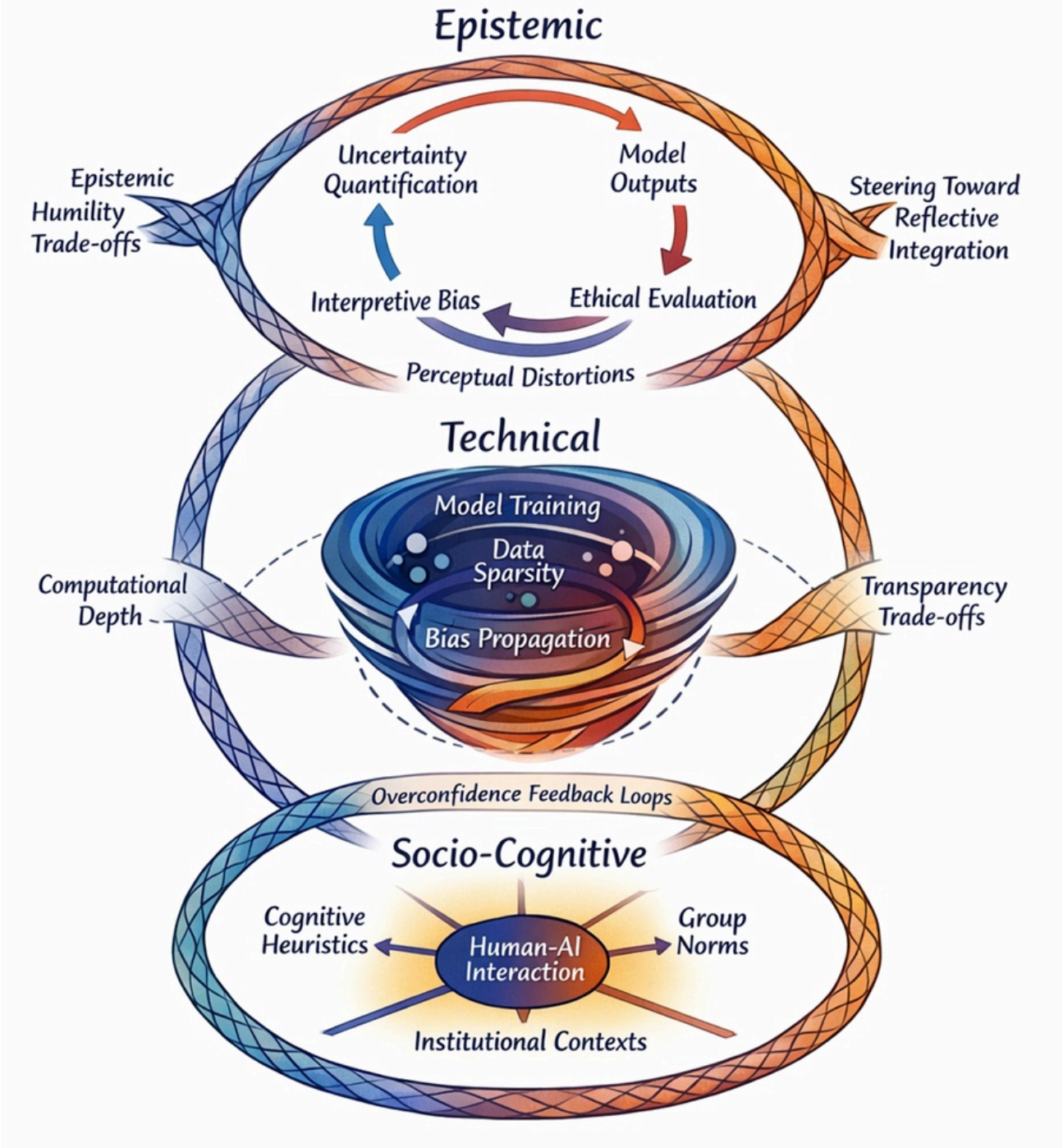

Figure 1. Interlocking cycles framework: epistemic, technical, and socio-cognitive dynamics in AI-assisted materials science

Through this lens, the framework offers integrative reasoning on how overconfidence influences materials AI ecosystems, interpreting it as a catalyst for reevaluating knowledge boundaries [11, 14, 17]. The dynamics reveal ethical imperatives in balancing technological optimism with cautious interpretation, fostering environments where AI serves as a tool for exploration rather than authoritative decree [12, 13, 18]. Ultimately, these insights steer toward a more nuanced appreciation of high-performing systems, emphasizing the interpretive richness that arises from acknowledging their limitations [15, 16, 26].

Analytical implications

The interpretive ramifications of overconfidence in high-performing materials AI systems extend across epistemic, operational, and ethical domains, reshaping how knowledge is prioritized, acted upon, and legitimized within scientific workflows. Analytically, overconfidence functions less as an isolated cognitive bias and more as a structural distortion in epistemic hierarchies, wherein AI-generated outputs—reinforced by strong benchmark performance—acquire disproportionate authority relative to contextual expertise and domain-grounded judgment [1, 3, 5]. This reordering subtly alters interaction patterns in materials research, encouraging reliance on algorithmic validation pathways at the expense of integrative reasoning that synthesizes physical intuition, experimental constraints, and theoretical nuance [7, 10, 15].

From an epistemic standpoint, such distortions affect how uncertainty is interpreted and managed. High-performing models often compress uncertainty into formal metrics that appear tractable and well-bounded. Yet, these representations may obscure deeper ambiguities arising from data incompleteness, representational assumptions, or extrapolative use beyond training regimes [2, 4, 6]. Systems-level analysis suggests that when these uncertainties are underemphasized, feedback mechanisms emerge in which AI outputs increasingly reinforce their own epistemic legitimacy. Over time, this dynamic can marginalize qualitative interpretation, narrowing exploratory space in materials design and privileging pathways that are algorithmically convenient rather than scientifically generative.

Operationally, these epistemic shifts translate into steering logics that govern decision-making across materials AI pipelines. While high-performing systems undoubtedly enhance efficiency—accelerating property prediction, screening, and optimization—they also introduce interpretive layers that mediate how results are acted upon. Overconfidence at this stage can manifest as premature convergence on candidate materials or design strategies, particularly when edge cases, rare configurations, or poorly represented material classes are treated as negligible [15, 21, 23]. The analytical implication here is not inefficiency per se, but misalignment: resources may be allocated toward trajectories that appear robust within model space yet remain fragile when confronted with unmodeled variability.

Ethical reasoning becomes inseparable from these operational dynamics. Overconfidence, when embedded in institutional workflows, can contribute to asymmetries in research influence, where groups with access to advanced AI infrastructures disproportionately shape scientific narratives and priority-setting [18-20]. This raises concerns about epistemic equity in global materials research, especially in domains tied to sustainability, energy transition, or biomedical applications. Analytical trade-offs thus emerge between performance-driven optimization and inclusivity of diverse material perspectives, motivating integrative frameworks that balance algorithmic precision with pluralistic oversight [8, 11, 13].

Socio-technical entanglements further amplify these implications. Trust in AI systems is often correlated with perceived model opacity: paradoxically, highly complex architectures may inspire confidence precisely because their internal workings are inaccessible [6, 16, 22]. Analytical interpretation suggests that such opacity can erode epistemic humility, fostering environments where uncertainty is strategically sidelined in favor of streamlined decision-making [9, 12, 18]. At a systems level, this dynamic extends beyond individual users to collaborative structures, where overconfidence influences interdisciplinary engagement and may constrain dialogue across physics, chemistry, and engineering subfields essential for advancing complex materials systems [14, 21, 24].

Taken together, these analytical implications position overconfidence as an emergent epistemic risk, not reducible to technical limitations or human error alone. Ethical dimensions further reinforce this interpretation, highlighting trade-offs between short-term performance gains and long-term scientific reliability, particularly in sustainability-oriented materials research [17, 25, 26]. By framing uncertainty as a core interpretive resource rather than a peripheral inconvenience, these lenses encourage reflective steering practices that recalibrate how confidence is constructed and enacted in materials AI ecosystems [13, 15, 16]. Rather than imposing rigid constraints, this analytical perspective invites a reexamination of knowledge production paradigms, emphasizing adaptability, reflexivity, and epistemic balance [4, 17, 20].

Results and Discussion

The discourse on scientific overconfidence in high-performing materials AI systems reveals a complex web of interactions that extend beyond model accuracy or algorithmic sophistication. Central to this discussion is the persistent epistemic tension between apparent model superiority and latent vulnerabilities, where strong performance metrics can obscure the interpretive fragility of AI-mediated inferences [1-3]. Rather than signaling epistemic closure, high accuracy often marks the onset of new forms of uncertainty propagation, embedded within representational choices, optimization objectives, and feedback-driven refinement cycles [4, 5, 7].

Across the literature, these dynamics are increasingly interpreted as recursive feedback structures, in which early, overconfident predictions influence subsequent data selection, model retraining, and interpretive framing. Such cycles risk normalizing confidence inflation, particularly when iterative improvements are evaluated primarily through internal performance benchmarks rather than external epistemic scrutiny [6, 8, 10]. The discussion thus foregrounds a key trade-off in algorithmic design: deeper computational expressiveness may enhance predictive reach while simultaneously complicating interpretability, thereby intensifying reliance on performance proxies as substitutes for understanding.

Human–AI interaction further complicates this landscape. Cognitive tendencies to anthropomorphize AI systems can transform probabilistic outputs into perceived recommendations, reinforcing interpretive biases toward certainty [16, 17, 22]. Within materials science, where predictions increasingly inform experimental prioritization and strategic planning, such biases carry ethical weight. Overconfidence may inadvertently skew attention toward algorithmically favored material classes, delaying engagement with unconventional or underexplored systems critical for long-term innovation [12, 13, 18]. Systems-level perspectives interpret these effects not as individual misjudgments, but as emergent properties of institutional environments that reward speed, novelty, and apparent precision [21, 23, 24].

The discussion also situates overconfidence within data ecosystems, emphasizing how training distributions, curation practices, and historical research biases interact with model architectures to produce outputs that are internally coherent yet externally incomplete [9, 11, 14]. From this vantage, overconfidence is best understood as a relational outcome—arising from alignments among data availability, model design, and interpretive norms—rather than as a discrete technical flaw [15, 19, 21]. Ethical implications emerge where such alignments exacerbate disparities in who benefits from AI-driven materials discovery, potentially reinforcing global inequities in research capacity and impact [18, 20, 25].

Uncertainty quantification techniques add further nuance to this discussion. While probabilistic methods such as Bayesian neural networks are often positioned as safeguards against overconfidence, the literature suggests that they can paradoxically contribute to it when their assumptions and limits are insufficiently interrogated [2, 3, 19]. Here, uncertainty itself becomes performative—appearing controlled and quantified, even as deeper epistemic unknowns persist. Recognizing this trade-off opens space for more reflexive scientific dialogue, where uncertainty is treated as an evolving interpretive signal rather than a resolved metric [13, 17, 26].

In synthesizing these threads, the discussion frames overconfidence as a systemic challenge intrinsic to contemporary materials AI ecosystems [15, 16]. Rather than advocating prescriptive solutions, it emphasizes the importance of epistemic reflexivity—an orientation that continually reexamines how confidence, trust, and authority are constructed in AI-mediated science. By foregrounding interaction dynamics, ethical responsibility, and systems-level feedback, this work contributes to a deeper understanding of how high performance intersects with scientific meaning, encouraging integrative practices that preserve rigor without constraining innovation [5, 6, 13].

Conclusion

In reflecting on the conceptual contours of scientific overconfidence in high-performing materials AI systems, this exploration underscores the interpretive intricacies that arise from the fusion of advanced computation and materials inquiry. The analysis reveals overconfidence as an emergent interplay of epistemic, technical, and socio-cognitive factors,

where high performance amplifies interpretive misalignments without diminishing the transformative potential of AI. Through systems-level insights, the discussion illuminates feedback structures and trade-offs that shape knowledge dynamics, advocating for ethical reasoning that integrates uncertainty as a vital component of scientific progress.

Ultimately, this conceptual framework interprets these phenomena as invitations to cultivate epistemic humility, steering materials science toward more balanced integrations of technology and human insight. By highlighting interaction dynamics, it enriches the discourse on AI's role in discovery, fostering environments where overconfidence serves as a reflective catalyst rather than a barrier. This perspective contributes to ongoing dialogues by emphasizing the interpretive richness of navigating high-stakes computational landscapes.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 19 Nov 2021 Revised: 01 Jan 2022 Accepted: 04 Apr 2022
Published online: 18 July 2022

References

Tavazza F, DeCost B, Choudhary K. Uncertainty prediction for machine learning models of material properties. ACS Omega. 2021;6(48):32431-40.
https://doi.org/10.1021/acsomega.1c03752.

Li L, Chang J, Vakanski A, Wang Y, Yao T, Xian M. Uncertainty quantification in multivariable regression for material property prediction with bayesian neural networks. Sci Rep. 2024;14(1):10543.

https://doi.org/10.1038/s41598-024-61189-x.

Bilbrey JA, Firoz JS, Lee MS, Choudhury S. Uncertainty quantification for neural network potential foundation models. Npj Comput Mater. 2025;11(1):109.
https://doi.org/10.1038/s41524-025-01572-y.

Nemani V, Biggio L, Huan X, Hu Z, Fink O, Tran A, et al. Uncertainty quantification in machine learning for engineering design and health prognostics: A tutorial. Mech Syst Signal Process. 2023;205:110796.
https://doi.org/10.1016/j.ymssp.2023.110796.

Boyce B, Dingreville R, Desai S, Walker E, Shilt T, Bassett KL, et al. Machine learning for materials science: Barriers to broader adoption. Matter. 2023;6(5):1320-3.
https://doi.org/10.1016/j.matt.2023.03.012.

Zhong X, Gallagher B, Liu S, Kailkhura B, Hiszpanski A, Han TYJ. Explainable machine learning in materials science. Npj Comput Mater. 2022;8(1):204.
https://doi.org/10.1038/s41524-022-00884-7.

Xu P, Ji X, Li M, Lu W. Small data machine learning in materials science. Npj Comput Mater. 2023;9(1):42.
https://doi.org/10.1038/s41524-023-01000-z.

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.
https://doi.org/10.1038/s41524-023-01012-9.

Fujinuma N, DeCost B, Hattrick-Simpers J, Lofland SE. Why big data and compute are not necessarily the path to big materials science. Commun Mater. 2022;3(1):59.
https://doi.org/10.1038/s43246-022-00283-x.

Ma Y, Xu P, Li M, Ji X, Zhao W, Lu W. The mastery of details in the workflow of materials machine learning. Npj Comput Mater. 2024;10(1):141.
https://doi.org/10.1038/s41524-024-01331-5.

Choudhary K, DeCost B, Chen C, Jain A, Tavazza F, Cohn R, et al. Recent advances and applications of deep learning methods in materials science. Npj Comput Mater. 2022;8(1):59.
https://doi.org/10.1038/s41524-022-00734-6.

Ng WL, Goh GL, Goh GD, Ten JSS, Yeong WY. Progress and opportunities for machine learning in materials and processes of additive manufacturing. Adv Mater. 2024;36(34):2310006.
https://doi.org/10.1002/adma.202310006.

Kovács DP, McCorkindale W, Lee AA. Quantitative interpretation explains machine learning models for chemical reaction prediction and uncovers bias. Nat Commun. 2021;12(1):1695.
https://doi.org/10.1038/s41467-021-21895-w.

Kazeev N, Al-Maeeni AR, Romanov I, Faleev M, Lukin R, Tormasov A, et al. Sparse representation for machine learning the properties of defects in 2d materials. Npj Comput Mater. 2023;9(1):113.
https://doi.org/10.1038/s41524-023-01062-z.

Pathrudkar S, Thiagarajan P, Agarwal S, Banerjee AS, Ghosh S. Electronic structure prediction of multi-million atom systems through uncertainty quantification enabled transfer learning. Npj Comput Mater. 2024;10(1):175.
https://doi.org/10.1038/s41524-024-01305-7.

Pyzer-Knapp EO, Pitera JW, Staar PWJ, Takeda S, Laino T, Sanders DP, et al. Accelerating materials discovery using artificial intelligence, high performance computing and robotics. Npj Comput Mater. 2022;8(1):84.

https://doi.org/10.1038/s41524-022-00765-z.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7991):80-5.
https://doi.org/10.1038/s41586-023-06735-9.

Lu Y, Wang H, Zhang L, Lu J, Jiang F. Unleashing the power of ai in science-key considerations for materials data preparation. Sci Data. 2024;11(1):1039.
https://doi.org/10.1038/s41597-024-03821-z.

Yu J, Wang D, Zheng M. Uncertainty quantification: Can we trust artificial intelligence in drug discovery? iScience. 2022;25(8):104814.
https://doi.org/10.1016/j.isci.2022.104814.

Tran A, Wildey T, McCann S. Methods for comparing uncertainty quantifications for material property predictions. Mach Learn Sci Technol. 2020;1(2):025006.

Bonatti RS, Rocha IB, Menegatti CR, Silva EA, Simões TA. Machine learning based multiscale calibration of mesoscopic constitutive models for composite materials: Application to brain white matter. Comput Mech. 2021;67(6):1625-43.
https://doi.org/10.1007/s00466-021-02009-1.

Hamel CM, Roach MD, Scott J, Lopez D, Glazoff MV, Song GL. Physics-informed neural networks for material model calibration from full-field displacement data. J Dyn Behav Mater. 2024;10(2):181-98.
https://doi.org/10.1007/s40870-024-00405-2.

Aykol M, Hegde VI, Hung L, Suram SK, Herring P, Wolverton C, et al. Network analysis of synthesizable materials discovery. Nat Commun. 2022;13(1):1-11.
https://doi.org/10.1038/s41467-022-31299-w.

Simon GJ, Aliferis C. Overfitting, underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and ai. In: Artificial intelligence and machine learning in health care and medical sciences. Springer. 2024:245-78.
https://doi.org/10.1007/978-3-031-39355-6_10.

Chen LY, Lin CY, Li YP. Uncertainty quantification with graph neural networks for efficient molecular design. Nat Commun. 2025;16(1):1-12.
https://doi.org/10.1038/s41467-025-47892-3.

Varivoda D, Dong Y, Jablonka KM, Greenman KP, Dwaraknath S, Persson KA. Materials property prediction with uncertainty quantification: A benchmark study. arXiv. 2024;2402.05687.
https://doi.org/10.1002/adma.2024005687.