ORIGINAL RESEARCH Open access

Algorithmic Forgetting as a Design Choice: A Conceptual Analysis of Memory in Materials AI

Maria Hernandez[1*], Carlos Vega[1]

Abstract

The term “forgetting” appears throughout the materials artificial intelligence literature in multiple, often contradictory senses: as a catastrophic failure that destroys previously acquired knowledge of structure–property relations, as an unexamined side effect of data deletion or replay buffer limits, and occasionally as an implicit consequence of model capacity constraints. This conceptual ambiguity impedes precise communication, obscures design decisions, and prevents the field from treating forgetting as a controllable parameter rather than an inevitable defect. The present boundary/definitional paper proposes a precise definition of algorithmic forgetting as a deliberate design choice, distinct from both catastrophic forgetting and passive capacity limits. It distinguishes algorithmic forgetting from five nearby concepts—catastrophic forgetting, data deletion, privacy preservation, capacity saturation, and regularization-induced compression—by clarifying intent, mechanism, epistemic consequences, and reversibility. The paper further articulates the conditions under which forgetting becomes beneficial (adaptation to distribution shift in experimental data streams, selective retention under resource constraints, and controlled deletion for intellectual property or safety) versus harmful (loss of rare but physically valid examples). Finally, it supplies a materials-specific conceptual framework for deciding what to forget and what to retain, grounded in rarity, recency of validation, and relevance to the current search space. By reframing forgetting as an explicit design lever, this analysis offers materials AI practitioners a shared vocabulary and a systematic approach to engineering memory policies that enhance rather than undermine long-term scientific utility.

Keywords Materials informatics, Continual learning, Algorithmic forgetting, Catastrophic forgetting, Design choice, Memory mechanisms

*Correspondence:
Maria Hernandez
maria.hernandez@gmail.com
[1] Department of Materials Science and Machine Learning, Polytechnic University of Valencia, Valencia, Spain

Introduction

The term “forgetting” now permeates the materials of artificial intelligence literature, yet its usage remains strikingly inconsistent [1, 2]. In some contexts, it denotes a catastrophic failure that erases previously learned structure–property mappings when new data arrive, a phenomenon first formalized in connectionist networks and later observed in graph neural networks trained on crystal property prediction [3-5]. In other contexts, the same word describes the deliberate discarding of training examples through replay buffer limits or active data deletion routines. Elsewhere, it appears as an unexamined byproduct of finite model capacity or regularization schemes that compress representations without any explicit decision to forget. This definitional sprawl is not merely terminological; it reflects deeper confusion about whether forgetting is something that happens to a model, something the model does, or something the designer should actively choose [6-8].

The ambiguity carries practical consequences for materials discovery pipelines. When a neural network potential is incrementally updated with new synthesis data, the sudden degradation of accuracy on older polymorphs is routinely labeled catastrophic forgetting and treated as a bug to be patched [4, 5]. Yet the same literature rarely asks whether certain older data—perhaps obtained under superseded measurement protocols—should be intentionally removed rather than retained indefinitely. Similarly, replay buffers are routinely capped in size to fit hardware constraints, but the decision to evict particular examples is seldom justified beyond "to prevent memory overflow" [9-19]. The result is a literature in which forgetting is simultaneously feared, tolerated, and ignored without a shared conceptual vocabulary that would allow designers to treat it as a tunable design parameter.

This study argues that the absence of such a vocabulary impedes progress in materials AI precisely because materials data streams are heterogeneous, noisy, and temporally evolving. Experimental campaigns produce data under changing instruments, evolving synthesis routes, and shifting application targets [7, 20-23]. Retaining every datum forever is neither feasible nor epistemically desirable. At the same time, indiscriminate forgetting risks discarding rare but physically critical examples that lie at the boundary of known phase space [9]. The field, therefore, requires a precise, materials-aware definition of algorithmic forgetting as a deliberate design choice.

Figure 1 synthesizes the conceptual reframing of forgetting by illustrating its transformation from an emergent failure mode into a structured design architecture with explicit policy dimensions.

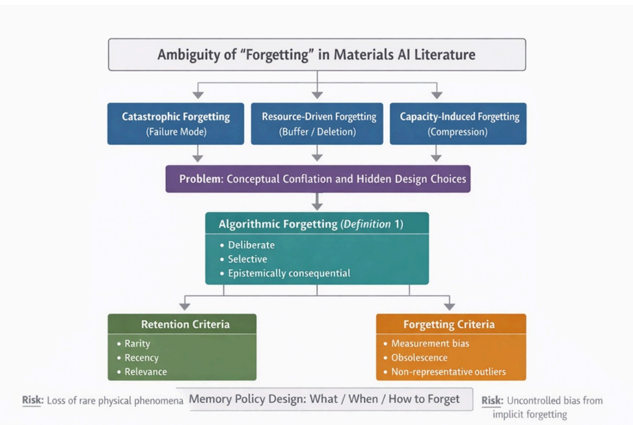

Figure 1. The conceptual reframing of forgetting by illustrating its transformation from an emergent failure mode into a structured design architecture with explicit policy dimensions.

The goal is to supply the boundary conditions and definitional clarity that materials AI currently lacks.

Forgetting in AI: Current Usage

Within the artificial intelligence literature applied to materials science, the term "forgetting" appears in several recognizably distinct yet overlapping senses [1, 2]. Historically, the dominant usage has treated forgetting as a catastrophic failure mode [1]. Even after techniques such as elastic weight consolidation emerged to mitigate interference in deep networks, the underlying diagnosis remained unchanged: forgetting is understood as an unintended degradation of prior knowledge [2]. This framing translates directly into materials-specific applications. Catastrophic forgetting has been documented, for instance, in graph neural networks trained incrementally on crystal property prediction tasks [5], where the introduction of new compositional data leads to abrupt loss of accuracy on previously learned oxide systems. A similar pattern of degradation arises in continual learning scenarios for materials property prediction under distribution shift [4].

A related but less explicitly theorized usage treats forgetting instead as a side effect of resource-management decisions. Replay buffers are routinely truncated to fit within memory budgets, and the eviction of older examples is described simply as "forgetting" without further scrutiny of its epistemic status [19, 24-28]. In the materials domain, where training data often combine high-fidelity density-functional-theory calculations with lower-fidelity experimental measurements, pruning decisions are typically justified on computational rather than scientific grounds—that is, based on memory limits rather than the relative value of retained versus discarded points [23]. Experience replay techniques in continual learning likewise accept controlled forgetting as a practical necessity [3, 28].

Beyond these two framings lies a third, even more implicit usage. Here, forgetting arises from model capacity constraints or regularization schemes that compress representations [17, 18]. Finite hidden-layer widths, weight pruning, and information-bottleneck regularization all induce selective information loss. In solid-state materials science, where machine-learning potentials must generalize across vast compositional spaces, such compression is often welcomed as a form of regularization that improves generalization [8, 26]. Yet the literature rarely labels the resulting information loss explicitly as “forgetting,” thereby obscuring the fact that the model is actively discarding aspects of the training distribution [24].

What emerges from this survey is a literature in which three conceptually distinct meanings—catastrophic failure, resource-driven side effect, and capacity-induced compression—coexist without clear demarcation [29]. Works operating primarily within the first framing include those on synaptic intelligence and lifelong learning reviews [8, 24]. Materials-oriented studies illustrate how the same term migrates across failure-mode and engineering contexts [9, 10, 23]. Replay buffer limits further blur the boundary, functioning simultaneously as a mitigation strategy and as a source of forgetting [19]. The net result is a conceptual ambiguity that impedes precise communication: the same word is asked to carry three fundamentally different meanings without explicit disambiguation [4, 5].

The Problem with Current Usage

The coexistence of these three usages generates several interlocking problems that extend beyond mere terminological inconvenience [1, 2]. One immediate consequence is the collapse of a normative distinction: forgetting is treated uniformly as a defect, even in situations where it might be scientifically advantageous [14]. Consider, for example, older experimental data that reflect superseded measurement artifacts—retaining such points indefinitely can pollute downstream predictions more severely than discarding them [23]. Yet the dominant catastrophic-forgetting framing provides no vocabulary for distinguishing beneficial from harmful loss [5].

A related difficulty emerges from unexamined side-effect forgetting, such as arbitrary replay buffer eviction. This practice introduces hidden biases whose epistemic consequences remain invisible to the practitioner [19]. A materials discovery campaign may silently lose access to rare high-pressure phases simply because those examples arrived early in the data stream, with no deliberate evaluation of their continued relevance [9]. What makes this particularly insidious is that the loss goes undiagnosed; the model ceases to generalize to those conditions without any alert to the user.

Beyond these conceptual and practical concerns, the literature lacks a shared language for discussing trade-offs. Generalization–stability analyses, bias–variance decompositions, and data-deletion studies each touch on aspects of information loss, yet they remain siloed from one another [17, 18]. Without a unifying construct, designers cannot compare the epistemic cost of forgetting one class of examples against another [15, 16]. This fragmentation has a further consequence: the absence of precise terminology actively hinders cumulative progress. Each new continual-learning paper for materials must reinvent its own ad-hoc heuristics rather than building on a common framework [4, 29].

Treating all forgetting as failure, therefore limits the field’s ability to engineer memory policies that are responsive to the unique characteristics of materials data—heterogeneous provenance, evolving experimental techniques, and the presence of both universal physical principles and context-specific outliers [7, 10]. What the situation demands, then, is not a refinement of existing taxonomies but a more precise vocabulary capable of capturing when forgetting serves scientific goals and when it undermines them.

Proposed Definition: Algorithmic Forgetting

Algorithmic forgetting is more productively understood not as an incidental byproduct of learning dynamics but as a deliberate intervention into the representational commitments of a machine-learning system. Within this framing, forgetting denotes the targeted removal or suppression of specific learned associations or training instances, enacted through explicit design choices and operating irreversibly within the model’s current state. The emphasis here is not merely on the absence of information, but on the structured reconfiguration of what the model is permitted to retain and act upon. Such an interpretation shifts attention away from passive degradation toward intentional epistemic control, where the boundaries of knowledge are actively drawn rather than passively eroded.

This reframing implies a set of conditions that distinguish algorithmic forgetting from more diffuse forms of information loss. Central among these is intentionality, insofar as the system—or more precisely, its designer or governing policy—must specify which elements of the training signal are to be removed and on what grounds [14].

This introduces a normative layer into model design, where decisions about retention and deletion are guided by explicit criteria rather than emerging implicitly from optimization processes. At the same time, the intervention must operate with specificity, targeting identifiable associations or examples rather than inducing a global reduction in representational capacity [8]. What is at stake is not compression in the aggregate, but the selective reorganization of knowledge. Equally important is the epistemic consequence of such interventions: the system's behavior must change in a predictable and controllable manner when confronted with queries that depend on the removed information [16]. Under these conditions, forgetting becomes observable not merely as an internal modification, but as a transformation in the model's inferential profile.

The analytical clarity of this definition becomes more apparent when contrasted with adjacent phenomena that might superficially appear similar. When a model fails to learn from rare or underrepresented examples, no act of forgetting has taken place; the absence of knowledge reflects a failure of acquisition rather than its removal. Similarly, uniform regularization mechanisms such as weight decay operate by compressing representations across the parameter space, thereby redistributing informational capacity without selectively erasing particular content [18]. In contrast, scenarios in which data are explicitly removed according to a defined policy—such as the exclusion of pre-2020 synthesis routes known to embed systematic measurement bias—exemplify the defining features of algorithmic forgetting, as the intervention is both targeted and justified within a broader epistemic rationale [19]. The distinction that emerges is therefore not reducible to statistical patterns of loss or retention, but hinges on whether the model's internal structure has been intentionally reconfigured in relation to specific knowledge claims [3].

Viewed through this lens, forgetting becomes a controllable dimension of system design rather than a pathological outcome to be avoided. This shift opens a conceptual space in which memory is treated as an actively governed resource, capable of being aligned with the evolving characteristics of material data streams. In practice, such an approach enables the construction of models that can adapt not only by acquiring new information, but by selectively relinquishing outdated, biased, or contextually irrelevant knowledge. Under these conditions, forgetting is no longer synonymous with failure; it becomes an instrument through which the epistemic integrity of materials AI systems can be continuously recalibrated [4, 23].

Distinctions from Nearby Terms

The conceptual boundaries of algorithmic forgetting become clearest when contrasted, term by term, with five neighboring constructs that also involve information loss [1, 2]. The comparison can be structured along four axes: intent (deliberate versus emergent), mechanism (targeted removal versus diffuse degradation), epistemic consequence (predictable change in future inferences versus uncontrolled degradation), and reversibility (permanent within model state versus recoverable through retraining).

Table 1 establishes the conceptual boundary conditions of algorithmic forgetting by systematically distinguishing it from adjacent mechanisms of information loss along dimensions of intent, mechanism, epistemic consequence, reversibility, and operational locus.

Table 1. Conceptual boundary matrix: algorithmic forgetting vs. adjacent mechanisms of information loss

Construct	Intent	Mechanism	E con
Algorithmic forgetting	Deliberate	Targeted removal of specific associations or examples	Pr co c ir b
Catastrophic forgetting	Unintended	Weight overwriting during new learning	Un deg prior
Data deletion/unlearning	Externally mandated (privacy, compliance)	Removal from the dataset or retraining pipeline	Glo di
Capacity saturation	Emergent constraint	Diffuse compression	Sta

https://doi.org/10.51847/w540537414

		due to finite model capacity	repre
Regularization-induced compression	Generalization-oriented	Uniform parameter shrinkage or pruning	S repre noi infor
Replay buffer eviction (unstructured)	Resource-driven	Arbitrary removal due to memory limits	Hic un infor

Catastrophic forgetting, as originally characterized [1] and later analyzed in materials contexts [4, 5], is entirely emergent and unintended; its mechanism is overwriting of consolidated weights, its epistemic consequence is uncontrolled loss of prior accuracy, and it is reversible only through explicit mitigation strategies such as those proposed [2]. Algorithmic forgetting, by contrast, is deliberate and targeted.

Data deletion, examined in studies of machine unlearning and approximate removal [14, 15], operates at the dataset level before or during training and is usually motivated by privacy or compliance rather than by ongoing model performance. Its mechanism is removal from the training corpus, its epistemic effect is a global shift in the empirical risk minimizer, and it is irreversible only in the sense that the data are gone; the model can be retrained from scratch. Algorithmic forgetting, however, occurs inside an already-trained or incrementally learning model and can be far more granular [16].

Privacy preservation via machine unlearning [16] shares the goal of removal but is triggered by external legal or ethical requirements rather than by scientific utility. Its mechanism must satisfy certified unlearning guarantees, whereas algorithmic forgetting in materials AI is driven by epistemic criteria such as data quality or distribution shift [23].

Capacity saturation, implicit in finite-model analyses and discussed in generalization–stability trade-offs [17, 18], induces diffuse information loss through architectural limits. The mechanism is global compression rather than selective removal, the epistemic consequence is statistical rather than targeted, and the process is not under direct designer control once the architecture is fixed.

Finally, regularization-induced compression, as seen in synaptic intelligence approaches [8] or progress-and-compress frameworks [26], deliberately reduces representational redundancy but does so uniformly rather than example-specifically. Its intent is improved generalization, not the surgical excision of particular associations [24].

Taken together, these distinctions reveal algorithmic forgetting as a distinct design primitive: intentional, granular, epistemically motivated, and situated within the model's ongoing life cycle [29]. By separating it from these neighbors, the field gains the conceptual precision required to treat memory management as an engineering discipline rather than an accidental byproduct of learning.

Forgetting as Design Choice

The shift advanced here reconceptualizes algorithmic forgetting as an intentional design variable rather than an incidental consequence of learning dynamics, thereby altering the questions that can be meaningfully posed within materials AI [2, 8]. Once forgetting is situated within the framework established by definition 1, the problem space expands beyond mitigation toward deliberation: the practitioner is no longer confined to preventing loss but is compelled to consider the conditions under which selective removal of knowledge enhances model performance and epistemic integrity. This perspective is not imposed externally but emerges from the intrinsic characteristics of materials data, which are heterogeneous in origin, temporally unstable, and subject to practical constraints that preclude indefinite retention.

Within this context, the capacity to respond to distributional change becomes a central motivation for engineered forgetting. Materials datasets evolve as experimental techniques mature, synthesis pathways diversify, and application targets shift, leading to a continual reconfiguration of the statistical landscape on which models operate [4, 23]. Under such conditions, the uncritical preservation of historical data risks entrenching correlations that have lost their empirical validity, thereby impairing predictive performance at the frontier of discovery [10]. A related implication is that relevance cannot be maintained solely through accumulation; it requires the capacity to relinquish outdated associations in a controlled manner. Algorithmic forgetting provides precisely this mechanism, enabling models to disengage from superseded patterns without relying on the disruptive effects of catastrophic interference [5]. In practice, this capacity aligns with observations from continual-learning systems in materials property prediction, where the pace of experimental innovation often exceeds the cadence of model retraining, rendering selective data removal a pragmatic requirement rather than an optional refinement [4, 23].

This emphasis on adaptation is inseparable from the material constraints under which computational systems operate. Memory capacity, training time, and inference efficiency impose hard limits on the volume of data that can be retained and processed, even within large-scale materials informatics infrastructures [19, 28]. As a result, mechanisms such as replay buffers and incremental learning pipelines must continuously negotiate which data to preserve and which to discard [3]. When these decisions are governed by explicit criteria rather than arbitrary overflow, the resulting process aligns with the definition of algorithmic forgetting and acquires a distinctly epistemic character [19]. The significance of this shift lies in its reorientation of constraint: what begins as a hardware-imposed limitation is transformed into a design opportunity, where the selective suppression of information is guided by its contribution to scientific inference. By privileging regions of composition–structure–property space that carry higher informational value, the model avoids dispersing its representational capacity across diminishing returns, thereby sustaining analytical focus within an expanding search domain [11].

Beyond considerations of adaptation and efficiency, the deliberate management of memory also intersects with emerging concerns around data governance in collaborative materials research. Datasets increasingly incorporate elements that are proprietary, sensitive, or ethically constrained, including unpublished synthesis protocols and biologically relevant materials interfaces. The continued retention of such information within trained models raises questions that extend beyond technical performance to encompass legal and ethical accountability [14, 16]. Techniques developed under the rubric of machine unlearning demonstrate that targeted removal of learned information can be achieved with formal guarantees [15, 16], and this capability acquires new significance when situated within a broader framework of scientific utility. In this setting, algorithmic forgetting functions not merely as a compliance mechanism but as a means of reconciling the openness of materials informatics with the legitimate need for controlled data stewardship [6].

Taken together, these considerations reposition forgetting as an affirmative design lever within materials AI, one that operates across multiple dimensions of system behavior. The relevant parameters are no longer implicit but subject to deliberate specification: the scope of removal, the temporal conditions under which it is triggered, and the technical pathways through which it is enacted [8, 19, 26]. This shift also introduces a deeper transformation in how memory itself is conceptualized, moving from passive accumulation toward active curation. Existing techniques—ranging from replay buffers to gradient episodic memory and related experience-replay strategies—already provide the operational substrate for such interventions [3, 19, 28]. Yet, their role has often been framed in terms of mitigation rather than intentional design. Reinterpreted through the lens developed here, these methods become instruments through which forgetting is not merely managed but strategically deployed. Under these conditions, algorithmic forgetting constitutes a new degree of freedom in the construction of adaptive, durable, and epistemically robust materials discovery systems, enabling models to evolve in tandem with the scientific environments they are intended to navigate [24, 29].

When to Forget, When to Retain

A materials-specific conceptual framework for deciding when to invoke algorithmic forgetting rests on two complementary sets of criteria—retention criteria and forgetting criteria—each grounded in the epistemic and practical realities of materials science. These criteria do not dictate automatic decisions but supply a shared vocabulary for justifying memory policies in research documentation and peer review [7, 23].

Table 2 formalizes the proposed decision lattice by integrating retention and forgetting criteria into a structured, operational framework for memory policy design in materials AI systems.

Table 2. Decision lattice for algorithmic forgetting in materials AI: integrating retention and forgetting criteria

Decision dimension	Criterion	Epistemic rationale	Risk if ignored
Retention	Rarity of example	Preserves coverage of	Loss of physically

		sparsely sampled phase space	critical edge cases
Retention	Recency of validation	Aligns model with current experimental consensus	Drift toward outdated scientific assumptions
Retention	Relevance to search space	Supports active discovery objectives	Reduced predictive utility for ongoing campaigns
Forgetting	Measurement artifacts	Removes systematically biased observations	Persistent model bias and error propagation
Forgetting	Superseded synthesis routes	Eliminates obsolete process–property mappings	Misleading generalization to irrelevant regimes
Forgetting	Distributional outliers (artifact-driven)	Reduces overfitting to non-generalizable anomalies	Distorted model boundaries
Meta-decision layer	Policy review cycle	Enables adaptive memory governance	Static policies in dynamic environments
Meta-decision layer	Balance threshold (retain vs forget)	Ensures justified trade-offs	Arbitrary or biased pruning

The distinction between retention and forgetting in materials AI acquires practical significance only when it is anchored in criteria that reflect scientific utility rather than abstract data-management heuristics. Retention, in this sense, is not synonymous with accumulation but with the preservation of information that continues to expand the model's epistemic reach. Particularly consequential are examples that occupy sparsely sampled regions of phase space, such as high-pressure phases or metastable defect configurations whose empirical occurrence is intrinsically rare [9]. The disappearance of such data would not merely reduce dataset size but would contract the model's capacity to reason about extreme or boundary conditions. A related consideration concerns the temporal anchoring of knowledge. Data supported by recent experimental validation or high-fidelity computational benchmarks carry disproportionate weight because they encode the current state of disciplinary consensus, which itself evolves alongside instrumentation and methodological refinement [10, 23]. Prioritizing such examples stabilizes the model against drift toward outdated correlations. This orientation toward present relevance extends further to the spatial structure of discovery efforts, where examples situated within the compositional or structural neighborhoods under active investigation exert a direct influence on ongoing experimental decisions. Preserving these locally informative instances ensures that the model remains aligned with the immediate trajectory of materials exploration rather than dispersing its attention across less consequential regions of the search space [11].

This emphasis on retention necessarily invites its counterpart: the identification of conditions under which continued storage becomes counterproductive. Information does not remain neutral over time; under certain circumstances, it actively distorts inference. Data derived from measurement regimes later shown to introduce systematic bias exemplify this dynamic, as their persistence can propagate artifacts that are no longer scientifically defensible [23]. Similarly, examples generated through synthesis pathways that have fallen out of relevance—whether due to technological obsolescence or shifts in research priorities—contribute little to the representation of the contemporary materials landscape and may instead anchor the model to trajectories no longer pursued [7, 10]. Beyond these cases, a subtler category emerges in the form of apparent outliers that were initially interpreted as meaningful deviations but are subsequently recognized as artifacts of limited sampling or experimental noise. Retaining such points risks overfitting the model to idiosyncrasies that lack generalizable significance, thereby undermining predictive robustness in broader contexts [4, 9].

The practical implementation of these criteria resists static prescription and instead unfolds as an ongoing evaluative process. Memory management becomes an iterative exercise in aligning the model's internal representation with a shifting epistemic environment, where decisions about retention and removal are revisited as new data, methods, and interpretations emerge. In this setting, it is entirely plausible for a system to preserve rare configurations—

such as high-entropy alloy states that remain underexplored—while simultaneously discarding earlier experimental entries compromised by calibration drift or methodological limitations. The governing logic resembles a structured decision surface rather than a fixed rule set, where each datum is assessed in relation to multiple, sometimes competing dimensions of value. Forgetting is triggered not by arbitrary thresholds but by a cumulative assessment in which the informational cost of retention outweighs its contribution to inference. Such an approach enables a balance between indiscriminate pruning and unbounded accumulation, ensuring that the model's memory evolves in step with the scientific landscape it is designed to interrogate [24, 29].

Objections and Replies

Positioning algorithmic forgetting as an intentional design choice rather than an error condition inevitably invites resistance, particularly in a field where predictive reliability is closely tied to the accumulation of learned associations. A common concern is that any form of forgetting compromises the stability of structure–property mappings, thereby eroding trust in model outputs [1, 2]. This concern, however, rests on an implicit equivalence between all forms of information loss. Once a distinction is drawn between uncontrolled degradation and deliberate, policy-driven removal, the conceptual terrain shifts. Catastrophic forgetting reflects an unregulated collapse of prior knowledge, whereas algorithmic forgetting operates through targeted intervention guided by epistemic criteria. Under this interpretation, the removal of outdated or misleading associations does not diminish knowledge but reorganizes it, allowing models to remain aligned with evolving data distributions and experimental realities [4, 5]. What appears as loss from one perspective becomes, under closer scrutiny, a mechanism of continual recalibration.

A related hesitation emerges from the unpredictability of future relevance. Materials discovery often proceeds through unexpected pathways, and examples that appear marginal at one moment may later acquire central importance [9, 11]. From this vantage point, any policy of forgetting risks foreclosing avenues of inquiry that cannot yet be anticipated. This uncertainty is real, yet it does not preclude structured intervention. The criteria governing retention and removal—particularly those grounded in rarity, recency, and contextual relevance—function precisely to guard against the indiscriminate elimination of potentially valuable information. Moreover, the temporal dimension of policy design introduces an additional layer of flexibility. Although forgetting is irreversible within a given model state, it need not be final at the level of system governance. Periodic reassessment of previously discarded data against new validation benchmarks enables selective reintegration when warranted, thereby transforming forgetting into a revisable commitment rather than a definitive erasure [3, 28]. In practice, this approach acknowledges uncertainty while embedding mechanisms for its correction [16].

Another line of critique interprets algorithmic forgetting as conceptually redundant, suggesting that its effects are already subsumed under established techniques such as regularization or capacity control [17, 18]. This interpretation overlooks a critical distinction in both intent and granularity. Regularization operates by imposing global constraints on model complexity, promoting generalization through diffuse smoothing across the parameter space. Algorithmic forgetting, by contrast, intervenes at the level of specific examples or associations, guided by considerations that extend beyond statistical efficiency to include scientific validity and contextual relevance [8, 26]. The two mechanisms are therefore neither interchangeable nor mutually exclusive; they address different dimensions of model behavior and can, in fact, be deployed in tandem. Recognizing this distinction clarifies that forgetting is not a disguised form of compression but a qualitatively different operation with its own conceptual and practical implications [24].

Engaging with these concerns does more than defend a particular methodological proposal; it reveals a broader shift in how memory is understood within materials AI. Rather than treating retention as an unquestioned good and loss as an inherent defect, the field is beginning to confront the necessity of selective curation in environments characterized by continual change. Within this emerging perspective, algorithmic forgetting is neither reckless nor redundant. It represents an extension of the discipline's evolving understanding of how knowledge should be managed, updated, and, when necessary, relinquished to sustain reliable and adaptive materials discovery systems [29].

Implications for Materials AI Practice

Adopting algorithmic forgetting as a design choice requires concrete changes in how materials AI research is conducted and reported. First, every published model description should include an explicit forgetting policy section that states which data were deliberately removed, according to which criteria, and at what points in the training lifecycle [14, 19]. Second, researchers must distinguish catastrophic forgetting from algorithmic forgetting in method sections and ablation studies, treating the latter as a tunable hyperparameter rather than a failure to be minimized [2, 5]. Third, system architectures should incorporate configurable retention parameters—such as replay-buffer eviction policies or synaptic-intelligence decay schedules—so that forgetting can be adjusted without retraining from scratch [8, 26]. Fourth, documentation must record not only retained data but also the rationale for forgotten content, creating an auditable memory ledger that supports reproducibility and peer scrutiny [16, 23].

These practices elevate memory management from an implicit afterthought to a core engineering discipline. Journals and conferences in materials informatics can reinforce the shift by requiring forgetting-policy statements in supplementary materials, much as they already require dataset and code availability. Over time, such norms will produce a cumulative literature in which forgetting is no longer hidden but openly engineered for scientific benefit [7, 11]. The result is more transparent, adaptable, and trustworthy materials discovery pipelines that treat memory as a resource to be curated rather than an infinite archive to be endured.

Conclusion

This boundary/definitional paper has identified the ambiguous usage of “forgetting” across the materials AI literature, surveyed its three dominant forms, diagnosed the resulting epistemic limitations, and proposed a precise definition of algorithmic forgetting as a deliberate design choice (definition 1). By distinguishing it from catastrophic forgetting, data deletion, privacy preservation, capacity saturation, and regularization-induced compression, the analysis supplies the conceptual boundaries the field has lacked. The framework for deciding when to forget versus when to retain—anchored in rarity, recency, and relevance—offers practitioners a practical decision lattice tailored to the evolving nature of materials data.

Algorithmic forgetting, once recognized as a design lever rather than a defect, opens new avenues for adaptive, resource-aware, and epistemically responsible systems. The call is therefore straightforward: future materials AI research must treat forgetting policies as a standard, documented component of system design, reported with the same rigor as architecture or loss functions. Only then can the field move beyond fearing memory loss toward engineering memory with intention and precision.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 29 Jul 2021 Revised: 14 Oct 2021 Accepted: 15 Nov 2021
Published online: 18 January 2022

References

Shurkova EY. Symbolic-connectionist model of relation learning and visual reasoning. Available from: https://era.ed.ac.uk/items/f094402a-bdb3-4cf4-b673-3ae23dba0657/full

Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, et al. Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci U S A. 2017;114(13):3521-6.

Lopez-Paz D, Ranzato MA. Gradient episodic memory for continual learning. Adv Neural Inf Process Syst. 2017;30:1-10.

Stern M, Pinson MB, Murugan A. Continual learning of multiple memories in mechanical networks. Phys Rev X. 2020;10(3):031044.

Liu H, Yang Y, Wang X. Overcoming catastrophic forgetting in graph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence; 2021. 35(10):8653-61.

Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018;559(7715):547-55.

Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of machine learning in solid-state materials science. npj Comput Mater. 2019;5(1):83.

Zenke F, Poole B, Ganguli S. Continual learning through synaptic intelligence. In: Proceedings of the 34th International Conference on Machine Learning (ICML). PMLR; 2017. p. 3987-95.

Cai J, Xu M, Li W, Xiong Y, Xia W, Tu Z, et al. MeMOT: Multi-object tracking with memory. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022. p. 8090-100.

Go SX, Wang Q, Wang B, Jiang Y, Bajalovic N, Loke DK. Continual learning electrical conduction in resistive-switching-memory materials. Adv Theory Simul. 2022;5(8):2200226.

Chang R, Wang YX, Ertekin E. Towards overcoming data scarcity in materials science: Unifying models and datasets with a mixture of experts framework. npj Comput Mater. 2022;8(1):242.

Zhang Z, Mondal S, Mandal S, Allred JM, Aghamiri NA, Fali A, et al. Neuromorphic learning with Mott insulator NiO. Proc Natl Acad Sci U S A. 2021;118(39):e2017239118.

Liu N, Ihalage A, Zhang H, Giddens H, Yan H, Hao Y. Interactive human-machine learning framework for modelling of ferroelectric-dielectric composites. J Mater Chem C. 2020;8(30):10352-61.

Ginart A, Guan M, Valiant G, Zou JY. Making AI forget you: Data deletion in machine learning. Adv Neural Inf Process Syst. 2019;32:1-10.

Izzo Z, Smart MA, Chaudhuri K, Zou J. Approximate data deletion from machine learning models. In: Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS). PMLR; 2021. p. 2008-16.

Sekhari A, Acharya J, Kamath G, Suresh AT. Remember what you want to forget: Algorithms for machine unlearning. NeurIPS. 2021;34:18075-86.

Bartoldson B, Morcos A, Barbu A, Erlebacher G. The generalization-stability tradeoff in neural network pruning. NeurIPS. 2020;33:20852-64.

Yang Z, Yu Y, You C, Steinhardt J, Ma Y. Rethinking bias-variance trade-off for generalization of neural networks. In: International Conference on Machine Learning; 2020. p. 10767-77.

Kong K, Li G, Ding M, Wu Z, Zhu C, Ghanem B, et al. Robust optimization as data augmentation for large-scale graphs. CVPR. 2022. p. 60-9.

Diethe T, Borchert T, Thereska E, Balle B, Lawrence N. Continual learning in practice. arXiv preprint arXiv:1903.05202. 2019.

Wang B, Ouyang Y, Li Y, Pan M, Cui H, Wang Y, et al. MoMa: A simple modular learning framework for material property prediction. In: Fourteenth International Conference on Learning Representations.

Bang J, Koh H, Park S, Song H, Ha JW, Choi J. Online continual learning on a contaminated data stream with blurry task boundaries. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 9275-84.

Wu YC, Feng JW. Development and application of artificial neural network. Wirel Pers Commun. 2018;102(2):1645-56.

Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: A review. Neural Netw. 2019;113:54-71.

Aljundi R, Babiloni F, Elhoseiny M, Rohrbach M, Tuytelaars T. Memory aware synapses: Learning what (not) to forget. In: Proceedings of the European Conference on Computer Vision (ECCV); 2018. p. 139-54.

https://doi.org/10.51847/w540537414

Schwarz J, Czarnecki W, Luketina J, Grabska-Barwinska A, Teh YW, Pascanu R, et al. Progress & compress: A scalable framework for continual learning. In: International Conference on Machine Learning; 2018. p. 4528-37.

Chaudhry A, Ranzato MA, Rohrbach M, Elhoseiny M. Efficient lifelong learning with A-GEM. arXiv preprint arXiv:1812.00420. 2018.

Rolnick D, Ahuja A, Schwarz J, Lillicrap T, Wayne G. Experience replay for continual learning. NeurIPS. 2019;32.

De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, et al. A continual learning survey: Defying forgetting in classification tasks. IEEE Trans Pattern Anal Mach Intell. 2021;44(7):3366-85.