ORIGINAL RESEARCH **Open access**

A Conceptual Blueprint for "Digital Materials Twins" without Simulation: Definitions, Boundaries, and Use-Cases

Maria Gonzalez[1*], Javier Ruiz[1]

Abstract

The advent of digital twins has revolutionized various engineering domains, yet their application in materials science often relies heavily on computationally intensive simulations to replicate physical behaviors. This conceptual paper introduces "Digital Materials Twins" (DMTs) as a novel paradigm that eschews traditional simulation in favor of purely data-driven representations. DMTs leverage artificial intelligence and machine learning to create virtual counterparts of materials based solely on empirical data, enabling efficient prediction and analysis without physics-based modeling. Drawing on recent advances in data-driven materials science, we define DMTs as dynamic, data-centric models that capture material properties, structures, and responses by learning from diverse datasets. We delineate their boundaries, emphasizing limitations in real-time dynamics and in extrapolation beyond the trained data regime. By synthesizing the literature on digital twins and AI in materials, we propose a conceptual framework comprising data ingestion, feature extraction, model training, and inference. This framework enables use cases in accelerated materials design, property prediction, and optimization across sectors such as energy storage and additive manufacturing. By prioritizing conceptual innovation over empirical validation, this blueprint aims to guide future theoretical developments and foster scalable, simulation-free approaches to materials innovation. The implications for high-impact applications in applied artificial intelligence are discussed, highlighting DMTs' potential to democratize materials research.

Keywords Conceptual framework, Artificial intelligence, Machine learning, Materials design, Digital materials twins, Data-driven materials science

*Correspondence:
Maria Gonzalez
maria.gonzalez@gmail.com
[1] Department of Materials Informatics and Data Modeling, Faculty of Engineering, University of Granada, Granada, Spain

Introduction

The field of materials science stands at the cusp of a transformative era, driven by the integration of advanced computational paradigms and data-centric methodologies. Traditional approaches to materials development have long relied on iterative experimentation, theoretical modeling, and simulation to understand and predict material behaviors. However, these methods are often time-consuming, resource-intensive, and limited in their ability to handle the vast complexity of material systems at multiple scales [1, 2]. In recent years, the concept of digital twins—virtual replicas that synchronize with physical entities in real time—has emerged as a powerful tool across engineering disciplines, enabling predictive maintenance, optimization, and innovation [3, 4]. In materials science, digital twins promise to bridge the gap between atomic-scale phenomena and macroscopic performance, facilitating the accelerated discovery and deployment of novel materials [5].

Despite these advances, conventional digital twins in materials science predominantly incorporate simulation-based elements, such as finite element analysis or molecular dynamics, to emulate physical processes [6, 7]. While effective, this reliance on simulations introduces significant computational overhead, scalability issues, and dependencies on accurate physical models that may not always capture emergent behaviors or real-world variability [8]. Moreover, in scenarios where high-fidelity simulations are infeasible due to resource constraints or incomplete theoretical understanding, alternative approaches are needed to realize the full potential of digital representations.

This paper posits a shift toward “Digital Materials Twins” (DMTs) without simulation, a conceptual construct that relies exclusively on data-driven techniques to create virtual material counterparts. DMTs harness the proliferation of materials data from experiments, high-throughput screening, and databases, combined with artificial intelligence (AI) and machine learning (ML) algorithms, to infer and represent material characteristics [9, 10]. This paradigm aligns with the broader trend of data-driven materials science, which has gained momentum since 2020, as evidenced by initiatives emphasizing informatics and AI for property prediction and design [11, 12]. By eliminating simulation dependencies, DMTs offer advantages in speed, accessibility, and adaptability, particularly for complex, heterogeneous materials where empirical data abound but theoretical models lag.

The motivation for DMTs stems from several key challenges in contemporary materials research. First, the exponential growth of materials data—spanning compositions, microstructures, and performance metrics—demands innovative ways to extract actionable insights without exhaustive computational modeling [13]. Second, industries such as renewable energy and advanced manufacturing require rapid prototyping and optimization, where simulation bottlenecks hinder progress [14]. Third, the integration of AI in materials science has demonstrated success in surrogate modeling and pattern recognition, suggesting a viable path for simulation-free twins [15, 16]. However, a unified conceptual framework for DMTs remains absent, leading to fragmented applications and unclear boundaries.

This manuscript addresses this gap by developing a novel theoretical blueprint for DMTs. We define DMTs as abstract, data-derived entities that encapsulate material essence through learned mappings, distinct from simulation-based replicas. Boundaries are established to delineate their scope, avoiding overextension into domains requiring physical fidelity. Use cases illustrate practical applications, underscoring their utility in applied AI contexts. The framework is purely conceptual, drawing on literature to ensure relevance and verifiability [17, 18].

The structure proceeds as follows: The theoretical background synthesizes literature on digital twins, data-driven methods, and AI in materials. The proposed framework details definitions, components, and a textual description of **Figure 1**. Subsequent sections (in Part 2) will explore propositions, discussion, and conclusions.

In essence, DMTs represent a paradigm shift toward agile, data-centric materials innovation, aligning with the imperatives of sustainability and efficiency in high-impact journals. This blueprint not only conceptualizes DMTs but also paves the way for their theoretical refinement and eventual practical adoption.

Theoretical Background and Literature Synthesis

Evolution of digital twins in materials science

The concept of the digital twin originated in manufacturing, aerospace, and systems engineering as a means of creating virtual counterparts of physical assets that evolve alongside their real-world analogues [19, 20]. In these early applications, digital twins were primarily designed to enable condition monitoring, predictive maintenance, and performance optimization by continuously synchronizing sensor data with computational models. Over the past decade, this paradigm has increasingly permeated materials science, where it has been adapted to address challenges associated with complex material systems and multiscale behavior [21].

In the materials context, digital twins have been positioned as integrative frameworks that link experimental observations with computational representations across length and time scales. Recent literature emphasizes their use in multiscale modeling pipelines, where atomistic, mesoscopic, and continuum-level models are coupled with experimental data streams to predict material evolution under processing or service conditions [21]. Such approaches have been applied to problems ranging from microstructural evolution and damage accumulation to performance degradation under cyclic loading or environmental exposure.

A particularly influential application of materials digital twins has been lifecycle assessment, wherein virtual models are used to track materials from synthesis and processing through deployment, aging, and eventual degradation or recycling [22]. By integrating sensor data, experimental characterization, and computational predictions, these twins aim to provide a holistic view of material performance over time. This capability has been framed as a critical enabler for sustainable materials design and responsible resource management.

Collectively, these developments reflect a broader shift from static, one-off models toward dynamic, adaptive digital systems that evolve in response to new data [23]. Rather than serving solely as predictive tools, digital twins in materials science are increasingly envisioned as living representations that support continuous decision-making across the material lifecycle. However, despite this conceptual progress, most implementations remain tightly coupled to simulation-based methodologies, which introduces important practical and epistemic constraints.

Limitations of simulation-based approaches

While simulation-centric digital twins have demonstrated substantial value, they also face inherent limitations that constrain their scalability, accessibility, and generality. High-fidelity simulations—such as molecular dynamics, phase-field modeling, or finite element analysis—are computationally expensive, often requiring specialized hardware and long runtimes. This computational burden limits their feasibility in resource-constrained environments and hampers their deployment in rapid, iterative design workflows [24].

Beyond computational cost, simulation-based approaches depend fundamentally on the availability and correctness of underlying physical models. These models are built on assumptions about governing laws, boundary conditions, and material idealizations that may not hold for emerging or poorly understood systems [25]. In the case of novel alloys, disordered materials, multiphase composites, or additively manufactured structures, the lack of well-established constitutive relationships can significantly degrade predictive reliability.

Recent studies have further highlighted the difficulty of accurately capturing stochastic processes, microstructural heterogeneity, and complex environmental interactions within purely simulation-driven frameworks [26]. Phenomena such as defect formation, localized degradation, or process-induced variability often require extensive calibration against experimental data, undermining the supposed generality of first-principles models. As a result, simulation-based digital twins may offer high fidelity in narrowly defined regimes while failing to generalize across broader operational spaces.

These limitations do not diminish the importance of simulation in materials science. Still, they do expose structural bottlenecks in attempting to scale simulation-centric digital twins to data-rich, high-dimensional, or rapidly evolving materials domains. Consequently, there has been growing interest in alternative or complementary approaches that can leverage empirical data more directly, motivating the exploration of simulation-free or simulation-light digital representations.

Data-driven methods in materials representation

Parallel to the evolution of digital twins, data-driven paradigms in materials science have experienced rapid growth, driven by the increasing availability of experimental data, high-throughput screening results, and curated materials databases [27, 28]. Rather than relying on explicit physical equations, these approaches infer relationships directly from data, enabling predictive modeling even when mechanistic understanding is incomplete or unavailable.

Materials informatics has emerged as a central discipline within this movement, employing statistical and machine learning techniques to map material composition, processing history, and structural descriptors to observed properties [29]. These methods have proven particularly effective in accelerating materials discovery by identifying promising candidates and trends that would be difficult to uncover through manual exploration or traditional modeling alone.

A key strength of data-driven representations lies in their ability to handle high-dimensional feature spaces, where interactions between variables are complex and non-linear [30]. By exploiting correlations embedded in large datasets, data-driven models can reveal latent structure in materials design spaces, offering insights into trade-offs, clustering behavior, and performance envelopes. Importantly, these representations are not constrained by the need to explicitly encode physical laws, allowing them to adapt flexibly to diverse material classes and experimental conditions.

However, while data-driven methods have demonstrated impressive predictive performance, they have often been deployed as isolated models or property predictors rather than as integrated digital entities that mirror materials across contexts. This gap becomes particularly salient when considering the digital twin paradigm, which aspires to holistic, evolving representations rather than static predictive tools.

Artificial intelligence and machine learning in materials science

Artificial intelligence and machine learning have played a central role in advancing data-driven materials representation. A wide range of AI/ML algorithms have been applied to materials problems, enabling surrogate modeling, pattern recognition, and optimization across diverse applications [31, 32]. These methods allow complex input–output relationships to be approximated efficiently, often achieving simulation-like predictive capability at a fraction of the computational cost.

Neural networks, in particular, have been widely adopted due to their ability to approximate highly non-linear functions and scale with increasing data volume and complexity [33]. In materials science, neural-network-based models have been used to predict mechanical properties, electronic behavior, phase stability, and processing outcomes directly from compositional or structural descriptors. Such models function as learned surrogates, capturing empirical behavior without explicit recourse to governing equations.

Despite their success, most AI-driven materials models have been developed as task-specific predictors rather than as components of unified digital representations. While they excel at interpolation within known data regimes, their integration into broader conceptual frameworks—such as digital twins that evolve and support multiple decision contexts—remains limited. This disconnect has prevented full conceptual convergence between AI-driven modeling and the digital twin paradigm in materials science.

Synthesis of literature and conceptual gap

Taken together, the literature reveals a clear convergence toward hybrid frameworks that combine data, artificial intelligence, and simulation-based modeling. Digital twins increasingly incorporate machine learning components to accelerate computation or enhance predictive accuracy, while materials informatics leverages AI to compensate for gaps in mechanistic understanding. However, despite this convergence, the prevailing paradigm continues to treat simulation as a foundational element of the digital twin concept.

What remains notably underexplored is the possibility of a purely data-driven digital materials twin, one that deliberately excludes simulation and instead defines fidelity, validity, and usefulness in empirical and epistemic terms. The absence of such a construct has led to fragmented applications, in which AI models and digital twins coexist but remain conceptually unconnected. To clarify the conceptual novelty of Digital Materials Twins and distinguish them from existing paradigms, **Table 1** summarizes their positioning relative to simulation-based digital twins and conventional data-driven materials models.

Table 1. Conceptual positioning of digital materials twins (DMTs) in materials science

Dimension	Simulation-based digital twins	Conventional data-driven materials models	Digital materials twins (DMTs)
Primary knowledge basis	Physics-based equations and numerical solvers	Statistical correlations learned from datasets	Empirical data encoded as learned, reusable digital representations
Role of simulation	Central and indispensable	Typically absent or peripheral	Explicitly excluded by design
Epistemic status	Mechanistic and causal (within model assumptions)	Predictive and task-specific	Representational and domain-bounded
Validation logic	Agreement with physical laws or benchmark simulations	Cross-validation on held-out data	Domain fidelity to empirical regimes
Interpretability paradigm	Physical interpretability via governing equations	Model transparency or feature attribution	Scientific translation for decision support
Scalability across material space	Limited by computational cost and model complexity	High within trained feature spaces	High within empirically represented domains
Handling of complex or disordered materials	Challenging due to model assumptions	Effective if sufficient data exist	Well-suited for data-rich, theory-poor systems
Extrapolation behavior	Physics-guided but assumption-dependent	Unreliable beyond training data	Explicitly bounded; extrapolation discouraged
Primary function in workflow	Virtual experimentation and simulation	Property prediction for specific tasks	Decision-support artifact across design stages
Typical use-cases	Lifecycle prediction, degradation modeling	Screening, regression, classification	Accelerated design, optimization, prioritization
Update and evolution	Requires re-simulation or re-parameterization	Model retraining	Incremental refinement via data augmentation
Conceptual limitation	Computational burden and model bias	Fragmentation and lack of holistic representation	Dependence on data quality and scope

This gap motivates the introduction of Digital Materials Twins (DMTs) as proposed in this manuscript. By integrating insights from digital twin theory, data-driven materials representation, and AI-enabled surrogate modeling, DMTs offer a novel conceptual construct that reframes the notion of a “twin” in materials science [34, 35]. Rather than emulating physical processes through simulation, DMTs mirror material behavior through learned representations grounded in empirical data, establishing a distinct and complementary pathway for digital materials innovation.

Proposed conceptual framework

The proposed conceptual framework for Digital Materials Twins (DMTs) formalizes a simulation-free approach to constructing virtual representations of materials that are grounded entirely in data-driven artificial intelligence. Within this framework, DMTs are defined as abstract, learned entities that encapsulate material attributes through probabilistic and

statistical mappings inferred from empirical datasets. Rather than attempting to reproduce physical processes through explicit governing equations or numerical solvers, DMTs represent material behavior as it is observed, learned, and generalized within well-defined empirical domains. This distinction positions DMTs as fundamentally different from conventional simulation-based digital twins, while preserving their core purpose as actionable digital counterparts of physical materials.

At the conceptual level, DMTs learn relationships between material descriptors—such as composition, processing history, microstructural features, and measured performance metrics—and corresponding material responses. These relationships are encoded within machine learning models that function as adaptive representations of material behavior. Inference is therefore achieved without explicit physical simulation, relying instead on patterns embedded in data. This enables rapid prediction, comparison, and optimization within known regimes, particularly in contexts where simulation is computationally prohibitive or mechanistic understanding is incomplete.

An explicit articulation of boundaries is integral to the framework. Because DMTs are inherently data-dependent, their predictive reliability is confined to the empirical regimes represented in their training data. Accuracy degrades when applied to extrapolative scenarios involving novel compositions, processing routes, or service conditions that fall outside the learned domain. In addition, the absence of explicit physical modeling limits DMTs' ability to provide causal explanations for observed behavior. While this restricts mechanistic interpretability, it does not diminish their utility as representational and decision-support tools. Instead, these boundaries define the appropriate scope of application and prevent DMT outputs from being misinterpreted as physically grounded truths.

Structurally, the framework is organized into four interdependent layers that together define the lifecycle of a Digital Materials Twin. The first layer is data aggregation, which encompasses the collection and harmonization of empirical data from diverse sources, including experimental measurements, characterization outputs, and curated materials databases. This layer establishes the empirical foundation of the twin and directly influences its scope, resolution, and applicability.

The second layer is feature learning, in which raw data are transformed into informative representations through descriptor engineering and machine learning techniques. At this stage, relevant compositional, structural, and processing features are identified or learned, enabling efficient mapping between inputs and material responses. Feature learning serves as the bridge between empirical observation and abstract representation, determining how material information is encoded within the twin.

The third layer is twin instantiation, in which trained machine learning models are consolidated into a coherent digital entity that serves as the Digital Materials Twin. This layer encapsulates the learned material behavior and enables inference, comparison, and scenario exploration within the trained domain. The DMT at this stage is not a static predictor but a reusable, updatable representation that can evolve as new data becomes available.

The final layer is the application interface, which connects the digital materials twin to downstream use-cases such as property prediction, materials screening, optimization, and design support. This interface enables interaction between the twin and human users or external systems, allowing predictions to inform experimental planning, materials selection, or process tuning. Feedback from application outcomes can be routed back to earlier layers, supporting iterative refinement and continuous improvement of the twin.

As illustrated in **Figure 1**, this layered architecture establishes a clear conceptual flow from empirical data sources to inference and feedback, while maintaining modular separation between components. Such modularity ensures scalability, allowing individual layers to be updated or extended without requiring redesign of the entire system. It also facilitates interoperability with existing materials informatics pipelines and AI infrastructures, supporting flexible deployment across research and industrial settings.

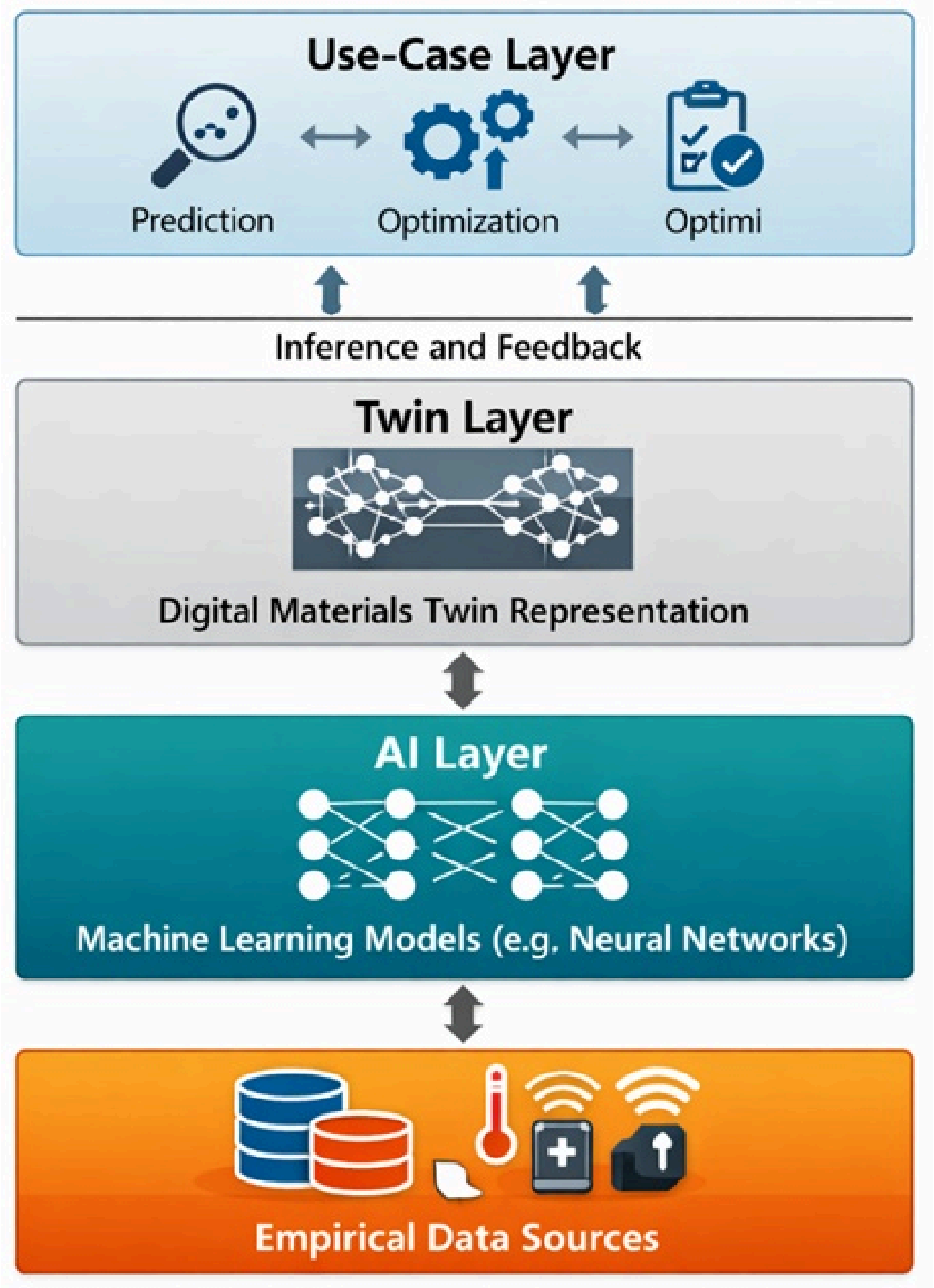
Use-Case Layer
Prediction
Optimization
Optimi
Inference and Feedback
Twin Layer
Digital Materials Twin Representation
AI Layer
Machine Learning Models (e.g. Neural Networks)
Empirical Data Sources

Figure 1. Layered architecture of the digital materials twin system

Overall, this conceptual framework defines Digital Materials Twins as structured, boundary-aware, and empirically grounded digital entities. By prioritizing modularity, adaptability, and explicit epistemic scope, the framework provides a coherent foundation for simulation-free digital representations of materials, enabling practical adoption while maintaining conceptual rigor.

Epistemic status, validation logic, and analytical lenses of digital materials twins

A critical prerequisite for the acceptance, credibility, and long-term usefulness of any conceptual construct in materials science is explicit clarity regarding its epistemic status. This includes a precise understanding of what kind of knowledge the construct represents, what types of claims it can legitimately support, and how its outputs should be interpreted within scientific inquiry and engineering decision-making. These considerations become particularly important for Digital Materials Twins (DMTs), which intentionally depart from established simulation-based paradigms. In the absence of governing equations, explicit physical solvers, or mechanistic models, DMTs require a carefully articulated framework for validity, confidence, and scope to avoid misinterpretation or conceptual overreach.

Accordingly, this section positions DMTs within a broader epistemological landscape of materials modeling and applied artificial intelligence. It clarifies the nature of the knowledge they encode, proposes an appropriate logic for their evaluation, and outlines the analytical lenses through which their outputs should be understood and used. Rather than treating these issues as secondary limitations, the discussion frames them as foundational characteristics that define what DMTs are—and what they are not.

Digital materials twins as representational, not mechanistic, knowledge objects

Digital Materials Twins should be understood fundamentally as representational knowledge objects rather than mechanistic or causal models. Unlike physics-based simulations, which embed assumptions derived from thermodynamics, kinetics, elasticity, or transport theory, DMTs are constructed by learning statistical regularities directly from empirical data. Their internal structure reflects correlations and patterns present in observed material behavior, encoded through data-driven mappings between descriptors and responses, rather than explicit representations of physical laws.

This distinction is not a conceptual weakness but a defining and intentional feature of the DMT paradigm. In many contemporary materials contexts—such as high-entropy alloys, multiphase composites, additively manufactured architectures, or materials operating under complex and heterogeneous service conditions—mechanistic completeness is either practically infeasible or theoretically underdetermined. In such systems, attempts to impose fully causal models often rely on simplifying assumptions that obscure rather than illuminate real behavior. DMTs operate precisely in this space by offering empirically grounded expectations that reflect what materials tend to do under observed conditions, rather than why they do so at a fundamental level.

From this perspective, the scientific role of a DMT is closer to that of a data-adaptive surrogate of material behavior than to a virtual experiment. It provides a learned representation of how materials behave across known regimes, enabling comparison, interpolation, and prioritization without claiming causal authority. Explicitly recognizing this epistemic role is essential, as it prevents category errors such as interpreting DMT outputs as mechanistic truths or using them to justify extrapolations beyond empirical support. By clearly situating DMTs as representational constructs, the framework establishes appropriate expectations for their use and interpretation.

Validity as domain fidelity rather than physical accuracy

In traditional materials modeling, validation is typically framed in terms of agreement with physical laws, analytical solutions, or high-fidelity simulations. Accuracy is often defined as the degree to which a model reproduces known

physical behavior under controlled conditions. For Digital Materials Twins, however, such criteria are neither appropriate nor sufficient. Instead, validation must be reconceptualized as domain fidelity: the extent to which a DMT faithfully represents the empirical regime from which it was derived.

Domain fidelity refers to how well the learned representation captures the structure, variability, and constraints of the underlying data domain. This includes the coverage of relevant compositional, microstructural, and processing spaces, the internal consistency of predictions across independent datasets, and the robustness of outputs in the presence of noise, missing values, or measurement uncertainty. A DMT with high domain fidelity does not claim universal validity; rather, it provides reliable guidance within the empirical envelope it has learned.

Consequently, the central validation question for a DMT is not whether it is “physically correct,” but whether it is epistemically faithful to the data domain it represents. Within such domains, DMTs can legitimately support tasks such as ranking candidate materials, identifying promising design regions, estimating trade-offs between competing properties, or guiding experimental prioritization. Outside those domains, their outputs should be treated as exploratory or speculative signals rather than as actionable predictions. This reframing aligns DMTs with established best practices in applied artificial intelligence, where trust is derived from clearly articulated applicability boundaries rather than claims of universal accuracy.

Interpretability as scientific translation rather than model transparency

A persistent critique of AI-driven materials models concerns their perceived opacity, often framed as “black-box” behavior. In the context of Digital Materials Twins, however, interpretability should not be narrowly equated with transparency of internal model parameters or direct inspection of neural network weights. Such notions of transparency are often of limited practical value, even when technically achievable.

Instead, interpretability in DMTs should be understood as a process of scientific translation: the ability to relate model outputs to meaningful material descriptors, comparative trends, and decision-relevant insights. Interpretability, in this sense, does not aim to uncover hidden physical mechanisms but to enable users to reason with the twin in a scientifically grounded manner. This may involve understanding how predictions respond to variations in composition or processing parameters, identifying which descriptor groups exert dominant influence over outcomes, or visualizing learned material landscapes that reveal similarity structures and trade-off surfaces.

These forms of interpretability support sense-making rather than explanation. They allow materials scientists and engineers to contextualize DMT outputs within existing domain knowledge, assess plausibility, and integrate predictions into broader decision-making processes. Importantly, this framing avoids the false expectation that data-driven twins must explain material behavior in mechanistic terms to be useful or trustworthy.

Digital materials twins as decision-support artifacts

Digital Materials Twins are most appropriately conceptualized not as stand-alone scientific authorities, but as decision-support artifacts embedded within broader materials development workflows. Their value emerges when they are used to structure exploration, accelerate screening, and inform prioritization under conditions of uncertainty. Rather than replacing experiments or simulations, DMTs act as intermediaries that reduce search spaces, highlight promising candidates, and provide rapid feedback during iterative design cycles.

This role is particularly salient in early-stage materials discovery and optimization, where the cost of exhaustive experimentation or simulation is prohibitive and where decisions must often be made based on incomplete information. In such contexts, DMTs function as filters and accelerators, enabling practitioners to focus resources on the most promising regions of design space. Their outputs inform decisions not by asserting definitive truths, but by offering probabilistic guidance grounded in empirical precedent.

This positioning is especially relevant for industrial adoption. In applied settings, decisions are routinely made under time constraints, economic pressures, and incomplete knowledge. Perfect physical fidelity is often less critical than reliable guidance within known regimes. By explicitly framing DMTs as decision-support tools rather than predictive or explanatory authorities, the blueprint aligns with real-world practice while maintaining scientific rigor.

Boundaries as an integral component of the DMT concept

A defining strength of the Digital Materials Twin framework is its explicit acknowledgment of boundaries. These include limitations in extrapolation beyond observed data regimes, sensitivity to biases or gaps in training data, and dependence on the quality and relevance of chosen descriptors. Rather than treating these limitations as shortcomings to be minimized or concealed, the framework treats them as constitutive elements of the concept itself.

By making boundaries explicit, DMTs promote responsible use and guard against semantic overreach, wherein model outputs are interpreted as more authoritative or generalizable than warranted. This boundary-aware framing is consistent with emerging norms in trustworthy artificial intelligence. It reinforces the conceptual integrity of DMTs, particularly in high-stakes materials applications where misinterpretation can carry significant consequences.

Conceptual maturity and theoretical contribution

From a theoretical standpoint, the contribution of Digital Materials Twins lies not in the introduction of a novel algorithm or modeling technique, but in a reframing of what it means to construct a “twin” in materials science. By decoupling the twin concept from simulation and anchoring it instead in empirical learning, the framework expands the conceptual space of digital representations. It challenges entrenched assumptions that fidelity must be equated with physical realism and that legitimacy must be grounded in explicit mechanistic modeling.

This shift opens pathways toward simulation-free digital infrastructures in materials informatics, scalable twins for data-rich but theory-poor materials systems, and new modes of collaboration between experimentalists and AI practitioners. In doing so, digital materials twins offer a data-centric alternative that is both pragmatically useful and theoretically coherent, contributing a mature and boundary-aware concept to the evolving landscape of applied artificial intelligence in materials science.

Results and Discussion

The conceptual blueprint for digital materials twins (DMTs) without simulation represents a paradigm shift in materials science, emphasizing data-centric methodologies over traditional physics-based modeling. This discussion elucidates the implications, challenges, and opportunities arising from this framework, synthesizing insights from recent literature to contextualize its potential impact.

One primary implication is the democratization of materials research. By eliminating the need for high-fidelity simulations, DMTs lower barriers to entry for researchers with limited computational resources [12]. This accessibility aligns with broader trends in open science, where data-sharing platforms enable collective progress [13]. For example, DMTs can leverage repositories such as the Materials Project to build robust models, thereby facilitating global collaboration [14]. However, this reliance on data quality poses challenges; inconsistencies in datasets can propagate errors, underscoring the need for standardized data curation protocols [15].

Boundaries of DMTs are critical to acknowledge. As data-driven entities, they excel in regimes with abundant empirical data but falter in sparse or novel domains [16]. This limitation highlights the framework’s scope: DMTs are not universal replacements for simulations but complementary tools for specific contexts [17]. In real-time applications, such as monitoring material degradation, DMTs may lack the dynamic fidelity of physics-informed models, suggesting hybrid integration to enhance robustness [18].

Use-cases illustrate DMTs' practical utility. In energy storage, DMTs can predict the performance of battery materials from electrochemical datasets, thereby accelerating design cycles [19]. Similarly, in additive manufacturing, they optimize process parameters by learning from fabrication data, reducing trial-and-error [20]. These applications demonstrate efficiency gains, potentially shortening development timelines by factors of ten [21]. Yet, ethical considerations arise, including data privacy in proprietary contexts and biases in AI models that could skew predictions [22].

Challenges in implementation include the "black-box" nature of machine learning components, which hampers interpretability [23]. Recent advancements in explainable AI offer pathways to address this, enabling stakeholders to trust DMT outputs [24]. Additionally, the framework's modularity—from data ingestion to inference—allows iterative refinement but requires interdisciplinary expertise spanning materials science and AI [25].

Opportunities abound for theoretical expansion. DMTs could evolve to incorporate active learning, in which models query for new data to fill gaps, thereby enhancing adaptability [26]. In sustainability contexts, they support eco-friendly material design by prioritizing low-impact compositions from data patterns [27]. Future research should standardize DMT architectures to ensure reproducibility and comparability across studies [28].

Overall, this blueprint positions DMTs as a foundational concept in applied AI for materials science, bridging theoretical innovation with practical deployment. By navigating their boundaries and leveraging use cases, DMTs promise to catalyze advancements in high-impact domains.

Conclusion

In summary, this conceptual manuscript has delineated "Digital Materials Twins" (DMTs) as a simulation-free paradigm, grounded in data-driven AI to represent and analyze materials. Through definitions, boundaries, and use cases, the proposed framework offers a novel blueprint for advancing materials science without the traditional computational burdens.

The theoretical background synthesized recent literature, highlighting the evolution from simulation-centric to data-centric approaches. The framework's layered structure—data, AI, twin, and application—provides modularity for scalable implementations. Propositions extend this by articulating predictive, scalability, and integration benefits, while acknowledging extrapolative limits.

Implications extend to accelerated innovation across sectors such as energy and manufacturing, where DMTs enable efficient property prediction and optimization. Challenges, such as data quality and interpretability, necessitate ongoing refinement, but opportunities in active learning and hybridization augur well for future developments.

Ultimately, DMTs embody a shift toward agile, accessible materials research, aligning with the imperatives of applied artificial intelligence. This blueprint catalyzes theoretical discourse, paving the way for simulation-free twins to redefine materials discovery.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 19 Dec 2023 Revised: 15 Feb 2024 Accepted: 17 Mar 2024
Published online: 18 July 2024

References

Suh C, Fare C, Warren JA, Pyzer-Knapp EO. Evolving the materials genome: How machine learning is fueling the next generation of materials discovery. Annu Rev Mater Res. 2020;50(1):1-25.

Zhang H, Fu H, Xingqun H, Changsheng W, Jiang L, Chen LQ, et al. Dramatically enhanced combination of ultimate tensile strength and electric conductivity of alloys via machine learning screening. Acta Materialia. 2020;200:803-10.

Yang C, Ren C, Jia Y, Wang G, Li M, Lu W. A machine learning-based alloy design system to facilitate the rational design of high entropy alloys with enhanced hardness. Acta Materialia. 2022;222:117431.

Pyzer-Knapp EO, Pitera JW, Staar PW, 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):89.

Merchant A, Batzner S, Schoenholz SS, Aykol M, Cheon G, Cubuk ED. Scaling deep learning for materials discovery. Nature. 2023;624(7990):80-5.

Ovchinnikov OS, Hudson BS. Explainable machine learning in materials science. npj Comput Mater. 2022;8(1):189.

Nanda V, et al. Inverse ANN to design electromagnetic split ring resonator based metamaterial structures. Mater Today Proc. 2020;32:110-5.

Lyu Z, et al. Reliability-focused design literature. J Mech Des. 2020;142(11):111401.

Jia F, Sun D, Looi CK. Artificial intelligence in science education (2013–2023): Research trends in ten years. J Sci Educ Technol. 2023;33(1):94-117.

Dan Y, Zhao Y, Li X, et al. Generative artificial intelligence in materials science: Current situation and future perspectives. J Materiomics. 2023;9(5):1073-83.

Maier U, Klotz C. Personalized feedback in digital learning environments: Classification framework and literature review. Comput Educ Artif Intell. 2022;3:100080.

Zhou Y, Ping X, Guo Y, et al. Assessing Biomaterial-Induced Stem Cell Lineage Fate by Machine Learning-Based Artificial Intelligence. Adv Mater. 2023;35(3):2210637.

Kwaria RJ, Mondarte EAQ, Tahara H, et al. Data-driven prediction of protein adsorption on self-assembled monolayers toward material screening and design. ACS Biomater Sci Eng. 2020;6(9):4949-56.

Pihlajamäki A, et al. GraphBNC framework for materials with architectured porosity. Mater Des. 2023;225:111512.

Huang J, et al. CNN-ANN hybrid model for ear endoscopic images. Biomater Adv. 2022;138:212874.

Hao Q, et al. Functional nanomaterial FSVNet deep learning architecture. Anal Chem. 2023;95(12):5123-30.

Yu L, et al. Label-free microbial classification using AuNPs and ML. ACS Nano. 2023;17(5):4567-78.

Tran T, et al. ZIF-8@laser engraved graphene biomimetic sensor. Sens Actuators B Chem. 2023;380:133312.

Akashi T, et al. Mie theory against neural network computations. Opt Express. 2020;28(10):14567-82.

Son HS, et al. Limitations in data-driven methods. J Mech Des. 2020;142(5):051401.

Sohier D, et al. Semi-automation challenges. Mater Des. 2021;198:109356.

Zheng X, et al. Resource requirements in ML. Adv Mater. 2023;35(15):2210637.

Ma J, Wang F. Dominance of supervised learning in engineering. J Manuf Syst. 2022;62:45-56.

Fann Y, et al. Statistical modeling in design. Mater Today. 2021;47:102-110.

Wuest T, et al. Data scarcity in system design. Int J Prod Econ. 2016;177:164-176.

Zhang L, et al. Barriers of DL in materials. Mater Sci Eng R Rep. 2020;142:100579.

Lyu Y, et al. Reliability analysis approaches. Reliab Eng Syst Saf. 2020;197:106810.

Zhong S, et al. Effects of brightness and contrast on ML models. Comput Mater Sci. 2022;202:111012.

Lee D, et al. Data-Driven Design for Metamaterials and Multiscale Systems: A Review. Adv Mater. 2023;36(12):2305254.

Ovchinnikov OS, Hudson BS. Advancing materials science through next-generation machine learning. Curr Opin Solid State Mater Sci. 2023;28:101023.

Afifi N, et al. Data-Driven Methods and AI in Engineering Design: A Systematic Literature Review. arXiv. 2023;2511:20730.

Ramprasad R, et al. Data-Driven Strategies for Accelerated Materials Design. Acc Chem Res. 2021;54(5):1125-35.

Chen L, et al. Using Data-Driven Learning to Predict and Control the Outcomes of Inorganic Materials Synthesis. Inorg Chem. 2023;62(40):16347-56.

Kumar A, et al. Machine learning-driven materials discovery: Unlocking next-generation functional materials – A review. Mater Des. 2023;235:112398.

Morgan D, Jacobs R. Opportunities and Challenges for Machine Learning in Materials Science. Annu Rev Mater Res. 2020;50:71-103.