REVIEW **Open access**

Temporal Generalization in Materials AI — What We Know About Model Aging and Drift

Ravi Menon[1*], Arjun Nair[1], Meera Pillai[2], Suresh Varma[1]

Abstract

Artificial intelligence is rapidly reshaping materials science by accelerating property prediction, synthesis planning, and materials design. Yet most AI models for materials are developed and validated under implicit stationary assumptions, while real deployments unfold in time-varying environments where materials, sensors, and processes evolve. This review synthesizes what is currently known about temporal generalization in materials AI—the capacity of models to remain reliable as data distributions and underlying mechanisms change. We distinguish two dominant degradation pathways: drift, in which input statistics or input–output relationships shift over time, and model aging, in which learned representations become obsolete as systems evolve. Drawing on evidence across biosensing and wearables, electrochemical energy storage, polymer synthesis, automated laboratories, and industrial manufacturing, we summarize how temporal failures arise, how they are detected, and why they often remain silent until performance drops become consequential. We then evaluate mitigation strategies—including domain adaptation, incremental and continual learning, active data acquisition, uncertainty-aware prediction, and human–AI feedback loops—highlighting where they succeed, where they break down, and the constraints that limit their scalability in real-world settings. Finally, we identify key gaps: limited longitudinal datasets, weak standardization of temporal evaluation protocols, underexplored multimodal temporal fusion, and insufficient emphasis on prevention rather than detection. We conclude with a forward agenda for resilient materials AI built around lifecycle monitoring, benchmarkable temporal stress tests, and hybrid frameworks that integrate mechanistic knowledge with adaptive learning to sustain reliability over time.

Keywords Materials AI, Temporal generalization, Model aging, Drift, Concept drift, Data shift

*Correspondence:
Ravi Menon
ravi.menon@outlook.com
[1] Department of Materials Engineering and AI Modeling, Faculty of Engineering, University of Kerala, Thiruvananthapuram, India
[2] Department of Intelligent Materials Analytics, Faculty of Engineering, Manipal University, Manipal, India

Introduction

The advent of artificial intelligence (AI) in materials science—widely referred to as *materials informatics*—has inaugurated a transformative paradigm for discovering, characterizing, and optimizing materials. By leveraging large-scale datasets, high-throughput simulations, and advanced statistical learning architectures, AI systems now augment and, in some cases, partially automate scientific reasoning across the materials lifecycle. Machine learning (ML) and deep learning approaches have demonstrated remarkable capability in predicting thermodynamic stability, electronic structure, mechanical performance, and functional properties, thereby reducing reliance on time-intensive experimentation and computationally expensive first-principles simulations [1-8]. Beyond property prediction, AI has reshaped synthesis planning, inverse materials design, and process optimization, enabling accelerated innovation in domains such as energy storage, catalysis, structural composites, smart sensors, and advanced manufacturing systems.

Recent methodological advances have further expanded the exploratory capacity of materials AI. Graph neural networks (GNNs), for instance, encode atomistic and crystallographic relationships directly into learning architectures, allowing models to infer structure–property relationships with unprecedented fidelity. In parallel, generative frameworks—including variational autoencoders, generative adversarial networks, and diffusion-based models—have enabled inverse design, in which candidate materials are algorithmically proposed to meet predefined performance criteria. Such approaches have facilitated accelerated screening of battery electrodes, solid electrolytes, and polymer systems, enabling traversal of chemical spaces that would be infeasible through conventional combinatorial experimentation alone [2, 9-14]. Collectively, these developments position AI not merely as an analytical tool but as an epistemic engine capable of reshaping the production of material knowledge.

Despite these advances, the dynamic, temporally evolving nature of materials systems poses foundational challenges for AI deployment. Unlike static benchmark datasets common in other machine learning domains, materials data are inherently time-dependent. Properties evolve through degradation, phase transitions, environmental exposure, cyclic loading, corrosion, and microstructural aging. Manufacturing ecosystems likewise exhibit temporal variability arising from equipment wear, feedstock heterogeneity, calibration drift, and operational fluctuations. Consequently, the statistical distributions governing model inputs and outputs are rarely stationary. When AI systems trained on historical datasets encounter temporally shifted conditions, predictive reliability may deteriorate—sometimes subtly, sometimes catastrophically [1, 7, 15-21].

This temporal instability foregrounds the concept of temporal generalization—the capacity of a model to sustain performance when extrapolated across time or across evolving operational regimes. Temporal generalization extends beyond conventional out-of-sample validation by introducing diachronic uncertainty: models must not only generalize across unseen compositions or structures but also across future states of matter and process conditions [10, 12, 16, 18, 19, 22, 23]. Closely related is the phenomenon of model aging, wherein predictive accuracy degrades as embedded parameterizations become misaligned with current system realities. Over time, models trained on legacy datasets may encode obsolete correlations, leading to systematic bias or diminished sensitivity to emergent phenomena.

Within the machine learning literature, these degradations are frequently conceptualized through the lens of drift. *Data drift* denotes shifts in the statistical distribution of input variables—for example, evolving microstructural descriptors or compositional ranges—while *concept drift* refers to changes in the functional relationship between inputs and outputs, such as altered degradation pathways or emergent failure mechanisms [18, 20, 22, 23]. In materials contexts, concept drift may arise when new synthesis routes yield microstructures whose performance drivers differ fundamentally from those of historical analogues. These layered temporal effects complicate validation, benchmarking, and deployment, particularly in safety-critical or high-cost industrial applications.

The implications of temporal instability are far-reaching. In energy storage systems, electrochemical aging alters ion transport pathways, interfacial stability, and capacity retention, rendering static predictive models progressively unreliable [2, 4]. In structural alloys, prolonged thermal or mechanical exposure can induce microstructural coarsening, phase precipitation, or defect accumulation, shifting mechanical response profiles beyond trained regimes [7, 8]. Manufacturing environments introduce additional variability: tool wear, thermal cycling, and process recalibration can induce distributional shifts that invalidate quality-prediction models if left unmonitored [18, 20, 22]. These realities underscore that materials AI systems do not operate in temporally neutral environments; rather, they are embedded within evolving physico-chemical and industrial ecosystems.

Addressing such challenges requires methodological and conceptual adaptation. Strategies including incremental learning, transfer learning, continual retraining, and adaptive sampling have emerged as mechanisms to maintain model relevance under shifting data regimes. Complementary approaches—such as uncertainty quantification, drift-detection algorithms, and temporal validation protocols—aim to identify when models are approaching epistemic obsolescence [17, 19, 24-26]. Yet, despite growing technical attention, the literature remains fragmented across application domains, with limited synthesis of how temporal phenomena manifest differently across materials classes and operational contexts.

Against this backdrop, the objectives of this review are threefold. First, it provides a structured background on temporal challenges in materials AI, integrating insights from sensor-driven biosensing platforms [1, 6], energy materials systems [4, 5], and predictive maintenance infrastructures [16, 18]. Second, it synthesizes current understanding of model aging and drift through a thematic analysis of empirical and computational case studies spanning polymers [10, 27], alloys [7, 8], and health-monitoring technologies [6, 15]. Third, it identifies methodological gaps and future research directions to enhance temporal robustness, with particular emphasis on continual learning architectures, adaptive data governance, and probabilistic uncertainty modeling [17, 19, 24, 26].

By focusing on peer-reviewed journal literature published, this review consolidates dispersed advances into a coherent interpretive framework. It argues that temporal robustness must be treated not as a secondary optimization parameter but as a foundational design principle in materials AI. As the field advances toward autonomous laboratories, closed-loop discovery systems, and real-time industrial deployment, ensuring that AI models remain reliable under temporally evolving conditions will be essential to sustaining scientific validity, industrial trust, and translational impact.

Main Text

Foundations of temporal generalization in materials AI

Temporal generalization constitutes a cornerstone of robust artificial intelligence deployment in materials science, where data-generating processes are intrinsically time-dependent rather than stationary. Unlike canonical machine learning benchmarks that assume fixed statistical distributions, materials datasets evolve in response to physicochemical transformations, operational stresses, and environmental perturbations. Structural degradation, oxidation, fatigue accumulation, and microstructural coarsening introduce gradual yet consequential shifts in measured properties. External influences—including temperature gradients, humidity exposure, pressure cycling, and radiation environments—further modulate materials behavior over time, producing data distributions that diverge from those represented in training corpora [1, 5, 7].

Biosensing platforms provide a salient illustration of this temporal sensitivity. Sensor materials, particularly polymeric and nanostructured substrates, exhibit signal drift due to biofouling, surface degradation, and environmental interference. Such drift can compromise AI-enabled diagnostic predictions when the calibration states encoded during training no longer align with operational realities. Domain-adaptation architectures have therefore emerged as critical recalibration tools, enabling models to align latent feature spaces between historical and real-time sensor conditions [1]. Complementarily, the integration of temporal features—through recurrent neural networks, long short-term memory systems, and transformer-based sequence encoders—has demonstrated improved predictive stability by capturing evolving signal trajectories rather than static snapshots [6, 15, 16].

Energy storage materials present a more structurally complex temporal landscape. Lithium-ion battery systems, for instance, undergo progressive electrochemical aging characterized by growth of the solid–electrolyte interphase, lithium plating, cathode phase transitions, and electrolyte decomposition. AI models tasked with forecasting capacity retention or failure onset must therefore generalize across extended cycling horizons. Generative modeling and graph neural network frameworks have shown promise in predicting degradation pathways and lifetime performance. However, these systems frequently encounter drift when extrapolated to novel chemistries, atypical duty cycles, or unobserved thermal regimes [2, 4].

A critical insight emerging from recent work is the value of integrating multi-scale modeling. By coupling quantum mechanical simulations, atomistic modeling, and continuum-scale machine learning, hybrid frameworks can simulate long-horizon degradation behaviors with greater physical grounding. Such integration enhances temporal robustness by embedding mechanistic priors into data-driven predictions [2, 12]. Nonetheless, limitations persist when training datasets lack compositional diversity or operational breadth, constraining extrapolative reliability and reinforcing the risk of temporal overfitting [8, 9].

Manufacturing ecosystems introduce yet another dimension of temporal complexity. Predictive maintenance systems rely on machine learning models trained on historical equipment performance data. Over time, however, tool wear, lubrication breakdown, vibration pattern evolution, and calibration drift alter process signatures. These shifts often manifest as concept drift, degrading predictive accuracy if models remain static. Adaptive machine learning strategies—such as residual monitoring in time-series architectures and online parameter updating—have been proposed to dynamically recalibrate predictive systems in response to evolving operational states [19, 22].

Collectively, these domain-specific manifestations underscore that temporal generalization is not merely a performance metric but a systems-level requirement. Robust materials AI systems must not only generate accurate predictions but also detect, interpret, and adapt to temporally emergent shifts in the data landscape [17, 18, 21].

Mechanisms of model aging in materials systems

Model aging refers to the progressive erosion of predictive validity as machine learning systems become temporally misaligned with the phenomena they were designed to represent. In materials informatics, this degradation is exacerbated by the physical aging processes inherent to material systems. As materials evolve, the statistical relationships encoded within trained models may lose explanatory and predictive fidelity.

Polymeric systems exemplify this dual aging dynamic. Variability in monomer purity, crosslink density, curing kinetics, and additive dispersion can introduce compositional drift across synthesis batches. Over time, such variability propagates into performance deviations—mechanical, thermal, or electrical—that legacy models fail to anticipate. Empirical studies demonstrate that without retraining or adaptive recalibration, predictive accuracy deteriorates as polymer production landscapes evolve [10, 27].

Self-driving laboratories offer a partial mitigation pathway. By embedding closed-loop experimentation, automated synthesis, and real-time model retraining, these platforms create human–AI symbiotic feedback systems. Experimental outcomes continuously update model parameters, reducing epistemic lag between prediction and reality. This iterative retraining architecture attenuates model aging by ensuring representational alignment with current experimental regimes [10].

Battery systems again provide a high-resolution view of aging interactions. Capacity fade emerges from layered electrochemical phenomena influenced by temperature gradients, charge–discharge rates, and interfacial instabilities. AI models trained on early-life performance data often fail to account for late-stage degradation accelerants. Dynamic electrochemical impedance spectroscopy has enabled real-time tracking of degradation signatures, yet predictive systems must incorporate uncertainty quantification to avoid overconfident forecasts under aging conditions [4, 17].

Wearable and implantable sensing platforms introduce additional aging pathways. Material fatigue, biofluid exposure, and mechanical deformation degrade signals over prolonged deployment. Incremental learning frameworks—where models assimilate new labeled data streams without full retraining—have proven effective in sustaining predictive accuracy under such conditions [6, 14, 15].

In alloy informatics, model aging manifests through innovation itself. As novel alloy compositions and processing routes emerge, legacy datasets become less representative of current design spaces. Explainable AI methodologies have been deployed to detect aging signals by tracking shifts in feature importance, latent embeddings, and decision boundaries [7–9]. These interpretive diagnostics reveal when models are relying on obsolete structure–property correlations.

Ultimately, model aging arises from the intersection of data scarcity, environmental variability, and innovation-driven distributional expansion. Mitigation strategies—including active learning, targeted data acquisition, and continual retraining—serve as mechanisms to refresh model knowledge and sustain epistemic relevance over time [24, 26].

Concept drift and data shift in materials informatics

Concept drift represents one of the most consequential temporal challenges in deployed materials AI systems. It occurs when the functional mapping between input variables and predicted outputs evolves, even if input distributions remain partially stable. In materials science, such drift often reflects emergent physical mechanisms rather than mere statistical noise.

Injection molding processes illustrate this phenomenon vividly. Equipment aging alters thermal gradients, injection pressures, and cooling dynamics, thereby modifying the causal pathways linking process parameters to product quality. Machine learning models trained on early operational states may therefore produce systematically biased predictions as manufacturing ecosystems evolve [23]. Hybrid-feature autoencoders and deep drift detection frameworks have been developed to identify latent distributional deviations and trigger automated model updates [23, 27].

Pharmaceutical quantitative structure–activity relationship (QSAR) modeling offers a parallel case. Temporal expansion of chemical libraries introduces new structural motifs that alter activity relationships. Studies indicate that ensemble learning architectures can partially quantify predictive uncertainty under such drift, yet real-world variability continues to challenge calibration reliability [17].

Energy materials systems experience concept drift through operational heterogeneity. Variations in cycling regimes, temperature exposures, and load profiles reshape degradation kinetics and performance trajectories. Adaptive optimization frameworks—particularly those that leverage Bayesian inference—have been proposed to recalibrate predictive models as operational states evolve [2, 12, 24].

Industrial manufacturing data streams further illustrate the ubiquity of drift. Continuous sensor monitoring produces high-velocity temporal data that is susceptible to regime shifts. Fuzzy clustering algorithms, support vector machine adaptation layers, and hybrid online–offline learning systems have been deployed to detect and accommodate such changes in real time [19, 20, 22].

Taken together, these examples demonstrate that drift detection is indispensable for sustaining temporal generalization. Without systematic monitoring, predictive systems risk silent performance degradation, undermining reliability in both scientific and industrial decision-making contexts [18, 19, 21].

Mitigation strategies for drift and aging

Addressing temporal degradation in materials AI requires integrative mitigation architectures that combine statistical adaptation, data governance, and interpretability frameworks.

Incremental learning represents one of the most widely adopted approaches. By assimilating newly acquired data without discarding prior knowledge, incremental models maintain relevance under evolving conditions. In polymer process monitoring, such approaches have demonstrated substantial reductions in error—reportedly decreasing RMSE by over 40% following the detection of drift events [23, 27].

Active learning complements this paradigm by selectively acquiring the most informative data points for retraining the model. Rather than uniformly sampling new observations, active learners prioritize out-of-distribution or high-uncertainty instances. This strategy has proven effective in mechanical systems and materials design contexts, where targeted experimentation enhances temporal generalization efficiency [15, 21].

Uncertainty quantification provides an additional safeguard. Bayesian neural networks, probabilistic graphical models, and ensemble predictors estimate predictive confidence, enabling systems to flag unreliable outputs under drift conditions. Such frameworks have demonstrated value in pharmaceutical modeling and energy materials forecasting, where risk-sensitive decisions demand calibrated reliability [2, 17].

Explainable AI (XAI) tools further support aging diagnostics. Techniques such as SHAP (Shapley Additive Explanations) reveal evolving feature attributions, exposing when models begin to rely on spurious or outdated correlations. These interpretive signals can trigger retraining cycles or data audits [9].

Domain adaptation remains particularly critical in sensing and wearable materials systems. By aligning feature distributions across temporal domains, adaptation layers recalibrate predictive models without requiring full retraining pipelines [1, 6, 7]. At the systems level, lifecycle frameworks such as extended CRISP-DM pipelines now incorporate operational monitoring phases, embedding drift detection and retraining protocols directly into deployment infrastructures [18, 23]. A systems-level synthesis of these temporal shift drivers, degradation pathways, diagnostic signals, and adaptive mitigation mechanisms is presented in **Figure 1**, which conceptualizes temporal robustness as a layered lifecycle infrastructure spanning detection, intervention, and governance.

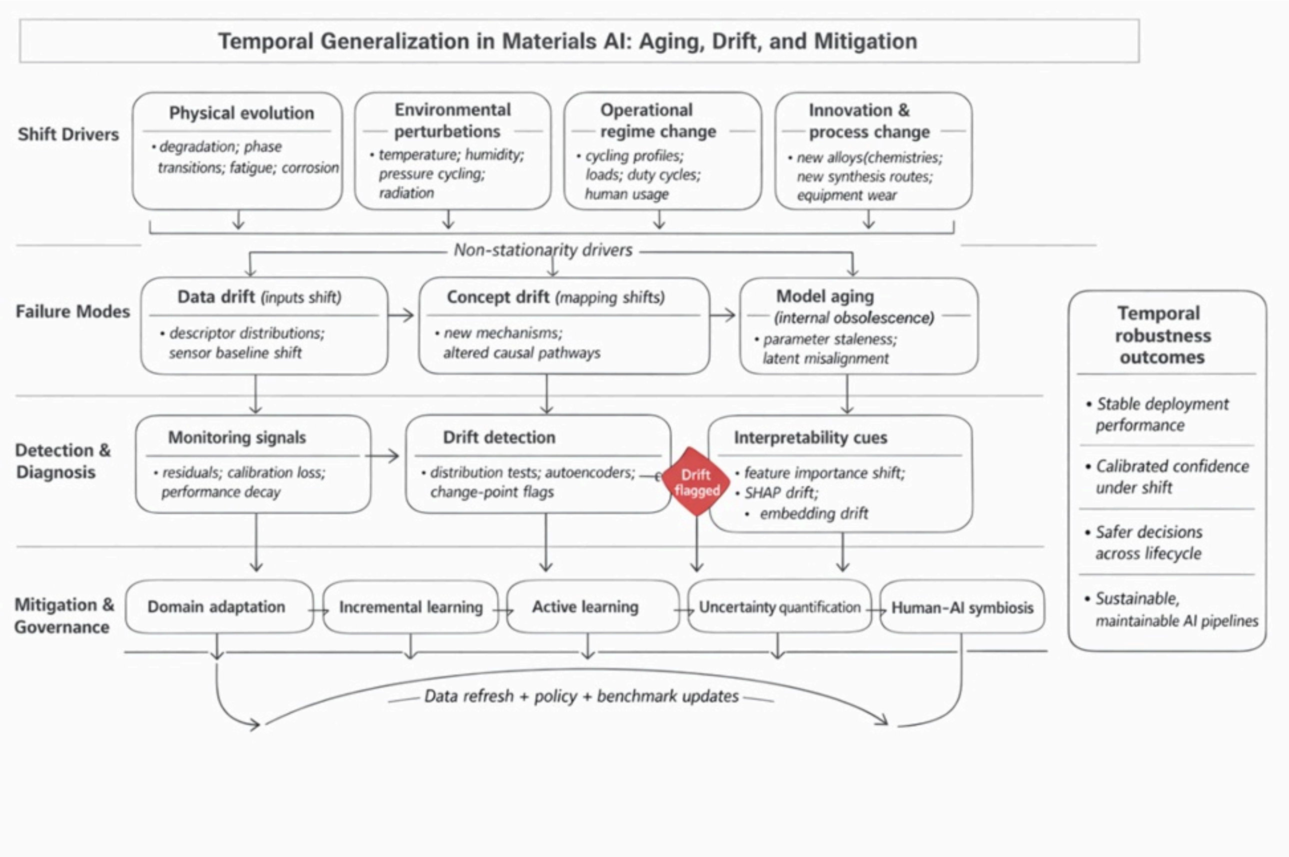

Figure 1. Temporal robustness stack in materials AI: shift drivers, failure modes, detection signals, and mitigation pathways. The schematic integrates upstream non-stationarity drivers (e.g., physical degradation, environmental perturbations, operational variability), downstream failure manifestations (data drift, concept drift, model aging), detection and interpretability signals, and adaptive mitigation strategies including domain adaptation, continual learning, uncertainty quantification, and human–AI governance. The framework visualizes temporal generalization as a lifecycle systems challenge rather than an isolated modeling limitation.

Empirical case studies across materials domains provide grounded insight into how temporal robustness challenges manifest and how adaptive AI frameworks are operationalized in practice. These applied implementations illuminate both the feasibility and limitations of drift-aware, aging-resilient materials intelligence systems.

In biosensing ecosystems, temporal instability is particularly pronounced because sensing substrates are sensitive to environmental perturbations. AI-powered drift compensation architectures have enabled the translation of laboratory prototypes into commercially viable diagnostic platforms. For example, adaptive calibration models embedded within olfactory and electrochemical biosensors dynamically realign prediction layers in response to humidity fluctuations, biochemical fouling, and substrate degradation. Such systems integrate domain-adaptation modules that recalibrate latent representations without interrupting operational continuity, thereby sustaining diagnostic reliability in real-world deployment settings [1].

Energy storage diagnostics offer another instructive case. Lithium-ion battery monitoring systems increasingly employ dynamic electrochemical impedance spectroscopy (DEIS) integrated with machine learning analytics. These hybrid frameworks capture time-resolved electrochemical signatures, enabling real-time tracking of degradation pathways, such as thickening of the solid–electrolyte interphase and an increase in charge-transfer resistance. By continuously feeding impedance spectra into predictive architectures, AI systems can generate temporally adaptive forecasts of remaining useful life and failure probability, thereby enhancing safety monitoring and lifecycle optimization [4].

Polymer research laboratories have begun operationalizing temporal robustness through human–AI collaborative infrastructures. In automated synthesis environments, machine learning models guide formulation pathways, yet human oversight remains essential for interpreting anomalous outputs induced by compositional drift or processing variability. This human–AI symbiosis introduces reflexive retraining loops, where experimental feedback recalibrates predictive systems in near real time. Such hybrid governance structures have proven effective in mitigating synthesis drift and sustaining predictive fidelity across evolving polymer chemistries [10].

Manufacturing systems present some of the most mature industrial applications of temporally adaptive AI. Predictive quality control and maintenance frameworks employ adaptive machine learning to monitor process parameters, including thermal gradients, injection pressures, vibration signatures, and tool wear. Online learning algorithms and residual-based monitoring systems dynamically update predictive coefficients, enabling sustained accuracy even as operational regimes evolve. These adaptive infrastructures have demonstrated measurable gains in yield stability, defect reduction, and maintenance forecasting across advanced manufacturing environments [18-20].

Collectively, these case studies underscore that temporal robustness is not an abstract modeling objective but a practical engineering requirement. Systems capable of detecting, interpreting, and adapting to time-dependent variation are more likely to achieve translational deployment across sensing, energy, polymer, and manufacturing ecosystems.

Results and Discussion

The synthesis of recent literature reveals that temporal generalization remains a pivotal yet comparatively underexamined dimension of materials artificial intelligence. While predictive accuracy and computational efficiency have historically dominated research priorities, the capacity of AI systems to remain reliable under temporally evolving conditions is emerging as a decisive determinant of real-world viability. Across domains—including biosensing [1, 5, 6], energy storage [2–4, 12], and manufacturing [16, 18, 20-22]—evidence converges on data shift as a principal driver of performance degradation.

Biosensing technologies exemplify this vulnerability. In olfactory and gustatory detection systems, environmental volatility—spanning humidity gradients, chemical interference, and substrate aging—induces progressive signal drift. Biomimetic sensor designs integrated with AI calibration layers have improved adaptive capacity, yet many deployed systems remain constrained by training regimes anchored in temporally narrow datasets [1]. Parallel challenges arise in wearable health monitoring platforms, where physiological variability—circadian rhythms, metabolic fluctuations, hydration levels—introduces temporal heterogeneity into biosignal streams. Machine learning frameworks must therefore incorporate dynamic feature selection and adaptive recalibration mechanisms to maintain diagnostic accuracy over extended deployment horizons [5, 6, 17].

Model aging emerges as an especially critical concern in long-duration monitoring infrastructures. Battery safety management systems illustrate how predictive architectures degrade when parameterizations fail to incorporate evolving degradation kinetics. Although dynamic electrochemical impedance spectroscopy enables high-resolution tracking of electrochemical aging, predictive models lacking incremental retraining pipelines remain vulnerable to parameter obsolescence and latent variable omission [3, 4, 23].

In physical metallurgy and alloy informatics, temporal challenges manifest differently. Data-driven mapping of alloy performance has achieved significant milestones; however, static predictive frameworks struggle to accommodate emergent alloy chemistries and novel processing routes. The temporal expansion of design spaces introduces distributional novelties that legacy models are ill-equipped to interpret. Explainable AI methodologies have therefore gained traction as diagnostic instruments, enabling researchers to trace how shifts in feature importance and latent embeddings correspond to predictive degradation [8, 9].

Concept drift further complicates deployment landscapes by altering causal relationships rather than merely shifting data distributions. Polymer injection molding systems demonstrate how equipment aging and process recalibration modify the mapping between operational parameters and product quality. Drift-detection frameworks paired with incremental learning have proven essential for maintaining predictive validity in such environments [23]. More broadly, adaptive manufacturing paradigms now embed machine-learning retraining cycles directly into operational workflows, enabling continuous refinement of process-control models [14, 19, 22].

Despite these advances, scalability remains a persistent challenge. Pharmaceutical informatics illustrates how temporal dataset expansion complicates uncertainty quantification in QSAR modeling. Ensemble and probabilistic frameworks can partially detect distributional shifts, yet they frequently produce overconfident predictions when confronted with structurally novel compounds [13]. Similarly, mechanics-focused AI systems highlight the difficulty of out-of-distribution generalization, prompting the development of hybrid architectures that embed physical constraints to stabilize predictive extrapolation under drift conditions [15].

Taken together, these cross-domain insights reveal that temporal robustness is not reducible to a single methodological solution. Rather, it requires a layered integration of adaptive learning, uncertainty modeling, explainability, and domain knowledge. Materials AI systems must evolve from static predictive engines into temporally reflexive infrastructures capable of sensing, diagnosing, and responding to epistemic misalignment over time.

Mitigation strategies designed to address temporal degradation in materials AI—while methodologically promising—exhibit uneven efficacy across application domains. Incremental learning and active learning frameworks, for example, have demonstrated measurable success in closed-loop discovery environments, where models are continuously retrained on newly generated experimental data streams [27]. Similar adaptive architectures underpin drift compensation in electronic nose systems, where real-time recalibration improves classification stability under sensor aging conditions [24]. Yet the scalability of such approaches remains constrained by computational overhead, infrastructure demands, and the availability of high-quality temporally labeled datasets [12, 21].

Bayesian active learning offers an instructive case of both promise and limitation. By probabilistically prioritizing the most informative candidate samples, these frameworks accelerate discovery pipelines while minimizing redundant experimentation. In relatively controlled discovery ecosystems, such approaches have yielded efficiency gains in materials screening. However, in highly dynamic operational environments—such as power electronics, where thermal loads, switching frequencies, and material fatigue evolve continuously—sporadic retraining cycles are insufficient. Embedding AI as a persistent cognitive layer within such systems requires continuous monitoring, adaptive memory architectures, and hierarchical updating protocols rather than episodic learning interventions [4, 11].

Hybrid governance models, particularly those centered on human–AI symbiosis, introduce an alternative mitigation paradigm. In polymer research laboratories, robotic synthesis platforms coupled with machine learning optimization

https://doi.org//k629752182

engines create reflexive experimental loops that counteract model aging. Human oversight remains critical for adjudicating anomalous predictions, curating retraining datasets, and embedding domain heuristics into automated workflows [9, 10]. Despite demonstrated efficacy, ethical and operational integration challenges remain unresolved. Questions surrounding accountability, interpretability, and decision authority persist, particularly as laboratory autonomy scales.

Beyond technical mitigation, temporal robustness carries broader ethical and sustainability implications. In AI-enabled drug delivery systems, predictive degradation of biomaterials over time has direct consequences for patient safety and therapeutic efficacy. Temporal miscalibration in such systems can compromise dosage accuracy or release kinetics, underscoring the need for robust lifecycle modeling [10]. Parallel concerns emerge in digital identity proofing infrastructures, where AI systems tasked with biometric verification or fraud detection must adapt to evolving adversarial tactics. Here, model aging is not merely a performance issue but a societal risk vector requiring continuous surveillance and recalibration [19].

The literature also reveals domain biases that shape current understanding of temporal AI robustness. Disproportionate research attention has been directed toward energy materials, polymers, and manufacturing systems, while comparatively limited investigation has addressed temporally dynamic radiological materials, biohybrid systems, or pathogen-responsive biomaterials [21, 25]. This imbalance constrains the generalizability of existing mitigation frameworks and signals the need for broader interdisciplinary expansion.

Several structural research gaps emerge from the synthesis. First, most existing studies emphasize drift detection rather than prevention. While anomaly detection, distributional monitoring, and uncertainty flagging are increasingly sophisticated, fewer frameworks seek to architect models inherently resilient to temporal volatility through anticipatory design principles [17, 18, 23]. Second, limited attention has been devoted to multimodal temporal integration. Materials systems generate heterogeneous data streams—spectroscopic, structural, electrochemical, mechanical—yet most predictive architectures treat these modalities in isolation, constraining holistic temporal modeling [5, 6, 15].

Third, although real-time monitoring infrastructures are widely advocated, practical deployment barriers persist. Continuous sensing, streaming analytics, and retraining pipelines demand computational and financial resources that may be infeasible in low-resource health systems, decentralized manufacturing networks, or emerging industrial economies [20]. These infrastructural asymmetries risk creating temporal robustness divides between technologically mature and resource-constrained environments. **Table 1** summarizes the temporal failure modes, detection signals, and mitigation strategies reported across five key materials AI application areas, highlighting domain-specific shift drivers and the prevailing practical constraints for achieving long-term robustness.

Table 1. Temporal generalization in materials AI: domains, temporal failure modes, and mitigation levers

Domain/application area	Typical temporal shift drivers	Dominant temporal failure mode(s)	Common detection signals	Mitigation strategies reported	I cons
Biosensors and e-nose/wearable sensing [1, 6, 7, 24]	Humidity/temperature variability; biofouling; substrate degradation; user context changes	Data drift + model aging	Baseline drift; calibration loss; rising residuals; embedding shift	Domain adaptation; drift compensation; incremental learning; temporal sequence models [1, 6, 7]	Limi re het depl
Energy storage (Li-ion diagnostics, lifetime prediction) [2–4, 12, 17]	Cycling regime changes; temperature gradients; aging	Model aging + concept drift	Capacity-fade mismatch; impedance pattern change;	DEIS-informed monitoring; hybrid multi-scale + ML; UQ (Bayesian/ensembles);	Data chen hor

	mechanisms evolving over life		uncertainty inflation/overconfidence	periodic retraining [4, 12, 17]	truth ret
Polymers and synthesis/self-driving labs [9, 10, 23, 26, 27]	Batch variability; compositional drift; process parameter instability; new formulations	Concept drift + model aging	Yield/performance decay; feature attribution drift; anomalous experimental outcomes	Closed-loop retraining; active learning; incremental learning; human–AI symbiosis [9, 10, 23, 27]	Ethic rep ac a infras
Alloys and physical metallurgy informatics [7–9]	New alloy classes; evolving processing routes; shifting microstructures under service	Model aging (innovation-driven) + data drift	Feature importance shift; embedding drift; rising OOD rate	Explainable AI diagnostics; transfer learning; targeted data acquisition [8, 9]	B s tem weal eme
Manufacturing and predictive maintenance [16, 18–20, 22, 23]	Tool wear; calibration drift; sensor replacement; operational regime changes	Concept drift (often primary) + data drift	Residual monitoring; change-point flags; quality metric shift	Online/streaming updates; residual-based adaptation; lifecycle monitoring frameworks (CRISP-DM extensions) [18, 23]	Sc r const go upd ris de
Cross-cutting/emerging and underrepresented systems [21, 25]	Radiological/biological exposures; multi-physics coupling; complex environments	Mixed (often uncharacterized)	Sparse or delayed feedback; weak observability	Transfer + UQ + multimodal fusion (proposed more than demonstrated) [21, 25]	Evi th lc dat inte be

Finally, the absence of standardized evaluation benchmarks for temporal generalization represents a critical methodological gap. Unlike computer vision or natural language processing—where temporally stratified datasets and drift challenges are increasingly institutionalized—materials AI lacks shared benchmarking ecosystems. The development of open, longitudinal datasets capturing degradation, drift, and lifecycle variability would enable comparative validation of mitigation strategies and accelerate methodological convergence [8, 13].

In synthesis, while mitigation strategies have advanced considerably, their implementation remains fragmented, domain-specific, and infrastructurally uneven. Achieving durable temporal robustness will require integrative frameworks that unify adaptive learning, anticipatory modeling, ethical governance, and scalable deployment architectures.

Conclusion

This review has examined the evolving landscape of temporal generalization in materials artificial intelligence, establishing that model aging and drift are not peripheral anomalies but foundational challenges embedded within the dynamic ontology of materials systems. Across sensing platforms, energy infrastructures, polymer synthesis ecosystems, and manufacturing environments, temporally induced distributional shifts systematically erode predictive reliability when left unaddressed.

Key findings underscore the ubiquity of concept drift in industrial process monitoring, the prevalence of model aging in long-horizon degradation forecasting, and the operational significance of mitigation strategies such as incremental learning, uncertainty quantification, and adaptive recalibration. These insights collectively affirm that temporal robustness is indispensable for translating AI innovations from controlled laboratory demonstrations into resilient, real-world deployment systems.

Looking forward, several research trajectories emerge as especially consequential.

First, the development of hybrid modeling architectures integrating physics-based simulation with machine learning offers a pathway toward anticipatory robustness. By embedding mechanistic priors—derived from quantum mechanics, thermodynamics, or continuum modeling—into predictive pipelines, such systems may preempt drift rather than merely react to it.

Second, advances in active, transfer, and continual learning hold promise for autonomous adaptation. Materials AI systems capable of self-directed data acquisition, domain transfer, and online retraining will be better equipped to operate within rapidly evolving environments such as biomimetic sensing networks, nanorobotic systems, and adaptive manufacturing lines.

Third, the creation of open temporal datasets represents an infrastructural imperative. Longitudinal materials datasets capturing degradation pathways, environmental exposures, and operational variability would enable benchmarking ecosystems analogous to those that catalyzed progress in other AI domains. Such resources would also foster collaborative standardization across academia, industry, and regulatory bodies.

Fourth, ethical and sustainability considerations must be foregrounded. Temporal bias in predictive systems—where performance varies across lifecycle stages, populations, or environmental contexts—poses risks for equitable deployment. Moreover, the computational demands of continuous retraining raise sustainability questions about energy consumption and the lifecycle carbon impact of AI infrastructure.

Ultimately, confronting temporal uncertainty will determine the durability of materials AI as a scientific and industrial paradigm. Systems capable of sensing their own epistemic obsolescence, adapting to emergent conditions, and integrating human and machine intelligence will define the next generation of resilient discovery platforms.

By advancing temporally robust frameworks, the field moves closer to realizing intelligent materials ecosystems—capable not only of accelerating innovation but of sustaining reliability across the unfolding timescales of real-world operation in energy, healthcare, manufacturing, and beyond.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 11 Feb 2023 Revised: 10 Apr 2023 Accepted: 11 May 2023
Published online: 18 July 2023

References

Khalid Z, Chen Y, Liu X, Noureen B, Chen Y, Wang M, et al. Recent advances and unaddressed challenges in biomimetic olfactory- and taste-based biosensors: Moving towards integrated, ai-powered, and market-ready sensing systems. Sensors (Basel). 2025;25(22):7000.
https://doi.org/10.3390/s25227000.

Wang Y, Zhang X, Li Z, Chen H, Liu J, Zhao Q. Convergence of computational materials science and ai for next-generation energy storage materials. J Electron Mater. 2025;55:45-114.

Du X, Meng J, Xue Z, Amirat Y, Gao F, Benbouzid M. Revolutionizing battery safety: Real-time insights with dynamic electrochemical impedance spectroscopy. ACS Energy Lett. 2025;10(5):2292-304.
https://doi.org/10.1021/acsenergylett.5c00484.

Hinov N. From energy efficiency to energy intelligence: Power electronics as the cognitive layer of the energy transition. Electronics. 2025;14(23):4673.
https://doi.org/10.3390/electronics14234673.

Shi XY, Mishra DK, Özcan ÖÖ, Karahan M, Panpan T, Zhongshan X, et al. Wearable smart sensors integration with ai and machine learning for tracking human health. Biosens Bioelectron X. 2025;27:100711.
https://doi.org/10.1016/j.biosx.2025.100711.

Ming M, Yin X, Luo Y, Zhang B, Xue Q. A point-line-area paradigm: 3d printing for next-generation health monitoring sensors. Sensors (Basel). 2025;25(18):5777.
https://doi.org/10.3390/s25185777.

Klenam DEP, Asumadu TK, Vandadi M, Rahbar N, McBagonluri F, Soboyejo WO. Data science and material informatics in physical metallurgy and material science: An overview of milestones and limitations. Results Mater. 2023;19:100455.
https://doi.org/10.1016/j.rinma.2023.100455.

Oviedo F, Lavista Ferres J, Buonassisi T, Butler KT. Interpretable and explainable machine learning for materials science and chemistry. Acc Mater Res. 2022;3(6):597-607.
https://doi.org/10.1021/accountsmr.1c00244.

Dadfar B, Alemdag B, Kabay G. The alchemist, the scientist, and the robot: Exploring the potential of human-ai symbiosis in self-driving polymer laboratories. Macromol Rapid Commun. 2025;46(23):e00380.
https://doi.org/10.1002/marc.202500380.

Suksaeree J. A review of artificial intelligence (ai)-driven smart and sustainable drug delivery systems: A dual-framework roadmap for the next pharmaceutical paradigm. Sci. 2025;7(4):179.
https://doi.org/10.3390/sci7040179.

Sharma PK, Chen CY. Ai-integrated micro/nanorobots for biomedical applications: Recent advances in design, fabrication, and functions. Biosensors (Basel). 2025;15(12):793.
https://doi.org/10.3390/bios15120793.

Hong S, Liow CH, Yuk JM, Byon HR, Yang Y, Cho E, et al. Reducing time to discovery: Materials and molecular modeling, imaging, informatics, and integration. ACS Nano. 2021;15(3):3971-95.
https://doi.org/10.1021/acsnano.1c00211.

Friesacher HR, Svensson E, Winiwarter S, Mervin L, Arany A, Engkvist O. Temporal distribution shift in real-world pharmaceutical data: Implications for uncertainty quantification in qsar models. Artif Intell Life Sci. 2025;8:100132.
https://doi.org/10.1016/j.ailsci.2025.100132.

Muhr D, Tripathi S, Jodlbauer H. An adaptive machine learning methodology to determine manufacturing process parameters for each part. Procedia Comput Sci. 2021;180:764-71.
https://doi.org/10.1016/j.procs.2021.01.099.

Vlassis NN, Ma R, Sun W. Towards out of distribution generalization for problems in mechanics. Comput Methods Appl Mech Eng. 2023;393:114782.
https://doi.org/10.1007/s00466-023-02282-z.

Çınar ZM, Nuhu AA, Zeeshan Q, Korhan O, Asmael M, Safaei B. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability. 2020;12(19):8211.
https://doi.org/10.3390/su12198211.

Sychugov AA. Monitoring and diagnostics of technological processes with the possibility of dynamic selection of variables. IOP Conf Ser Mater Sci Eng. 2020;971:022069.

Bokrantz J, Subramaniyan M, Skoogh A. Realising the promises of artificial intelligence in manufacturing by enhancing crisp-dm. Prod Plan Control. 2024;35(30):2234-54.
https://doi.org/10.1080/09537287.2023.2234882.

Kim JB. Enhancing personal identity proofing services through ai for a sustainable digital society in korea. Sustainability. 2025;17(23):10486.
https://doi.org/10.3390/su172310486.

Li T, Wandella A, Gomer R, Al-Mafazy MH. Operationalizing health data governance for ai innovation in low-resource government health systems: A practical implementation perspective from zanzibar. Data Policy. 2024;6:e63.
https://doi.org/10.1017/dap.2024.65.

Bapir-Bakr B, Kareem-Ali H, Gutiérrez-Serrano S, Del-Rey-Maestre N, Hernández-Fernández C. Subspace Complexity Reduction in Direction-of-Arrival Estimation via the RASA Algorithm. Sensors (Basel). 2025;25(19):6120.
https://doi.org/10.3390/s25196120.

Tripathi S, Muhr D, Brunner M, Jodlbauer H, Dehmer M, Emmert-Streib F. Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing. Front Artif Intell. 2021;4:576892.
https://doi.org/10.3389/frai.2021.576892.

Shen G, Li S, Zhang Y, Zhou H, Li M. Dynamic monitoring method of polymer injection molding product quality based on operating condition drift detection and incremental learning. Polymers (Basel). 2025;17(22):3025.
https://doi.org/10.3390/polym17223025.

Sun F, Sun R, Yan J. Cross-domain active learning for electronic nose drift compensation. Micromachines. 2022;13(8):1260. https://doi.org/10.3390/mi13081260.

Pan X, Xie Y, Li C, He Y, Zhang Y, Wang Y, et al. Convergence of computational materials science and ai for next-generation energy storage materials. J Electron Mater. 2025;55(1):45-114.

Kusne AG, Yu H, Wu C, Zhang H, Hattrick-Simpers J, DeCost B, et al. On-the-fly closed-loop materials discovery via bayesian active learning. Nat Commun. 2020;11(1):5966. https://doi.org/10.1038/s41467-020-19799-8.

Ngo TD, Kashani A, Imbalzano G, Nguyen KT, Hui D. Additive manufacturing (3d printing): A review of materials, methods, applications and challenges. Compos B Eng. 2018;143:172-96. https://doi.org/10.1016/j.compositesb.2018.02.012.