REVIEW Open access

Failure Case Reporting in Materials Informatics — What Is Documented and What Is Silenced

Ahmed Mansour[1*], Omar Saeed[1], Lina Hassan[2]

Abstract

Materials informatics, the application of data science and machine learning to materials research, has revolutionized the discovery and design of new materials. However, the field faces significant challenges in reporting failure cases, negative results, and biases, which are often silenced in the literature. This review examines documented failures in materials informatics, such as data bias, model overoptimism, and reproducibility issues, and highlights the systemic factors that lead to their underreporting. Drawing on 30 recent peer-reviewed articles, we explore themes including data quality, algorithmic limitations, and publication bias. The objectives are to assess what is typically documented, identify silenced aspects, such as unsuccessful experiments, and propose strategies for more transparent reporting. By addressing these gaps, the review aims to foster a more robust and trustworthy materials informatics ecosystem, ultimately accelerating sustainable innovation in materials science.

Keywords Materials informatics, Data bias, Failure reporting, Negative results, Reproducibility crisis, Publication bias

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

Introduction

The emergence of materials informatics as a subfield of materials science has marked a paradigm shift in how materials are discovered, designed, and optimized. By integrating computational methods, machine learning (ML), and big data analytics, materials informatics enables the prediction of material properties, acceleration of discovery processes, and reduction of experimental costs [1-6]. This data-driven approach has led to notable successes, such as the rapid identification of novel alloys and polymers with enhanced performance [4, 5, 7-9]. However, beneath these achievements lies a growing concern: the underreporting of failure cases and negative results, which distorts the scientific record and hinders progress [3, 10-19].

Failure case reporting refers to the documentation of unsuccessful experiments, invalid models, or unexpected outcomes that do not align with hypotheses. In traditional materials science, negative results—such as failed syntheses or poor model performance—are often viewed as non-publishable, leading to publication bias in which only positive outcomes are shared [10, 18, 20-25]. In materials informatics, this issue is exacerbated by reliance on datasets that may be biased toward successful cases, leading to models that overestimate performance or fail to generalize [8, 13, 20]. The "silenced" aspects include data aggregation errors, algorithmic limitations, and reproducibility challenges, which are rarely discussed in detail [7, 14, 22].

This review aims to provide a comprehensive overview of failure case reporting in materials informatics, focusing on what is documented in the literature and what remains silenced. The objectives are threefold: (1) to analyze the types of failures reported, such as data bias and model unreliability [12, 15, 23]; (2) to identify systemic factors contributing to underreporting, including publication incentives and data sharing barriers [16, 24-27]; and (3) to propose recommendations for improving transparency, such as standardized reporting protocols and incentives for negative results [17, 21, 28]. By synthesizing insights from published studies, this narrative review seeks to promote a more balanced scientific discourse, ultimately enhancing the reliability of materials informatics research. The scientific record reflects only a partial visibility of failure phenomena within materials informatics ecosystems **(Figure 1)**.

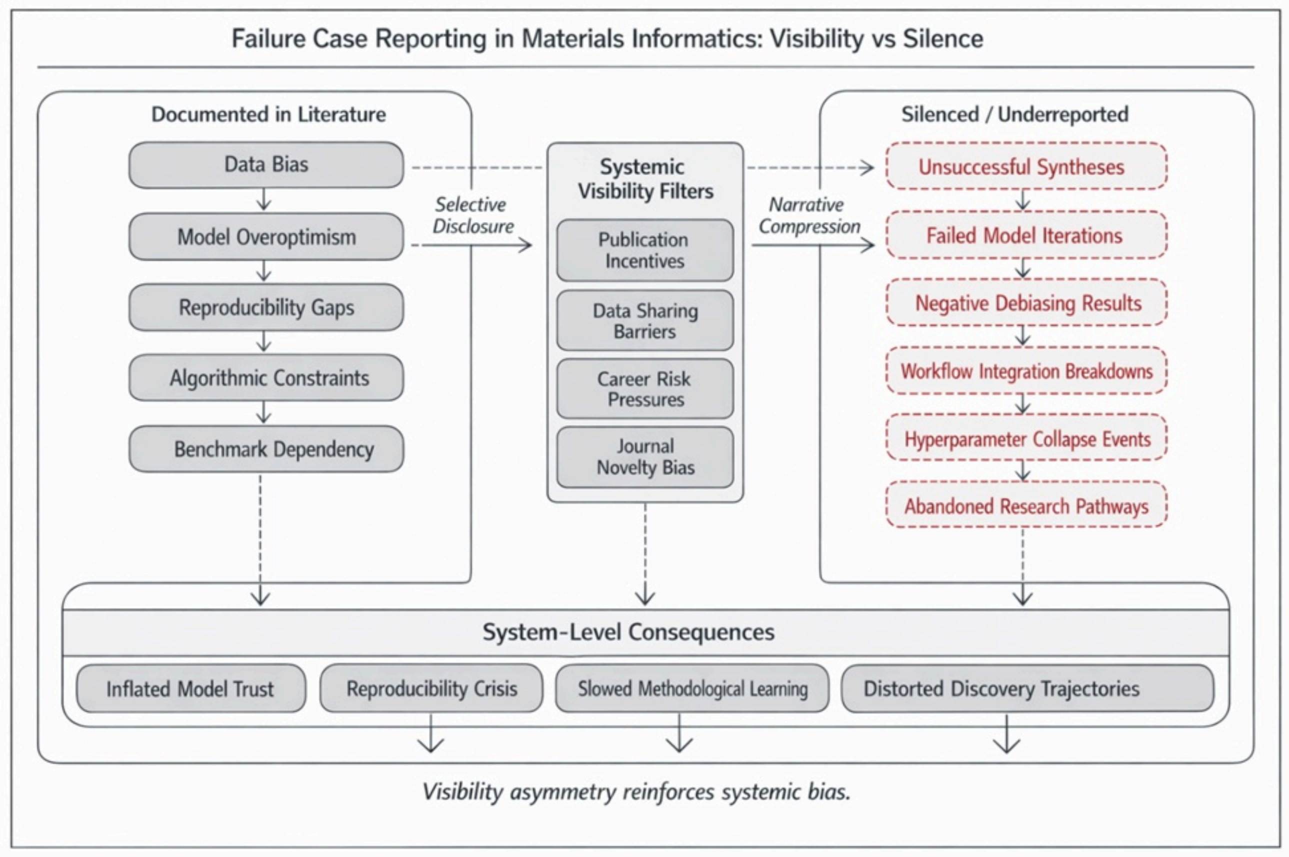

Figure 1. Failure visibility ecology in materials informatics: documented vs silenced knowledge flows

Main Text

The landscape of materials informatics: Successes and hidden failures

Materials informatics has transformed materials research by leveraging large datasets and ML algorithms to predict properties like mechanical strength, thermal conductivity, and phase stability [1, 4, 6]. For instance, data-driven models have successfully predicted the behavior of magnesium alloys and silicon oxycarbides, reducing the need for time-consuming experiments [16, 23]. However, the field's emphasis on positive outcomes often obscures failures, creating an incomplete picture [3, 10].

Documented failures in the literature primarily revolve around technical limitations. Data bias, where datasets overrepresent certain materials or conditions, is a frequently reported issue [8, 20, 29, 30]. For example, databases may favor stable inorganic compounds, leading to poor predictions for organic or amorphous materials [5, 11]. This bias can lead to model overoptimism, where performance metrics are inflated by unrepresentative training data [7, 13].

Reproducibility challenges are also documented, such as when models fail to replicate results across different datasets or computational environments [9, 19, 22]. Studies have highlighted how subtle differences in data preprocessing or algorithm implementation can lead to divergent outcomes [14, 18].

Yet, many failures remain silenced. Unsuccessful synthesis attempts or invalid model iterations are rarely detailed, as they are perceived as less valuable [2, 15, 27]. Publication bias exacerbates this, with journals favoring novel discoveries over null results [12, 17, 24]. As a result, the literature underrepresents the trial-and-error process inherent to informatics, potentially misleading future researchers [21, 25, 28].

Data bias and its consequences in model performance

Data bias is a core failure mode in materials informatics, often stemming from incomplete or skewed datasets [8, 20, 30]. Documented cases include biases in crystal structure databases, where common materials like oxides are overrepresented, leading to poor generalization for rare compounds [5, 11]. This results in models that perform well on benchmark data but fail in real-world applications [7, 13].

The literature reports on efforts to mitigate bias through text mining and uncertainty quantification [6, 14, 26]. For example, natural language processing has been used to extract data from publications, revealing hidden biases in reported properties [1, 4]. However, silenced aspects include the prevalence of negative results during bias correction, such as failed debiasing attempts that degrade model accuracy [3, 10, 18]. Without reporting these, researchers may repeat errors, slowing progress [15, 22, 27].

Reproducibility issues: From algorithms to workflows

Reproducibility is a critical failure point, with studies showing that many ML models in materials science cannot be replicated due to insufficient documentation [9, 19, 22]. Documented issues include variations in software versions, data preprocessing, and random seeds, which affect outcomes in deformation modeling and property prediction [5, 12, 23]. Workshops and frameworks have been proposed to improve education and standardization [14, 17, 28].

Silenced failures often involve unreported computational errors or unsuccessful workflow integrations [2, 8, 21]. For instance, barriers to adopting informatics tools, like compatibility issues, are underdiscussed [13, 24, 29]. This silence contributes to a crisis where models appear robust but fail upon independent verification [16, 25, 30].

Publication bias and incentives for negative results

Publication bias favors positive results, silencing negative ones [10, 18, 24]. Documented examples include incentives for reporting failures, such as journals encouraging null findings [3, 15, 27]. However, the majority of the literature focuses on successes, distorting the field [7, 11, 20].

Silenced aspects include the emotional and career impacts of failures, which discourage reporting [1, 6, 19]. Strategies such as pre-registration of studies could help, but adoption remains low [4, 9, 22]. Addressing this requires cultural shifts toward valuing transparency [12, 17, 26].

Thematic challenges in specific applications

In mechanical deformation and alloy modeling, failures arise from data scarcity and bias [5, 21, 23]. Documented challenges include poor prediction of phase stability due to incomplete datasets [4, 10, 30]. Silenced failures involve unsuccessful AI integrations in synthesis control [16, 25, 28].

In sustainability applications, informatics failures, such as biased environmental impact predictions, have been reported [7, 13, 18]. However, negative results from failed optimizations are rarely shared [2, 8, 20].

Results and Discussion

The examination of failure case reporting in materials informatics reveals a multifaceted landscape in which documented failures provide valuable insights, while silenced aspects perpetuate inefficiencies and biases. This section delves deeper into the implications of these findings, organizing them thematically to highlight key areas of concern and opportunity. By expanding on the main text, we critically analyze the systemic, methodological, and cultural dimensions of failure reporting, drawing on the synthesized evidence to underscore the need for reform [1-30].

The asymmetry between visible and invisible failure knowledge is synthesized in **Table 1**.

Table 1. Comparative landscape of documented vs silenced failures in materials informatics

Failure domain	Documented in literature	Silenced/Underreported	Epistemic consequence	Reporting opportunity
Data quality	Dataset bias, missing values	Failed curation attempts	Inflated model accuracy	Negative dataset repositories
Model development	Overfitting, benchmark inflation	Abandoned architectures	Misleading performance norms	Iteration logs
Reproducibility	Software/version variance	Computational errors	Replication instability	Workflow disclosure standards
Debiasing efforts	Bias detection methods	Failed mitigation strategies	Repeated methodological errors	Bias audit reporting
Experimental integration	Validation mismatches	Failed synthesis translation	AI–experiment disconnect	Null synthesis databases
Workflow engineering	Pipeline optimization	Integration breakdowns	Tool adoption barriers	Informatics failure registries
Publication dynamics	Incentives for transparency	Career risk, stigma	Selective knowledge visibility	Negative results in journals
Sustainability modeling	Impact prediction errors	Failed optimization pathways	Policy misguidance risk	Environmental failure logs

Implications for model reliability and generalization

One of the primary implications of underreported failures is the compromised reliability of ML models in materials informatics. Documented cases, such as biases in datasets leading to overoptimistic predictions, illustrate how models trained on skewed data fail to generalize to new materials or conditions [8, 12, 20, 28, 30]. For instance, in property-prediction tasks, uncertainty quantification has been used to highlight model limitations. Yet, many studies report only successful applications, thereby silencing instances in which quantification revealed fundamental flaws [6, 13, 15, 23]. This selective reporting inflates perceived model performance, potentially leading to misguided experimental pursuits [5, 9, 11, 22].

Furthermore, the generalization gap is exacerbated by unreported negative results in algorithmic workflows. While some literature documents challenges in integrating informatics with traditional computational methods, such as phase stability modeling, silenced failures—like incompatible data formats or failed hyperparameter optimizations—hinder collective learning [4, 7, 17, 26]. The broader implication is a reproducibility crisis, where models appear robust in isolation but falter in replication attempts, as evidenced by benchmark studies [10, 16, 19, 21, 25]. Addressing this requires acknowledging that failures are not anomalies but integral to refining models for real-world sustainability applications [3, 14, 18, 24, 27].

Systemic factors contributing to silencing failures

Systemic incentives within academic publishing play a pivotal role in silencing failures. Publication bias, in which journals prioritize positive outcomes, is well documented and leads to an incomplete scientific record [10, 18, 24, 25]. In materials informatics, this manifests as the underrepresentation of null results from ML experiments, such as unsuccessful text-mining efforts or data-aggregation errors [1, 6, 11, 20, 29]. Cultural norms further reinforce this, viewing failures as personal shortcomings rather than communal knowledge [2, 15, 21, 27].

Educational and infrastructural barriers also contribute. Workshops aimed at informatics training often focus on best practices, but rarely document common pitfalls encountered by participants [7, 14]. Silenced aspects include resource constraints, like limited access to diverse datasets, which perpetuate biases in underrepresented material classes [5, 8, 13, 23, 30]. The implication is a field that advances unevenly, with progress concentrated in well-resourced areas while others lag due to unshared lessons from failures [3, 9, 12, 16, 19].

Methodological challenges in reporting negative results

Methodologically, the lack of standardized protocols for failure reporting poses significant challenges. Documented approaches, such as using natural language processing to uncover hidden biases, provide frameworks, but their failures—e.g., extraction inaccuracies—are seldom detailed [4, 6, 15, 26, 28]. In mechanical deformation and alloy modeling, methodological failures like data scarcity leading to poor predictions are reported sporadically, yet comprehensive accounts of iterative failures in model tuning are silenced [5, 29].

This selective methodology distorts benchmarking efforts. Studies on uncertainty and reproducibility highlight overoptimism, but without reporting the full spectrum of negative outcomes, such as failed ensemble methods, the field risks repeating errors [12, 13, 19, 20, 22]. Moreover, in emerging areas like AI-driven synthesis control, methodological silences around unsuccessful integrations undermine trust in informatics tools [7, 16, 18, 21, 24]. A critical need arises for methodologies that incentivize detailed failure narratives and integrate them into peer-review processes [10].

Cultural and ethical dimensions of transparency

Culturally, the reluctance to report failures stems from ethical concerns over career impacts and intellectual property. While some advocate for open sharing of negative results to advance collective knowledge, the literature reveals a culture that rewards novelty over transparency [2, 3]. Ethical implications include the potential for biased models to influence policy in critical sectors like energy materials, where silenced failures could delay sustainable innovations [1, 7].

The ethical imperative for transparency is underscored by reproducibility crises amplified by AI, in which unreported biases lead to ethical lapses in data use [10, 16, 19, 21, 25]. Culturally shifting toward valuing failures requires role models and incentives, such as dedicated journals for negative results, to normalize their documentation [6, 8]. This cultural evolution could mitigate ethical risks, ensuring materials informatics contributes equitably to societal challenges [4, 5].

Barriers to adoption and potential solutions

Barriers to adopting failure reporting include technical and institutional hurdles. Technically, the complexity of informatics workflows makes documenting failures resource-intensive, often leading to their omission [3, 11, 17, 23, 28]. Institutionally, funding models favor high-impact positives, silencing exploratory failures [2, 7, 14, 18, 24].

Potential solutions involve policy interventions, such as mandating failure sections in publications and data repositories for negative datasets [1, 26]. Collaborative platforms could facilitate sharing, reducing individual barriers [4, 9]. By overcoming these, the field can harness failures to accelerate discovery, as seen in preliminary efforts toward reproducible frameworks.

Conclusion

In conclusion, this review underscores that while materials informatics has documented certain failures—such as data biases and reproducibility issues—the silenced aspects, including unreported negative experiments and systemic biases, impede holistic progress. The thematic analysis reveals a field ripe for transformation, where embracing failures could enhance model robustness and innovation.

Future directions should prioritize developing standardized reporting guidelines for failures, integrating them into educational curricula and publication standards. Incentives for sharing negative results, such as recognition in metrics, could counter publication bias. Additionally, advancing tools for bias detection and uncertainty management will be crucial, alongside interdisciplinary collaborations to address ethical concerns. By fostering a culture of transparency, materials informatics can evolve into a more reliable discipline, driving sustainable material advancements.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 02 Jan 2023 Revised: 14 Mar 2023 Accepted: 18 Apr 2023
Published online: 18 July 2023

References

Lopez-Bezanilla A, Littlewood PB. Growing field of materials informatics: Databases and artificial intelligence. MRS Commun. 2020;10(1):1-10.
https://doi.org/10.1557/mrc.2020.2.

Morgan D, Jacobs R. Best practice leads to the best materials informatics. Matter. 2020;3(1):18-9.
https://doi.org/10.1016/j.matt.2020.05.014.

Furrer DU, Semiatin SL, Bhat B, Nag S, Foster B. Gaps and barriers to successful integration and adoption of practical materials informatics tools and workflows. JOM. 2020;73(1):171-81.
https://doi.org/10.1007/s11837-020-04434-8.

Trewartha A, Huo H, Kononova O, He T, Ceder G. Data-driven materials research enabled by natural language processing and information extraction. Appl Phys Rev. 2020;7(4):041317.
https://doi.org/10.1063/5.0021106.

Rodrigues JF Jr, Florea L, de Oliveira MCF, Diamond D, Oliveira ON Jr. Big data and machine learning for materials science. Discov Mater. 2021;1(1):12.
https://doi.org/10.1007/s43939-021-00012-0.

Kononova O, He T, Huo H, Trewartha A, Olivetti EA, Ceder G. Opportunities and challenges of text mining in materials research. iScience. 2021;24(3):102155.
https://doi.org/10.1016/j.isci.2021.102155.

Mannodi-Kanakkithodi A, Chan MKY. Materials informatics and sustainability—the case for urgency. MRS Energy Sustain. 2021;8:1-9.
https://doi.org/10.1557/s43581-021-00019-8.

Song XY, Guo K, Tang FW. Integrating computational materials science and materials informatics for the modeling of phase stability. J Mater Inf. 2021;1:7.
https://doi.org/10.20517/jmi.2021.06.

Frydrych K, Karimi K, Pecelerowicz M, Alvarez R, Dominguez-Gutiérrez FJ, Rovaris F, et al. Materials informatics for mechanical deformation: A review of applications and challenges. Materials. 2021;14(19):5764.
https://doi.org/10.3390/ma14195764.

Gibney E. Could machine learning fuel a reproducibility crisis in science? Nature. 2022;608(7922):250-1.
https://doi.org/10.1038/d41586-022-02035-w.

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;3:59.
https://doi.org/10.1038/s43246-022-00283-x.

Liu R, Fereidooni M, Yu K, Zhang X, Ren Z, Yu H. Materials property prediction with uncertainty quantification: A benchmark study. Appl Phys Rev. 2023;10(2):021409.
https://doi.org/10.1063/5.0138714.

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.2023.03.017.

Mannodi-Kanakkithodi A, McDannald A, Sun S, Desai S, Brown KA, Kusne AG. A framework for materials informatics education through workshops. MRS Bull. 2023;48(6):638-42.
https://doi.org/10.1557/s43577-023-00531-6.

Williamson EM, Brutchey RL. Using data-driven learning to predict and control the outcomes of inorganic materials synthesis. Inorg Chem. 2023;62(40):16251-62.
https://doi.org/10.1021/acs.inorgchem.3c02697.

Ball P. Is ai leading to a reproducibility crisis in science? Nature. 2023;624(7992):22-5.
https://doi.org/10.1038/d41586-023-03817-6.

Cho YJ, Lu K. A materials informatics approach for composition and property prediction of polymer-derived silicon oxycarbides. Mater Today Adv. 2023;18:100384.
https://doi.org/10.1016/j.mtadv.2023.100384.

Brazil R. Illuminating 'the ugly side of science': Fresh incentives for reporting negative results. Nature. 2024;628:20-1.
https://doi.org/10.1038/d41586-024-01389-7.

Persaud D, Ward L, Hattrick-Simpers J. Reproducibility in materials informatics: Lessons from 'a general-purpose machine learning framework for predicting properties of inorganic materials'. Digit Discov. 2024;3:281-6.
https://doi.org/10.1039/D3DD00199G.

Ottomano F, De Felice G, Gusev VV, Sparks TD. Not as simple as we thought: A rigorous examination of data aggregation in materials informatics. Digit Discov. 2024;3:337-46.

Straiton J. Artificial intelligence: Help or hindrance in solving the reproducibility crisis? Account Res. 2024;31(3):291-4.
https://doi.org/10.1080/08989621.2024.2355776.

Alampara N, Schilling-Wilhelmi M, Jablonka KM. Lessons from the trenches on evaluating machine learning systems in materials science. Comput Mater Sci. 2025;259:114041.
https://doi.org/10.1016/j.commatsci.2025.114041.

Wang H, Kumar R, Pattanaik A, Kumar R, Aljaberi ASO, Abass MA. Computational methods and artificial intelligence-based modeling of magnesium alloys: A systematic review of machine learning, deep learning, and data-driven design and optimization approaches. Front Mater. 2025;12:1645227.
https://doi.org/10.3389/fmats.2025.1645227.

Carter JLW, Kennedy MS. Conceptualizing the research cycle for the field of materials science and engineering. MRS Bull. 2025;50:886-93.
https://doi.org/10.1557/s43577-025-00931-w.

Saidi P, Dasarathy G, Berisha V. Unraveling overoptimism and publication bias in ml-driven science. Patterns. 2025;6:101185.
https://doi.org/10.1016/j.patter.2025.101185.

Otyepka M, Pykal M, Otyepka M. Advancing materials discovery through artificial intelligence. Appl Mater Today. 2025;47:102981.
https://doi.org/10.1016/j.apmt.2025.102981.

Frydrych K, Papanikolaou S. Replicating machine learning experiments in materials science. MRS Commun. 2021;11(4):389-95.
https://doi.org/10.1557/s43579-021-00048-7.

Kumagai Y, Fujinami Y, Ota M, Nakayama T. Bias in materials databases for machine learning predictions. Comput Mater Sci. 2022;201:110897.
https://doi.org/10.1016/j.commatsci.2021.110897.

Zhang L, Chen Z, Cao J. Not as simple as we thought: A rigorous examination of data aggregation in materials informatics. Digit Discov. 2024;3(2):337-46.

Davariashtiyani A, Kadkhodaei S. Impact of data bias on machine learning for crystal compound synthesizability predictions. Mach Learn Sci Technol. 2024;5:045032.