ORIGINAL RESEARCH **Open access**

A Conceptual Framework for Anticipating Second-Order Effects of Materials AI Deployment

Khaled Mahfouz[1*], Rania Abdelaziz[1], Sherif Adel[2], Dina Mostafa[1], Tamer Nabil[2], Reem Saad[1]

Abstract

The deployment of artificial intelligence (AI) in materials science has overwhelmingly prioritized first-order effects—such as accelerated property predictions, high-throughput screening of candidate compounds, and the discovery of novel materials with targeted functionalities—while systematically neglecting second-order effects that arise indirectly from transformations in research practices, institutional incentives, and community norms. Second-order effects are defined here as the consequences of AI adoption that emerge not from the immediate technical outputs of models but from the behavioral, structural, and epistemic shifts these outputs induce among researchers, laboratories, funding bodies, and publishing ecosystems. This framework identifies six principal types of second-order effects (epistemic, behavioral, institutional, social, normative, and ecological). It delineates four mechanisms through which first-order successes propagate into these indirect outcomes, including attention reallocation, success amplification, skill substitution, and self-reinforcing feedback loops. It then proposes a five-component anticipation framework—baseline mapping, intervention specification, causal pathway mapping, stakeholder analysis, and scenario development—that equips materials AI practitioners to foresee and mitigate such effects before large-scale deployment. By embedding foresight into the innovation pipeline, the framework advances responsible materials AI practices that safeguard the long-term integrity, equity, and epistemic robustness of the field, ensuring that technological gains do not inadvertently undermine the very scientific ecosystem they seek to enhance. Ultimately, proactive anticipation of second-order effects will allow the materials community to harness AI's transformative power while preserving the diversity of inquiry, the balance between computation and experiment, and the human-centered values that have historically driven discovery.

Keywords Materials AI, Responsible innovation, Second-order effects, Unintended consequences, Technology assessment, Epistemic impacts

*Correspondence:
Khaled Mahfouz
khaled.mahfouz@gmail.com
[1] Department of Intelligent Materials Systems, Mansoura University, Mansoura, Egypt
[2] Department of Computational Materials Analytics, Helwan University, Cairo, Egypt

Introduction

Materials artificial intelligence has rapidly transitioned from a niche computational tool to a core driver of discovery, promising faster predictions, novel materials, and dramatically increased research throughput. Foundational works have cataloged the remarkable direct benefits: machine learning algorithms now routinely predict molecular and solid-state properties with unprecedented speed and scale, enabling inverse design strategies that invert the traditional Edisonian trial-and-error paradigm. Yet this narrow focus on first-order technical performance—model accuracy, discovery rate, and experimental validation throughput—has left the field largely blind to the second-order effects that may ultimately prove more consequential for the trajectory of materials science itself. These indirect consequences arise not from any single AI

output but from the subtle and cumulative ways in which AI reshapes researcher behavior, institutional priorities, collaboration networks, and even the epistemic standards that define credible knowledge within the discipline [1-5].

The prevailing literature in artificial intelligence for materials science remains dominated by demonstrations of first-order efficacy. Butler *et al.* [2] outlined how machine learning can accelerate molecular and materials discovery across diverse classes of compounds, emphasizing predictive power and data-driven insights. Subsequent reviews chronicled recent advances in solid-state applications and the promise of autonomous research platforms that integrate computation, robotics, and experimentation. Such contributions rightly celebrate the potential for closed-loop discovery systems capable of operating at speeds unattainable by human researchers alone. However, these accounts rarely pause to interrogate what happens after the first-order successes accumulate: how the very presence of powerful AI tools begins to reconfigure what problems are deemed worth pursuing, which skills are valued in hiring and promotion, which research outputs receive citations and funding, and which forms of evidence come to dominate peer review [6-10].

The present work contends that these second-order effects are not peripheral externalities but central features of AI deployment that demand explicit anticipation. Ignoring those risks amplifies unintended distortions — ranging from narrowed research agendas to eroded trust in non-AI methodologies— that could undermine the field's long-term vitality. Drawing on broader scholarship in technology assessment and the sociology of science, the paper argues that purposive technological action in materials AI carries unanticipated consequences that ripple far beyond immediate performance metrics. This conceptual framework, therefore, supplies a structured lens for foresight. It moves systematically from problem identification through definition, typology, propagation mechanisms, and an actionable anticipation architecture. By doing so, it fills a critical gap. While materials AI researchers excel at forecasting material properties, they have yet to develop comparable foresight for forecasting the systemic impacts of their own tools. The framework offered here is intended as both a scholarly contribution and a practical instrument, equipping authors, reviewers, funders, and community leaders to embed second-order anticipation into the earliest stages of AI system design and deployment. In an era when AI is poised to become infrastructure rather than experiment, such foresight is no longer optional but essential for responsible stewardship of the materials research enterprise.

Figure 1 presents the hierarchical propagation structure through which first-order AI outputs translate into second-order effects and ultimately reshape the materials science ecosystem.

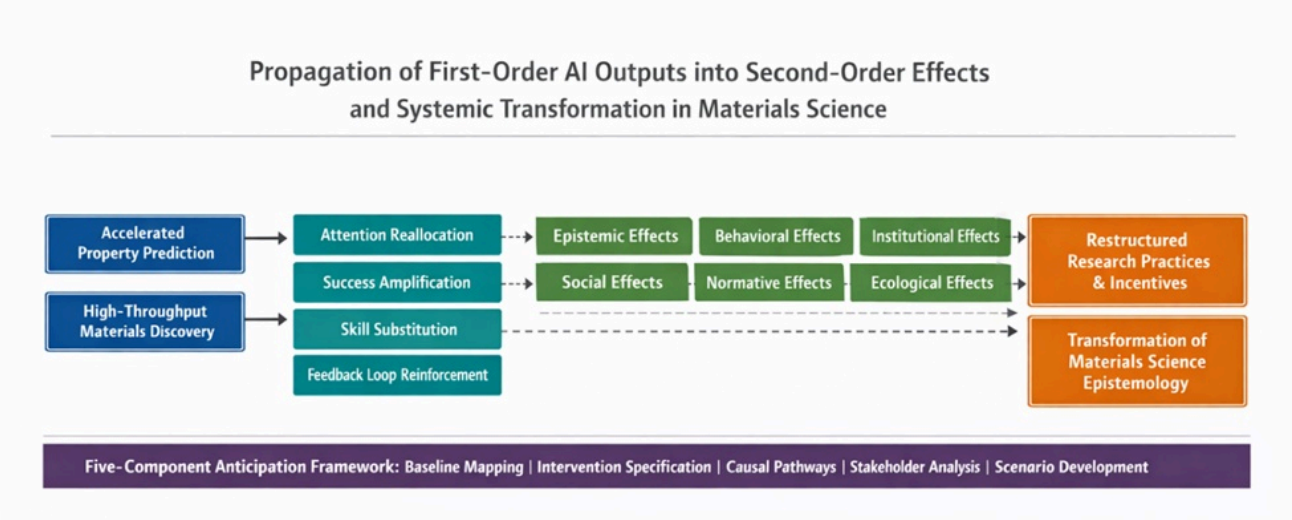

Figure 1. The hierarchical propagation structure through which first-order AI outputs translate into second-order effects and ultimately reshape the materials science ecosystem.

Defining Second-Order Effects

Second-order effects are consequences of AI deployment that arise not directly from AI outputs (such as a predicted crystal structure or property value) but from the changes in human behavior, research priorities, institutional structures, or community norms induced by the routine use of those outputs.

This definition distinguishes three hierarchical layers of impact. First-order effects are the direct, intended results of AI application: a model correctly classifies a material's band gap, an autonomous platform proposes a new synthesis route, or a generative system enumerates thousands of hypothetical compounds. These effects are typically measured by metrics of accuracy, speed, or novelty and constitute the explicit value proposition of materials AI. Second-order effects, by contrast, emerge one step removed: once researchers internalize the reliability of AI predictions, they may reallocate effort away from labor-intensive experimental validation toward prompt engineering and model fine-tuning; once funding agencies observe accelerated discovery pipelines, they may preferentially support AI-centric proposals; once journals begin to publish AI-generated hypotheses at scale, peer-review standards may quietly shift toward computational plausibility rather than empirical corroboration [11-14].

Third-order effects represent still broader systemic transformations: the cumulative second-order changes may ultimately redefine the identity of materials science itself, tilting the discipline toward computational fluency as the dominant form of expertise, reshaping graduate training curricula, and altering the very epistemology of what counts as a "contribution" to the field. The distinction is not merely taxonomic; it carries normative weight. First-order successes are visible and attributable, lending themselves to celebration in grant reports and press releases. Second- and third-order effects are emergent, distributed across many actors, and often visible only in retrospect—precisely the terrain that Merton famously mapped in his classic analysis of the unanticipated consequences of purposive social action [15-19].

Table 1 clarifies the conceptual distinction between first, second, and third-order effects, highlighting their differing temporal scales, visibility, and governance challenges.

Table 1. Conceptual differentiation of first, second, and third-order effects in materials AI

Dimension	First-order effects	Second-order effects	Third-order effects
Definition	Direct technical outputs of AI systems	Indirect consequences arising from behavioral, institutional, and epistemic shifts	System-wide transformations emerging from accumulated second-order changes
Temporal emergence	Immediate	Medium-term	Long-term
Visibility	High (measurable and attributable)	Moderate (diffuse and distributed)	Low (emergent and systemic)
Primary actors	AI systems and developers	Researchers, institutions, journals, and funders	Entire scientific ecosystem
Mechanism of formation	Model performance and outputs	Propagation via attention, incentives, and skill shifts	Recursive accumulation and structural lock-in
Evaluation metrics	Accuracy, speed, and discovery rate	Behavioral change, funding patterns, and publication norms	Epistemic transformation and disciplinary identity
Example (materials AI)	Predicting band gaps or generating compounds	Shift toward AI-friendly research problems	Redefinition of materials science as a computational discipline

Governance challenge	Optimization	Anticipation and monitoring	System-level stewardship

In the materials AI context, the stakes are concrete. An AI system that excels at screening metal-organic frameworks (first-order) may inadvertently train an entire cohort of early-career researchers to favor problems amenable to high-throughput computation while devaluing exploratory synthesis in under-studied chemical spaces (second-order). Over time, the community may come to view “materials discovery” as synonymous with database querying rather than laboratory intuition (third-order). These layers interact recursively; a second-order behavioral shift can itself become the input for further institutional adaptation. The framework, therefore, treats second-order effects not as rare anomalies but as predictable outcomes of the feedback between technology and its human and organizational context. By making this layered causality explicit, the definition supplies the conceptual foundation upon which typology, mechanisms, and anticipation strategies are subsequently built.

Types of Second-Order Effects

Six distinct yet interrelated types of second-order effects characterize the materials AI landscape. Each type is defined below, illustrated with materials-specific examples, and evaluated for its importance to the field’s future.

Epistemic Effects concern alterations in what the community accepts as knowledge, evidence, or valid understanding. When AI-generated predictions achieve publication without exhaustive experimental follow-up, the boundary between simulation and empirical fact may blur; researchers may begin to treat high-confidence model outputs as provisional truth, subtly elevating computational plausibility over traditional reproducibility standards. Such shifts matter because they risk eroding the empirical grounding that has historically distinguished materials science from purely theoretical domains [20-24].

Behavioral Effects encompass changes in individual researcher practices, attention allocation, and daily workflows. With AI tools delivering rapid candidate lists, investigators may instinctively gravitate toward problems that maximize model performance metrics, investing less time in hypothesis generation grounded in chemical intuition or in the painstaking characterization of outliers. The result is a gradual deskilling in certain experimental competencies and a corresponding over-reliance on prompt-crafting expertise.

Institutional Effects involve modifications to funding priorities, publication gatekeeping, hiring criteria, and promotion pathways. Grant panels may increasingly reward proposals that promise AI integration, while tenure committees begin to weigh AI-assisted publication volume more heavily than depth of mechanistic insight. Over time, these incentives can lock in path-dependent resource allocation that privileges certain subfields at the expense of others.

Social Effects pertain to transformations in collaboration patterns, community structure, and equity. AI platforms that require specialized coding or data-science fluency may concentrate influence among researchers with computational backgrounds, potentially marginalizing experimentalists or investigators from institutions with limited high-performance computing access. Collaboration networks may tighten around a small set of well-resourced AI hubs, altering the social topography of the discipline.

Normative Effects reflect evolving values, research priorities, and ethical commitments. As AI success stories proliferate, the community may implicitly adopt efficiency and scalability as overriding virtues, downgrading the intrinsic value of curiosity-driven exploration or long-term fundamental inquiry. Ethical discourse may narrow to bias mitigation within models while overlooking broader questions of environmental cost or dual-use potential in materials design.

Ecological Effects describe changes in the broader research ecosystem—journals, conferences, databases, and infrastructure. The surge in AI-derived manuscripts may prompt journals to create dedicated tracks or special issues, while conference programs tilt toward computational sessions; shared materials databases may become optimized for

machine readability at the expense of human interpretability. These infrastructural adaptations, once entrenched, become self-perpetuating features of the scientific environment.

Each type is consequential because it operates at a different scale, yet collectively reshapes the conditions under which future materials innovation occurs. Anticipating them requires moving beyond isolated case studies to a systematic mapping of their interdependencies [25-29].

Mechanisms of Effect Propagation

First-order AI successes do not automatically produce second-order effects; they propagate through four identifiable mechanisms that convert technical performance into behavioral, institutional, and epistemic change. Attention Reallocation occurs when the visibility and ease of AI-assisted tasks redirect the researcher's focus. A model that rapidly screens thousands of candidates makes those candidates cognitively salient, drawing time and resources away from less tractable problems that resist automation. In materials science, this can manifest as a preference for high-throughput virtual libraries over niche experimental systems whose properties lie outside current training distributions.

Success amplification arises because AI-generated results tend to accumulate citations, media attention, and follow-on funding more readily than traditional work. Early demonstrations of accelerated discovery create a Matthew effect: well-publicized AI milestones attract further investment, reinforcing the perceived superiority of the approach and marginalizing alternative methodologies even when they retain unique strengths. Skill Substitution takes place as AI systems assume roles previously performed by human experts, thereby altering the value placed on particular competencies. Routine property prediction or structure generation once required deep domain knowledge; once delegated to models, that knowledge may atrophy in the next generation of trainees, while new skills in model interpretation and uncertainty quantification rise in prestige. Feedback loops complete the propagation cycle when first-order outputs elicit community responses that themselves become inputs for further AI development or deployment. For example, an increase in AI-assisted publications generates more training data, which improves future models, which in turn encourages even greater reliance on AI—creating a self-reinforcing cycle whose second-order consequences (narrowed problem portfolios, homogenized research questions) may go unnoticed until deeply embedded.

The effect propagation pathway can be conceptualized as a sequential diagram: first-order effects (rectangles labeled "accelerated predictions" and "novel discoveries") feed rightward through arrows labeled with the four mechanisms into second-order effect clusters (ovals representing epistemic shifts, behavioral changes, etc.), which then cascade into third-order systemic outcomes (a large encompassing circle labeled "transformed materials science ecosystem"). Arrows also loop backward from second-order nodes to first-order inputs, illustrating the recursive nature of the process. This visual heuristic underscores that propagation is neither linear nor inevitable but contingent on the specific deployment context, making anticipation both possible and necessary.

A Framework for Anticipation

To move from recognition of second-order effects to practical foresight, this paper proposes a five-component conceptual framework specifically tailored to materials AI deployment. The framework is iterative and prospective, designed to be applied before, during, and after system rollout.

Baseline Mapping requires a systematic documentation of the pre-AI research landscape, including current problem-selection patterns, publication norms, collaboration networks, funding distributions, and epistemic standards. By establishing a clear snapshot of existing practices—drawing on bibliometric data, survey instruments, and stakeholder interviews—researchers create the reference point against which future deviations can be detected.

Intervention Specification involves a detailed characterization of the proposed AI system and its intended integration into workflows. This includes not only technical architecture but also the human–AI division of labor, data governance policies, and planned points of interaction with experimental pipelines. Clarity at this stage prevents vague claims of "AI assistance" from obscuring the precise loci where second-order propagation may begin. Causal Pathway Mapping traces plausible chains from first-order outputs through the four propagation mechanisms identified earlier to potential second- and third-order consequences. For each pathway, participants articulate boundary conditions, amplification factors, and early-warning indicators, transforming abstract risks into concrete, monitorable scenarios.

Stakeholder Analysis identifies all parties who may experience or influence second-order effects—principal investigators, graduate students, funding officers, journal editors, industry partners, and underrepresented groups—and maps their differential exposure and agency. This step ensures that anticipation efforts do not inadvertently privilege the perspectives of well-resourced AI developers. Scenario Development generates multiple plausible futures by combining outputs from the preceding components into narrative vignettes. Workshops or structured exercises produce optimistic, baseline, and pessimistic scenarios that highlight leverage points for mitigation and reveal hidden trade-offs. These scenarios serve as boundary objects for community dialogue and decision-making.

Collectively, the five components form a coherent architecture that renders second-order anticipation an explicit, repeatable practice rather than an afterthought. When applied diligently, the framework converts the abstract warning of unintended consequences into an operational capability, positioning materials AI as a domain that leads rather than lags in responsible technology governance.

Anticipation Strategies

The five-component anticipation framework gains practical traction through a suite of targeted strategies calibrated to each of the six types of second-order effects. These strategies are not generic checklists but context-sensitive practices that draw explicitly on baseline mapping, intervention specification, causal pathway mapping, stakeholder analysis, and scenario development. When enacted iteratively, they convert foresight into ongoing governance, allowing materials AI deployment to remain responsive to emergent ripple effects rather than reactive to entrenched distortions.

As summarized in **Table 2**, each second-order effect type is linked to specific propagation mechanisms, observable indicators, and targeted anticipation strategies."

Table 2. Mapping of second-order effect types to propagation mechanisms and anticipation strategies

Second-order effect type	Primary mechanism(s)	Observable indicator	Associated risk	Anticipation strategy
Epistemic	Success amplification; feedback loops	Increase in AI-only publications	Erosion of empirical validation norms	Bibliometric monitoring; contribution redefinition forums
Behavioral	Attention reallocation; skill substitution	Shift in researcher time allocation	Deskilling in experimental methods	Time-use surveys; problem-selection tracking
Institutional	Success amplification	Funding and hiring bias toward AI	Path dependency in resource allocation	Funding audits; hiring criteria reviews
Social	Attention reallocation	Concentration of collaborations in AI hubs	Inequality and marginalization	Network mapping; equity dashboards

Normative	Success amplification; feedback loops	Emphasis on efficiency over curiosity	Value drift in research priorities	Value audits; discourse tracking
Ecological	Feedback loops	Growth of AI-focused journals and databases	Infrastructure lock-in	Ecosystem mapping; database usage analytics

For epistemic effects, two complementary strategies stand out. First, practitioners should monitor changing publication norms by instituting annual bibliometric reviews that quantify shifts in the ratio of AI-only versus experimentally corroborated claims across high-impact journals in materials science. As Chubb and colleagues observed when examining AI's integration into broader research workflows [8], such monitoring reveals subtle redefinitions of evidentiary thresholds long before they become normalized. This strategy anchors in component 1 (baseline mapping) to fix pre-deployment publication patterns and component 3 (causal pathway mapping) to trace deviations back to specific model deployments. Second, research teams should track what counts as a "contribution" via structured community forums in which diverse stakeholders periodically renegotiate credit allocation between computational prediction and traditional mechanistic insight. These deliberations, informed by component 4 (stakeholder analysis), safeguard against the quiet elevation of model plausibility over empirical depth.

For behavioral effects, two strategies focus on the observable researcher conduct. First, regular surveys of time allocation —administered at project milestones—can quantify how researchers redistribute effort between prompt engineering, model interpretation, and hands-on experimentation once AI tools are embedded. Drawing on the feedback-loop dynamics highlighted by Glickman and Sharot in human–AI interaction studies [7], such surveys expose attention reallocation before it hardens into habit. The practice integrates directly with Component 2 (Intervention Specification) by documenting intended versus actual workflow divisions. Second, longitudinal tracking of problem-selection patterns through project registries or grant databases reveals whether AI availability steers investigators toward easily automatable chemical spaces and away from chemically intractable but scientifically rich domains. This approach leverages component 5 (Scenario Development) to test alternative futures and intervene with targeted incentives for exploratory work.

For institutional effects, anticipation requires vigilance over resource flows. First, funding-pattern analysis—conducted by comparing pre- and post-AI award distributions—can flag whether panels increasingly favor proposals that foreground machine-learning components. As Clarke and Whittlestone noted in their survey of AI's long-term impacts on science [9], such shifts can lock in path dependence unless detected early. The strategy aligns with Component 3 by mapping causal pathways from first-order success metrics to grant criteria. Second, hiring-criteria audits, performed by reviewing job advertisements and promotion rubrics every two years, ensure that computational fluency does not eclipse experimental expertise in evaluation matrices. Component 4 (Stakeholder analysis) ensures that early-career and experimentally oriented voices shape these audits.

For social effects, two network-oriented strategies are essential. First, collaboration-network mapping using co-authorship and funding graphs can detect whether AI platforms concentrate influence among computationally resourced groups. Building on the equity concerns raised in technology-assessment literature [12], this mapping flags emerging divides before they widen. It draws on component 1 to establish baseline diversity metrics. Second, equity-metric dashboards—tracking participation rates of underrepresented institutions in AI-driven materials consortia—provide early signals of marginalization. These dashboards, updated via Component 5 scenario exercises, support corrective actions such as targeted training or shared infrastructure.

For normative effects, value audits and discourse tracking form the core. First, periodic value audits—facilitated through anonymous surveys or ethics workshops—can surface whether efficiency and scalability are displacing curiosity-driven priorities in research statements and grant narratives. As Hagendorff reflected on the limits of machine-learning research norms [25], such audits prevent value drift from remaining tacit. The strategy links to component 2 by specifying normative guardrails at deployment. Second, tracking ethical discourse in conference proceedings and editorial boards

reveals whether discussions narrow to model bias while sidelining broader questions of environmental cost or dual-use materials. Component 3 causal mapping ties these shifts to specific AI success stories.

For ecological effects, ecosystem mapping and usage tracking close the loop. First, journal-and-conference ecosystem maps—updated biannually—can reveal whether dedicated AI tracks or special issues are reshaping submission volumes and session allocations. Informed by DeCost *et al.* [22] vision of sustainable scientific AI paradigms, these maps prevent infrastructural lock-in. They rely on Component 1 baseline data. Second, database-usage analytics can monitor whether shared materials repositories are being optimized exclusively for machine readability at the expense of human interpretability. Component 5 scenarios, then test mitigation pathways such as dual-interface designs.

Collectively, these twelve strategies—two per effect type—render the anticipation framework actionable. They are deliberately lightweight, scalable across laboratory, institutional, and community levels, and explicitly tied to the five components so that foresight remains systematic rather than ad hoc.

Relation to Existing Concepts

This conceptual framework does not emerge in isolation but extends and operationalizes several foundational ideas in the sociology of science, technology studies, and responsible innovation. It first builds directly upon Merton's seminal analysis of the unanticipated consequences of purposive social action [29]. Where Merton demonstrated that intentional interventions routinely generate unforeseen outcomes through complex social mechanisms, the present framework translates that abstract insight into a materials-AI-specific architecture. It supplies the missing operational layer—baseline mapping, causal pathways, and scenario development—that allows practitioners to anticipate rather than merely observe those consequences after the fact.

The framework also aligns closely with technology-assessment methodologies. Alami and co-authors, in their examination of artificial intelligence within health technology assessment [12], underscored the necessity of evaluating indirect complexity before deployment. The five-component structure advanced here adapts that evaluative discipline to materials science, where the stakes include not only clinical outcomes but the epistemic and institutional integrity of an entire discovery pipeline. By requiring explicit intervention specification and stakeholder analysis, the framework prevents the narrow first-order optimism that has historically characterized technology assessments in emerging domains.

Furthermore, the framework resonates with the principles of responsible innovation. Csernovszky *et al.* [20] articulated a system-level impact assessment for artificial intelligence and sustainability, emphasizing that innovation must proactively address ripple effects across ecological and social dimensions. The six-type typology and tailored anticipation strategies offered here provide the granular instrumentation needed to enact such responsibility within materials AI. Rather than treating ethical or societal considerations as post-hoc add-ons, the framework embeds them at the design stage, ensuring that materials discovery remains aligned with broader societal values.

Finally, the framework engages the notion of epistemic debt—the gradual accumulation of unexamined assumptions within complex technical systems. Hagendorff's reflections on forbidden knowledge and the limits of machine-learning research [25] warn that unmonitored reliance on AI can erode foundational epistemic practices. By distinguishing first-, second-, and third-order effects and by mandating ongoing baseline mapping, the framework functions as an early-warning system against precisely this form of debt. It complements broader analyses of AI's long-term impacts on scientific cooperation and epistemics [9] by furnishing a deployable tool where prior scholarship has remained largely diagnostic. In sum, the framework synthesizes these traditions into a cohesive, materials-specific instrument that transforms retrospective sociological insight into prospective governance.

Implications for Materials AI Practice

The conceptual framework carries direct and actionable implications for the four primary constituencies that shape materials AI deployment: authors, reviewers, funders, and the broader community. For authors, three practices become obligatory. First, every manuscript introducing or applying a new AI system must include a dedicated subsection discussing plausible second-order effects, grounded in the five-component anticipation process. Second, authors should outline concrete monitoring plans—specifying which strategies from Section 6 will be applied and at what intervals. Third, as effects emerge, authors must commit to iterative updates via follow-on commentaries or data statements, thereby treating second-order anticipation as an evolving responsibility rather than a one-time declaration. These steps elevate the discussion beyond first-order performance metrics and align with the call for transparency in scientific AI articulated by DeCost *et al.* [22].

Reviewers, in turn, must shift their evaluative lens. They should routinely ask whether the submission has addressed potential second-order effects and whether the claimed first-order gains are accompanied by foresight measures. A narrow focus on accuracy or discovery rate alone should be flagged as insufficient. Reviewers can invoke the typology and mechanisms presented here to probe whether authors have considered attention reallocation or success amplification. Such questioning, informed by the technology-assessment ethos [12], will gradually raise the standard of peer review and discourage the publication of unexamined AI deployments.

Funders bear perhaps the greatest leverage. They should require second-order effect analysis as a mandatory element of all proposals involving materials AI, using the five-component framework as a template. Funding calls can also earmark resources specifically for longitudinal monitoring studies that apply the strategies outlined in Section 6. By making such analysis a condition of support, funders prevent the institutional amplification mechanism from operating unchecked and ensure that public resources advance not only technical capability but also epistemic and social resilience.

At the community level, two collective actions are now urgent. First, the field should establish a shared second-order effects database—modeled on existing materials informatics repositories [27]—where practitioners deposit anonymized observations of behavioral, institutional, or normative shifts linked to specific AI interventions. Second, professional societies and workshops should develop and disseminate anticipation toolkits that package the framework, typology, and strategies into ready-to-use templates for laboratories and consortia. These toolkits, building on the responsible-innovation literature [20], will democratize foresight and prevent the burden of anticipation from falling solely on individual research groups. Together, these practice-level changes embed second-order anticipation as a core competency rather than an optional virtue, ensuring that materials AI evolves as a responsibly governed infrastructure.

Conclusion

This conceptual framework has demonstrated that the deployment of artificial intelligence in materials science cannot be evaluated solely through the lens of first-order technical success. By defining second-order effects, articulating six distinct types, identifying four propagation mechanisms, and proposing a five-component anticipation architecture, the paper supplies a structured and actionable approach to foresight. The tailored strategies for each effect type, together with their explicit linkages to baseline mapping, intervention specification, causal pathway mapping, stakeholder analysis, and scenario development, transform an abstract sociological warning into an operational practice.

The framework's integration with established concepts—Merton's unanticipated consequences, technology assessment, responsible innovation, and epistemic debt—grounds it in rigorous intellectual traditions while extending them into the unique context of materials informatics and autonomous discovery. Its implications for authors, reviewers, funders, and the community as a whole indicate a clear path forward: second-order effect anticipation must become standard operating procedure rather than an afterthought. Only then can the field harness the transformative power of AI without inadvertently reshaping research practices, institutional incentives, collaboration networks, and epistemic norms in ways that undermine long-term vitality.

The ultimate call is therefore simple yet urgent: every materials AI project, from initial conception through deployment and scaling, should incorporate the anticipation framework as an integral design element. In doing so, the community will not only accelerate discovery but also steward the scientific ecosystem that makes such discovery possible. The future of materials science depends not merely on what AI can predict, but on how wisely we anticipate the changes that prediction itself will bring.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 10 Nov 2024 Revised: 31 Dec 2024 Accepted: 07 Feb 2025
Published online: 18 July 2025

References

Bley F, Lapuschkin S, Samek W, Montavon G. Explaining predictive uncertainty by exposing second-order effects. Pattern Recognit. 2025;160:111171.

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.

Montoya JH, Aykol M, Anapolsky A, Gopal CB, Herring PK, Hummelshøj JS, et al. Toward autonomous materials research: Recent progress and future challenges. Appl Phys Rev. 2022;9(1):011405.

Zenil H, Tegnér J, Abrahão FS, Lavin A, Kumar V, Frey JG, et al. The future of fundamental science led by generative closed-loop artificial intelligence. Front Artif Intell. 2026;9:1678539.

Glickman M, Sharot T. How human–AI feedback loops alter human perceptual, emotional and social judgements. Nat Hum Behav. 2025;9(2):345-59.

Chubb J, Cowling P, Reed D. Speeding up to keep up: Exploring the use of AI in the research process. AI Soc. 2022;37(4):1439-57.

Clarke S, Whittlestone J. A survey of the potential long-term impacts of AI: How AI could lead to long-term changes in science, cooperation, power, epistemics and values. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society. AIES. New York (NY): ACM; 2022. p. 192-202.

Kieslich K, Diakopoulos N, Helberger N. Anticipating impacts: Using large-scale scenario-writing to explore diverse implications of generative AI in the news environment. AI Ethics. 2025;5(5):4555-77.

Ferrara E. The butterfly effect in artificial intelligence systems: Implications for AI bias and fairness. Mach Learn Appl. 2024;15:100525.

Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J, et al. Artificial intelligence and health technology assessment: Anticipating a new level of complexity. J Med Internet Res. 2020;22(7):e17707.

Bauer K, Von Zahn M, Hinz O. Expl (AI) ned: The impact of explainable artificial intelligence on users' information processing. Inf Syst Res. 2023;34(4):1582-602.

Choudhury A, Chaudhry Z. Large language models and user trust: Consequence of self-referential learning loop and the deskilling of health care professionals. J Med Internet Res. 2024;26:e56764.

Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA. 2017;318(6):517-8.

Kunze KN, Orr M, Krebs V, Bhandari M, Piuzzi NS. Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research: A call to emphasize data quality and indications. Bone Jt Open. 2022;3(1):93-7.

Yadav A, Pratap B, Shukla D. Prediction of second order effects structures using machine learning. Asian J Civ Eng. 2025;26(7):2933-59.

Palumbo D, De Finis R, Di Carolo F, Vasco-Olmo JM, Diaz FA, Galietti U. Influence of second-order effects on thermoelastic behaviour in the proximity of crack tips on titanium. Exp Mech. 2022;62(3):521-35.

Chang FR. Second-order effects of artificial intelligence. Issues Sci Technol. 2024;41(1).

Csernovszky A, Csete MS. Artificial intelligence and sustainability: A conceptual framework for system-level impact assessment. Cog Sust. 2026;5(1).
https://doi.org/10.55343/CogSust.22409.

Jeong J, Jeong I. Organizational AI investment and its ripple effects: An investigation of job insecurity, coworker conflict and work–family conflict. Pers Rev. 2025;54(7):1729-51.

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.

Prasittisopin L. Artificial intelligence-driven materiality: A systematic review and perspective. Eng Sci. 2025;35:1587.

Sen P. Artificial intelligence transforming materials research: Possible road ahead for India. Proc Indian Natl Sci Acad. 2025:1-9.

Hagendorff T. Forbidden knowledge in machine learning: Reflections on the limits of research and publication. AI Soc. 2021;36(3):767-81.

Righetti L, Madhavan R, Chatila R. Unintended consequences of biased robotic and artificial intelligence systems [ethical, legal, and societal issues]. IEEE Robot Autom Mag. 2019;26(3):11-3.

De Pablo JJ, Jackson NE, Webb MA, Chen LQ, Moore JE, Morgan D, et al. New frontiers for the materials genome initiative. npj Comput Mater. 2019;5(1):41.

Vasudevan R, Pilania G, Balachandran PV. Machine learning for materials design and discovery. J Appl Phys. 2021;129(7):070401.

Oehme E. Fostering collective action effectiveness for local rural development: Conceptualizing purposive interaction of social and human capital between decision-makers and project-owners/managers in northern Swedish NUTS-SE33 LEADER-territories [Master's thesis]. Linköping: Linköping University; 2022.