

An AI-Powered Personal Data Expiry Management System: Automated Classification, Retention Scoring, and Privacy-Preserving Deletion

¹Sanika Patil, ²Suhani Saste, ³Mihir Patil, ⁴Shubham Patil

¹²³⁴Department of Information Technology, Pillai College of Arts, Commerce and Science, Navi Mumbai, India

Abstract - Organizations now retain personal data across databases, file shares, software-as-a-service applications, backups, and machine learning training sets, often well past the point at which the data still serves a documented purpose. Manual data retention programs struggle to keep pace with this sprawl, which raises regulatory exposure under storage-limitation principles such as those in the General Data Protection Regulation, increases the cost and attack surface of unnecessary storage, and leaves organizations unable to answer a simple question: which personal data can safely be deleted today. Artificial intelligence offers a route to closing this gap by continuously discovering personal data, estimating how much retention value it still holds, and triggering compliant deletion or anonymization automatically. This paper proposes an AI-Powered Personal Data Expiry Management System (PDEMS), a framework in which personal data is automatically classified for sensitivity, scored for an expiry likelihood that combines legal retention requirements, consent status, purpose fulfillment, access recency, and data value decay, and then routed to retention, archival, anonymization, or deletion, with deletion extended to any downstream machine learning models through machine unlearning. We review the literature on data lifecycle security and privacy, automated personally identifiable information detection, data value decay, and machine unlearning; propose a six-layer PDEMS architecture; formalize an expiry scoring function and a constrained decision rule that legal holds and active consent can override; and present an operating algorithm for the full discovery-to-deletion cycle. Because no experiments were conducted for this paper, literature-reported results are clearly distinguished from the proposed methodology throughout, and the paper closes with the evaluation protocol, privacy safeguards, and limitations that further empirical work would need to address.

Keywords - data retention; data lifecycle management; personal data expiry; automated PII detection; machine unlearning; privacy-preserving computing; data minimization.

I. INTRODUCTION

Modern organizations collect personal data far faster than they retire it. Customer records, employee files, support tickets, application logs, and the training sets behind machine learning models all accumulate across databases, file shares, cloud storage, and third-party software, and much of this data continues to be stored long after the purpose for which it was collected has been fulfilled.

Industry data governance practice now routinely refers to this accumulated material as redundant, obsolete, or trivial data, and treats its identification and removal as a distinct governance problem from the initial task of securing data while it is in active use.

Over-retention is not a passive risk. Data protection regulations such as the European Union's General Data Protection Regulation embed a storage limitation principle directly into law, requiring that personal data be kept in identifiable form no longer than necessary for the purposes for which it was collected [1], [2]. Every record kept beyond that point is also a record that can be breached, and it is a record that, if used to train or fine-tune a machine learning model, may need to be actively unlearned rather than simply deleted from a database, since trained models can retain the statistical influence of individual records even after the original rows are removed [6]–[8]. At the same time, security and privacy surveys of the data lifecycle consistently find that collection and storage stages receive far more research and tooling attention than the later retention and destruction stages, leaving automated, defensible deletion comparatively underdeveloped relative to automated collection and classification [3]–[5].

Existing data lifecycle management tools address part of this problem. Commercial platforms can discover personal data across structured and unstructured stores, classify it by sensitivity, and apply administrator-defined retention schedules with scheduled deletion jobs. These systems are valuable, but the retention schedule itself is typically a static, manually authored rule rather than a continuously updated estimate of whether a specific record still has legal, operational, or business value.

They also generally stop at the data-store boundary: a scheduled deletion job can remove a row from a database, but it has no mechanism for propagating that deletion into any machine learning model that was trained on the row, which is precisely the gap that machine unlearning research has been developed to close [6]–[8]. This paper is concerned with the space between these two bodies of work: automated, AI-driven estimation of when personal data has expired, and the mechanisms needed to enforce that expiry consistently across both conventional data stores and the models trained on them.

We refer to the proposed system as an AI-Powered Personal Data Expiry Management System, or PDEMS. The central research problem is how an organization can continuously and automatically determine which personal data has passed the point of legitimate retention, and act on that determination in a way that respects legal holds, active consent, and business necessity, while also reaching into any derived machine learning artifacts. This raises several sub-questions that structure the remainder of the paper: how personal data can be discovered and classified for sensitivity at scale [9]–[11]; how an expiry or retention-value score can be constructed from legal, consent, usage, and data-value-decay signals [12], [13]; how such a score should be combined with hard legal and consent constraints rather than overriding them; how deletion decisions can be executed consistently across heterogeneous data stores without centralizing raw personal data during scoring [14]–[16]; and how deletion can be extended to machine learning models through unlearning rather than treated as a database-only operation [6]–[8].

The contribution of this paper is threefold. First, we bring together literature on data lifecycle security and privacy, automated PII detection, data value decay, and machine unlearning, which are rarely discussed together, to motivate and scope the PDEMS problem and identify a concrete gap between static retention scheduling and continuous, AI-driven expiry estimation. Second, we propose a six-layer PDEMS architecture, a mathematical formulation of an expiry score with a hard-constraint decision rule, and an operating algorithm that carries a deletion decision through to model unlearning and audit. Third, we distinguish literature-reported findings from the proposed methodology throughout, and set out the evaluation protocol, privacy safeguards, and limitations that empirical validation of such a system would need to address.

The remainder of this paper proceeds as follows. Section II reviews the literature on data lifecycle security and privacy, regulatory retention principles, automated personal data discovery and classification, data value decay, and machine unlearning, and synthesizes this literature into a research gap. Section III presents the proposed PDEMS methodology, covering its architecture, data considerations, feature engineering, expiry scoring formulation, decision and optimization logic, and the execution and feedback mechanism that closes the loop. Section IV discusses the framework conceptually, since no experiments were conducted, and presents evaluation metrics and a comparative analysis against existing data lifecycle management approaches. Section V addresses privacy, security, and ethical considerations; Section VI states limitations; Section VII outlines future work; and Section VIII concludes.

II. LITERATURE REVIEW

A. Regulatory Foundations of Data Retention

Data retention obligations in most jurisdictions are shaped by two competing pressures:

sector-specific rules that mandate keeping certain records for a minimum period, and general privacy law that limits how long personal data may be kept in identifiable form once its purpose has been served. The General Data Protection Regulation is the most widely referenced instance of the latter. It defines personal data broadly and requires, under its storage limitation principle, that personal data be kept in a form that permits identification of the data subject for no longer than is necessary for the purposes for which it is processed [1]. Legal commentary on the regulation notes that this principle does not specify a fixed retention period; instead, it obliges the data controller to justify, for each category of personal data, how long retention remains necessary, which in practice means the retention decision has to be made per data category and revisited over time rather than fixed once at collection [2]. This distinction between a scientifically or legally justified retention period and an operationally convenient one is central to the PDEMS design in Section III, where legal retention requirements are treated as a hard lower bound on the expiry decision rather than one input to be traded off against others.

B. Security and Privacy Across the Data Lifecycle

A separate line of survey research has examined the data lifecycle as a sequence of distinct security and privacy stages. One widely cited taxonomy divides the big-data lifecycle into collection, storage, analytics, utilization, and destruction phases, maps documented threats and standardization efforts onto each phase, and finds that established international standards address collection and storage far more thoroughly than the later utilization and destruction phases [3]. A systematic review of privacy-preserving techniques across the data lifecycle similarly organizes existing encryption, anonymization, and access-control techniques by lifecycle stage and highlights retention and deletion as a stage with comparatively fewer mature, standardized techniques relative to collection and storage [4]. A domain-specific survey of the health data lifecycle reaches a related conclusion:

privacy-preserving approaches in that domain are dominated by cryptographic and access-control techniques applied while data is active, with far less attention paid to the disposal stage once data is no longer clinically needed [5]. Taken together, this literature indicates that the destruction or expiry stage of the data lifecycle is a recognized, if under-served, research area, which supports treating expiry management as a distinct system to be designed rather than an afterthought bolted onto existing storage and access controls.

C. Automated Discovery and Classification of Personal Data

Any expiry system must first be able to find personal data reliably across unstructured and semi-structured sources, since manual tagging cannot scale to enterprise data volumes. Early automated approaches to this problem combined rule-based pattern matching, such as regular expressions for identifiers with a checkable structure, with dictionary lookups for names and locations, but these approaches degrade for uncommon, misspelled, or context-dependent entities [9].

More recent work has moved toward machine learning and deep learning classifiers. Research applying natural language processing together with unsupervised learning to personally identifiable information detection in large unstructured text corpora demonstrated that entity recognition techniques can locate personal data in free text without relying on a fixed dictionary of known identifiers [10]. Building on this direction, ensemble approaches that combine semantic features with multiple classifiers have been proposed for cross-domain personally identifiable information detection,

de-identification, and re-identification risk assessment, addressing the practical problem that a model trained on one document type or domain often performs poorly on another [11]. Most directly relevant to the discovery layer proposed in Section III, a recent ensemble system that combines rule-based patterns, dictionary matching, conditional random fields, a bidirectional long short-term memory sequence model, and a transformer-based token classifier under a class-specific weighted voting scheme reported a weighted F1 score of 0.9749 on a general-purpose benchmark, narrowly exceeding its strongest individual component, while also showing reduced accuracy on an out-of-distribution financial-document test set, which the authors attribute to calibration sensitivity for entity classes absent from the ensemble's weighting corpus [9]. This result, drawn directly from the cited study rather than from any experiment conducted for this paper, illustrates both the practical feasibility of ensemble-based personal data discovery and the generalization limits that a production expiry system would need to manage.

D. Data Value Decay and the Economics of Retention

A second body of relevant work asks a different question: even where retention is legally permitted, does keeping older personal data continue to be useful. Economic analysis of the relationship between data age and model value has argued that when data loses relevance over time, accumulating an ever-larger stock of historical data can be actively counterproductive, since retraining on stale, less relevant records can measurably harm a model's accuracy relative to using a smaller, more recent sample, and the rate at which relevance decays differs substantially across data domains [12].

Complementary work in data management has formalized data decaying as the progressive, deliberate loss of resolution or detail in stored data as it ages, proposing that recent data be kept at full resolution while older data is compacted, aggregated, or discarded on a schedule tied to its diminishing operational value, and demonstrating this approach on large-scale telecommunication data [13]. These findings jointly support treating retention value as a quantity that changes continuously over time rather than a binary property fixed at the moment of collection, which motivates the explicit value-decay term

included in the expiry scoring function proposed in Section III-E.

D. Machine Unlearning as an Extension of Deletion

Deleting a record from a database does not remove its influence from a machine learning model that was trained on it, since model parameters can encode statistical traces of individual training examples; this observation is the basis of the field of machine unlearning, which several regulatory commentators connect directly to the right to erasure found in data protection law [6]–[8]. A comprehensive survey of the field distinguishes exact unlearning, which reconstructs a model as though the deleted data had never been included, typically through sharded or incremental retraining strategies, from approximate unlearning, which adjusts model parameters to reduce the influence of deleted data without full retraining, at the cost of weaker removal guarantees [6]. A second survey organizes the same space by the type of model and the verification mechanism used to check that unlearning occurred, and notes that few existing studies compare unlearning methods under realistic, non-uniform data distributions, which is precisely the setting a production expiry system would need to operate in in [7]. The foundational sharded, isolated, sliced, and aggregated retraining approach, commonly referred to as SISA, demonstrated that partitioning training data in advance can make exact unlearning computationally tractable by limiting retraining to only the shard containing the deleted record, rather than the full training set [8]. This body of work is the basis for treating model unlearning as a mandatory component of the execution layer in Section III-H, rather than an optional extension of database-level deletion.

E. Privacy-Preserving Computation for Cross-Store Classification

Because a PDEMS must classify and score personal data across many distributed stores without unnecessarily concentrating that data in one place, techniques originally developed for privacy-preserving machine learning are directly applicable to the scoring layer itself. Differential privacy provides a formal guarantee that the presence or absence of any single individual's data has a bounded effect on the output of a computation, which allows aggregate statistics to be published or shared without revealing individual records [14]. Federated learning allows a shared model to be trained across many decentralized data holders by exchanging model updates rather than raw data, which keeps personal data local to the store that holds it while still allowing a central model to benefit from patterns observed across stores [15]. Earlier anonymization work on k -anonymity established that a released dataset can be structured so that each record is indistinguishable from at least k minus one other records on any combination of quasi-identifying attributes, a technique still used as a lightweight privacy safeguard for classification metadata that must be aggregated centrally [16].

These three techniques are complementary rather than interchangeable, and Section III-F treats them as candidate building blocks for the privacy-preserving classification layer rather than as a single prescribed solution.

F. Synthesis and Research Gap

Table IV, presented after the methodology section for ease of comparison, situates the proposed PDEMS relative to the literature reviewed above. Regulatory and legal scholarship establishes that indefinite retention is not compliant and that retention justification must be revisited over time, but it does not specify a computational mechanism for doing so [1], [2]. Data lifecycle security surveys confirm that the destruction and disposal stage is comparatively under-addressed relative to collection and storage, without proposing an automated scoring mechanism to close that gap [3]–[5]. Automated PII detection research supplies a mature discovery layer but stops at classification rather than extending to a retention or expiry decision [9]–[11]. Data value decay research demonstrates that retention value changes continuously with time and can be estimated, but has not been integrated into a governance decision loop [12], [13]. Machine unlearning research supplies the mechanism needed to extend deletion into trained models, but is generally studied independently of the upstream question of which data should be deleted and when [6]–[8]. Limited research to date has combined continuous, AI-driven expiry scoring, a hard-constrained decision rule that respects legal holds and consent, privacy-preserving cross-store classification, and deletion that is propagated into downstream machine learning models, within a single closed-loop system. This is the gap the proposed PDEMS framework is intended to address.

III. PROPOSED METHODOLOGY

This section presents the proposed AI-Powered Personal Data Expiry Management System. Unless a specific literature citation is given, the architecture, scoring formulation, and algorithm described below constitute the proposed contribution of this paper rather than an established or previously validated system, and are presented as a conceptual framework requiring empirical validation, as discussed in Section IV.

A. Conceptual Data Lifecycle

Personal data in an organization typically passes through a small number of recognizable stages: it is created or collected, it is actively used to serve the purpose for which it was collected, its purpose is gradually fulfilled while its operational value declines, and at some point it crosses a threshold beyond which continued retention in identifiable form is no longer justified. Fig. 1 depicts this progression and the two conditions, an active legal hold or an unwithdrawn active consent basis, that can block the transition from the expiry threshold to an actual deletion or anonymization action. This conceptual picture, adapted from the general lifecycle staging used in prior data lifecycle security surveys [3], underlies the scoring and decision logic developed in the remainder of this section.

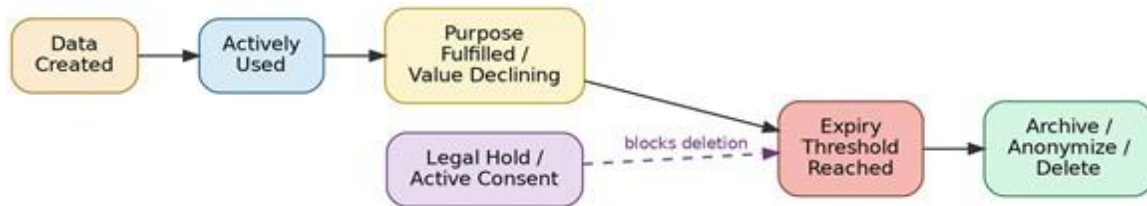


Fig. 1. Conceptual personal data lifecycle and the point at which an expiry decision is made.

B. System Architecture

The proposed PDEMS follows a six-layer architecture. The first layer is the personal data source layer, comprising the organization's databases, file shares, software-as-a-service applications, backup systems, application logs, and machine learning training sets. The second layer is a discovery and classification engine that scans these sources for personal data and assigns sensitivity labels, drawing on the automated PII detection techniques reviewed in Section II-C [9]–[11]. The third layer is the expiry scoring engine, which computes the continuously updated expiry score defined in Section III-E for each identified record or record group. The fourth layer is a policy and compliance engine that encodes legal retention minimums, sector-specific rules, and consent and legal-hold status as hard constraints, separate from the scored, softer signals. The fifth layer is the expiry decision engine, which combines the score from the fourth layer with the constraints from the fourth layer to select an action: retain, archive, anonymize, or delete. The sixth layer is the execution and audit layer, which carries out the selected action against both the originating data store and any machine learning model trained on the affected data, and logs the action for audit and human review. Fig. 2 depicts this layered flow.

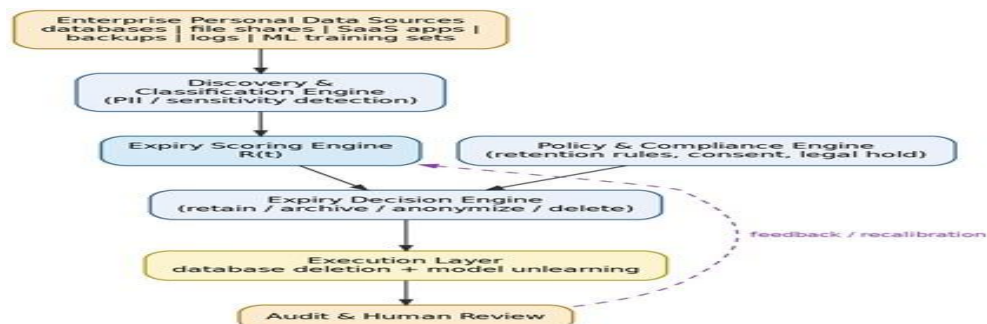


Fig. 2. Layered architecture of the proposed AI-Powered Personal Data Expiry Management System.

Fig. 3 focuses on the data-flow perspective across distributed stores. Because centralizing raw personal data from every store purely to compute an expiry score would itself increase privacy risk, classification and feature extraction are performed locally at each store, and only aggregated, privacy-preserving metadata, protected using techniques such as federated model updates or differentially private aggregation, is passed to the central scoring model, consistent with the privacy-preserving computation techniques discussed in Section II-F [14], [15]. Deletion and anonymization commands are then routed back to each originating store individually.

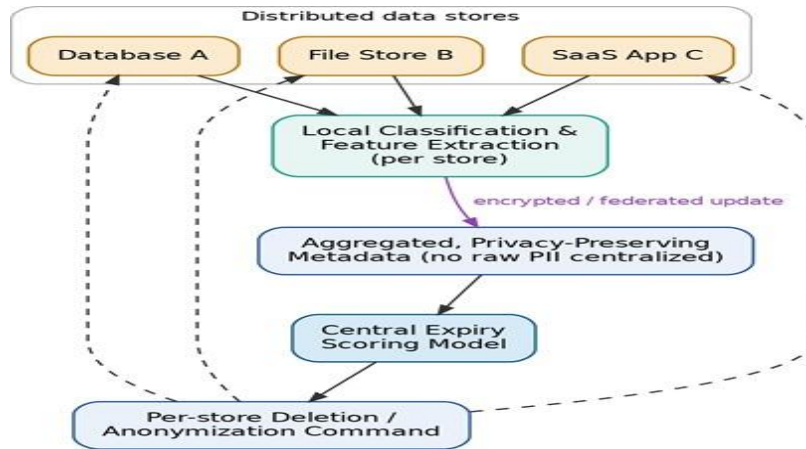


Fig. 3. Privacy-preserving data flow across distributed personal data stores.

C. Dataset and Data Considerations

No public dataset was identified that jointly provides, for the same records, ground-truth sensitivity labels, purpose and consent metadata, access recency logs, legal retention requirements, and an independently verified correct expiry decision. The closest partial precedents are the benchmark corpora used in automated PII detection research, such as the ai4privacy PII-masking corpus and synthetic financial-document collections used to train and evaluate ensemble PII classifiers [9], which provide sensitivity labels for text spans but no retention, consent, or legal-hold metadata, and the telecommunication datasets used in data-decaying research, which provide usage-recency and value-decay signals but no PII sensitivity labels [13]. Given this gap, an empirical evaluation of the full PDEMS pipeline would need to combine an existing PII detection benchmark with a separately constructed, clearly labeled synthetic or simulated retention-metadata layer, generated from parameterized models of purpose duration, consent lifetime, and legal retention periods calibrated against publicly documented regulatory minimums. Any results obtained from such a synthetic layer must be reported as simulated or synthetic results and not represented as an evaluation on real organizational data, consistent with the requirement that literature-reported, proposed, and experimental findings be kept clearly distinct throughout this paper.

D. Feature Engineering

Table I lists the variables used to construct the expiry score. Several of these variables, such as the composite data value score, are proposed derived quantities rather than directly observed signals and are marked accordingly.

TABLE I. SUMMARY OF VARIABLES USED IN THE PROPOSED EXPIRY SCORING FUNCTION

Variable	Description	Data Type	Source / Basis
Sensitivity level (S)	Classified sensitivity tier of the record	Ordinal	Discovery and classification engine [9]–[11]
Purpose fulfillment (P)	Estimated proportion of the original collection purpose already served	Continuous [0,1]	Business process metadata

Legal retention minimum (L)	Hard lower bound on retention derived from applicable law or sector rule	Duration (hard constraint)	Policy and compliance engine [1], [2]
Consent / legal-hold status (H)	Whether an active consent basis or legal hold currently applies	Boolean (hard constraint)	Consent management system / legal hold registry
Access recency (A)	Time elapsed since the record was last accessed or used	Continuous (days)	Access and audit logs
Data value decay (V, proposed)	Estimated remaining operational or model-training value as a function of age	Continuous [0,1]	Derived, grounded in data value decay literature [12], [13]
Redundancy duplication flag (D)	Whether the record is a duplicate or derivable from another retained record	Boolean	Deduplication / data catalog
Derived-model presence (M)	Whether the record contributed to a currently deployed machine learning model	Boolean	Model lineage / training-data provenance log
Storage cost weight (C)	Relative cost of continued storage for the record's storage tier	Continuous	Infrastructure metadata

E. Expiry Scoring Formulation

We define the expiry score of a record at time t as a weighted combination of the soft signals in Table I, evaluated only when the hard constraints described below permit deletion at all:

$$R(t) = w_1 (1 - P(t)) + w_2 \text{Decay}(A(t)) + w_3 S - w_4 V(t) + w_5 D - w_6 M \quad (1)$$

where $\text{Decay}(A(t))$ is a monotonically increasing function of time since last access, such as a normalized exponential or logistic curve, and w_1 through w_6 are non-negative weights calibrated per data category rather than fixed globally, since the relative importance of sensitivity, staleness, and value decay is expected to differ between, for example, marketing contact records and financial transaction records. A higher $R(t)$ indicates a stronger case for moving the record toward archival, anonymization, or deletion. The presence of a currently deployed model trained on the record, M , is included with a negative sign because deletion of such a record carries the additional cost and complexity of triggering model unlearning, discussed in Section III-H, and this cost is a legitimate factor in deciding whether immediate deletion or a scheduled batch unlearning cycle is more appropriate.

Critically, $R(t)$ is a soft signal used to prioritize and schedule review, not a value that can by itself authorize deletion. The decision rule is instead a constrained one:

$$\text{Action}(t) = \text{Delete/Anonymize only if } [t - t_0 \geq L] \text{ AND } [H(t) = \text{false}] \text{ AND } [R(t) \geq \tau] \quad (2)$$

where t_0 is the collection date, L is the legal retention minimum for the record's category, $H(t)$ is true whenever an active consent basis or legal hold currently applies, and τ is a category-specific decision threshold. This formulation deliberately places the legal retention minimum and the consent or legal-hold flag as binding constraints that the expiry score cannot override, reflecting the principle, established in data protection law, that a retention justification must be affirmatively absent before deletion is appropriate, rather than merely outweighed by a data-driven score [1], [2]. When either hard constraint blocks deletion, the record is instead flagged for archival or restricted access rather than left with no action taken.

F. Decision and Optimization Logic

Where $R(t)$ exceeds the threshold τ for a large batch of records simultaneously, for example following a bulk data migration or a policy change, the execution layer must also decide how to sequence the resulting deletions and anonymizations, since immediate execution of every eligible action could overwhelm downstream systems or exceed a scheduled maintenance window. We propose ranking eligible actions by a combination of $R(t)$, storage cost weight C , and estimated execution risk,

and processing them in descending order within each maintenance window. This is presented as a proposed scheduling heuristic rather than a novel optimization contribution, since batch scheduling under capacity constraints is a well-studied problem in its own right and is not the primary focus of this paper.

G. Privacy-Preserving Classification and Scoring

As introduced in Section III-B and Fig. 3, the discovery, classification, and feature-extraction steps that feed the expiry score are performed locally at each data store rather than by centralizing raw records. Where a shared classification or scoring model benefits from patterns observed across multiple stores or business units, that model can be trained using a federated learning protocol, in which only model updates, rather than raw personal data, are exchanged with a central aggregator [15]. Where aggregate statistics about a data category, such as the proportion of records exceeding a given age, must be shared more broadly, for instance in a compliance dashboard, differentially private mechanisms can be used to bound the extent to which any individual record influences the published statistic [14]. Where classification metadata for a small set of records must be pooled centrally for manual review, k-anonymity or a comparable generalization technique can be applied to that metadata before it leaves the originating store [16]. These are proposed as complementary safeguards to be selected according to the sensitivity of the specific data flow, not as a single mandatory pipeline.

H. Execution and Machine Unlearning

The final architectural component extends deletion beyond the originating data store. When the decision rule in Eq. 2 authorizes deletion or anonymization of a record flagged with an active derived-model presence, M , the execution layer must also trigger an unlearning operation against every model identified in the training-data provenance log as having used that record. Consistent with the taxonomy in Section II-E, the execution layer supports two unlearning modes. Exact unlearning retrains only the shard of a model that contained the deleted record, following a sharded training design such as SISA, and is preferred when the model was originally partitioned to support this [8].

Approximate unlearning instead adjusts the existing model's parameters to reduce the deleted record's estimated influence without full retraining, and is used when exact retraining is computationally impractical, at the cost of a weaker removal guarantee that should be disclosed in the audit log [6], [7]. Fig. 4 depicts the resulting feedback cycle, and Algorithm 1 specifies the full discovery-to-deletion process procedurally.

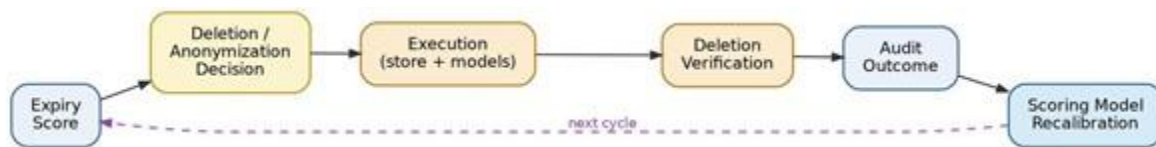


Fig. 4. Expiry scoring and execution feedback loop, including deletion verification and model recalibration.

Algorithm 1: AI-Powered Personal Data Expiry and Deletion Cycle

Input: connected personal data sources; policy and legal retention rules; consent and legal-hold registry; model lineage log; category-specific weights $w_1..w_6$ and threshold τ

Output: per-record action log (retain / archive / anonymize / delete) and, where applicable, model unlearning confirmations

```

1: for each connected data source do
2:   Run local discovery and classification to identify personal data and assign sensitivity S
3:   Extract local features: P, A, D, M, C (Table I)
4:   Send only aggregated / encrypted feature updates to the central scoring model (Fig. 3)
5: end for
6: for each candidate record or record group do
7:   Compute expiry score  $R(t)$  using Eq. 1
8:   Check hard constraints: legal retention minimum  $L$  and consent / legal-hold flag  $H(t)$ 
9:   if  $(t - t_0 < L)$  or  $(H(t) = \text{true})$  then
10:    Flag record for archival / restricted access; continue
11:   end if
12:   if  $R(t) \geq \tau$  then

```

- 13: Select action in {Anonymize, Delete} based on category policy and storage cost weight C
- 14: if M = true then
 - 15: Identify all models trained on this record from the provenance log
 - 16: Trigger exact or approximate unlearning for each identified model (Section III-H)
- 17: end if
- 18: Execute deletion/anonymization at the originating store 19: Verify execution and record outcome

IV. RESULTS AND DISCUSSION

As established in Section III-C, no dataset available to the authors jointly provides the sensitivity, purpose, consent, and legal-retention metadata needed to evaluate the full PDEMS pipeline, and no experiments were conducted for this paper. This section presents a conceptual and literature-grounded discussion of expected system behavior, a proposed evaluation protocol, and a comparison with existing data lifecycle management approaches. Any quantitative figure mentioned below is explicitly attributed to the literature that reported it; no accuracy, deletion-rate, or storage-savings figures are original to this paper.

A. Proposed Evaluation Protocol

A rigorous empirical evaluation of the PDEMS would require, at minimum, four components. First, a labeled corpus combining an existing PII detection benchmark, such as the ai4privacy PII-masking corpus used in recent ensemble PII detection research, with a synthetic retention-metadata layer specifying purpose duration, consent lifetime, and applicable legal retention minimums [9]. Second, a held-out temporal evaluation in which the scoring model is trained on records collected before a cutoff date and evaluated on records collected after it, since retention-relevant data is inherently time-ordered rather than exchangeable. Third, a deliberately constructed adversarial or edge-case test set containing records with conflicting signals, such as a low-sensitivity record still covered by an active legal hold, to verify that the hard-constraint logic in Eq. 2 is never overridden by the soft expiry score. Fourth, an unlearning verification step that, following each simulated deletion, checks whether the targeted model's predictions on held-out data still reflect the deleted record's influence, using the verification approaches surveyed in the machine unlearning literature [6], [7]. None of these steps have been executed for the present paper; they are specified as the minimum protocol required before the framework's classification, scoring, or unlearning claims could be empirically supported.

B. Evaluation Metrics

Table III lists metrics proposed for evaluating a future PDEMS implementation, distinguishing classification-quality metrics from decision-outcome and privacy-verification metrics.

TABLE III. PROPOSED EVALUATION METRICS FOR AN AI-POWERED PERSONAL DATA EXPIRY MANAGEMENT SYSTEM

Metric	Applies To	Description	Justification
Precision / Recall / F1	PII discovery and classification	Standard detection-quality metrics for identified personal data spans	Directly comparable to reported ensemble PII detection results [9]
False-deletion rate	Decision outcome	Proportion of deleted or anonymized records later found to require retention	Directly measures the cost of an overly aggressive threshold
Over-retention rate	Decision outcome	Proportion of eligible records not yet actioned within a defined review window	Measures compliance and storage-cost risk from under-action
Legal-hold violation rate	Hard-constraint compliance	Proportion of actions that would have violated an active legal hold or consent flag absent the hard-constraint check	Verifies Eq. 2's constraint logic is never bypassed

Unlearning verification pass rate	Model unlearning	Proportion of triggered unlearning operations that pass a post-hoc influence-removal check	Assesses whether deletion genuinely propagates to derived models [6]–[8]
Storage cost reduction	System outcome	Reduction in storage cost attributable to executed actions	Ties expiry management to its operational business case
Scoring latency	System	Time to compute an updated expiry score after new access or usage data arrives	Assesses feasibility of near-real-time scoring at enterprise scale
Cross-domain generalization gap	Classification	Difference in discovery accuracy between in-distribution and out-of-distribution document types	Reflects the domain-shift limitation reported in [9]

C. Discussion of Literature-Reported Evidence

Although the PDEMS framework itself has not been evaluated, related literature provides indicative, though not directly transferable, evidence that its component techniques are individually feasible. The ensemble PII detection system most closely related to the discovery layer proposed in Section III-B reported a weighted F1 of 0.9749 on a general-purpose benchmark, narrowly exceeding its strongest individual transformer-based component, but its accuracy on an out-of-distribution financial-document set was noticeably lower, which the original authors attribute to calibration sensitivity for entity classes underrepresented in the ensemble weighting corpus [9]. This is a literature-reported result from a system evaluated on its own benchmarks; it is cited here only as evidence that ensemble-based personal data discovery is feasible in principle, not as a performance claim for the discovery layer of the PDEMS proposed in this paper, whose accuracy on any specific organizational dataset remains an open empirical question. Separately, economic analysis of data aging has reported that retraining on excessively old data can measurably reduce model accuracy relative to using a smaller, more recent sample, with the rate of relevance decay varying substantially across data domains [12]; this finding supports the inclusion of a value-decay term in Eq. 1 but does not by itself establish the correct decay function or weight for any specific organizational data category.

D. Conceptual Walkthrough of the Decision Rule

To illustrate, not to validate, the intended behavior of the decision rule, consider a hypothetical, illustrative record representing a former customer's contact details. Suppose the record's legal retention minimum L has already elapsed, no legal hold or active consent applies, so $H(t)$ is false, and the record has not been accessed in a long period, giving a high $\text{Decay}(A(t))$ term, while its sensitivity S is moderate and it is not linked to any deployed model, so M is false. Under Eq. 1, these conditions would be expected to push $R(t)$ above a typical category threshold τ , and under Eq. 2 the hard constraints would already be satisfied, so the record would be routed to anonymization or deletion. If instead the same record were still covered by an active marketing consent, $H(t)$ would be true and the record would be retained regardless of how high $R(t)$ climbed. This example uses illustrative, non-empirical values and must not be read as a demonstrated or measured outcome.

E. Comparative Analysis

Table IV situates the proposed PDEMS relative to the main categories of prior work reviewed in Section II, along five qualitative dimensions: whether the approach continuously scores retention value rather than applying a static schedule; whether it performs automated, learning-based personal data discovery; whether deletion decisions respect legal and consent constraints as hard limits; whether deletion is propagated into trained machine learning models; and whether the system closes an explicit feedback loop back into its own scoring model. The comparison is conceptual and feature-based rather than a benchmark of measured performance, since the underlying systems were evaluated, where evaluated at all, on different tasks, datasets, and organizational contexts.

TABLE IV. CONCEPTUAL COMPARATIVE ANALYSIS OF EXISTING APPROACHES AND THE PROPOSED PDEMS

Approach	Continuous Expiry Scoring	Automated PII Discovery	Hard Legal / Consent Constraints	Model Unlearning Integration	Feedback Loop
Commercial data lifecycle management platforms	No; static administrator schedules	Yes; discovery and classification are mature	Partial; policy-based, not scored	No; database-only deletion	Limited; manual policy review
Automated PII detection research [9]–[11]	No; classification only	Yes; primary focus of this literature	No	No	No
Data value decay research [12], [13]	Yes, for value; not for legal/consent status	No	No	No	No
Machine unlearning research [6]–[8]	No	No	No; assumes deletion request already exists	Yes; primary focus of this literature	No
Proposed PDEMS	Yes; continuously updated score	Yes; via discovery and classification layer	Yes; explicit hard-constrained decision rule	Yes; integrated execution layer	Yes; audit-driven recalibration

The comparison indicates that the proposed framework's contribution lies less in any single novel technique. Automated discovery, value-decay modeling, and machine unlearning each have precedent individually. Its contribution lies instead in combining continuous, AI-driven scoring with a hard-constrained decision rule and model-level unlearning within one closed-loop system, a combination that Sections II and IV suggest has received limited attention to date.

V. PRIVACY, SECURITY, AND ETHICAL CONSIDERATIONS

A system that continuously scores and acts on personal data across an entire organization concentrates significant power and risk, and this must shape the design rather than being addressed only after deployment. The privacy-preserving classification techniques described in Section III-G, namely local feature extraction, federated model updates, differential privacy for aggregate statistics, and k-anonymity for pooled classification metadata, should be treated as required safeguards rather than optional enhancements, since the scoring model itself becomes a new, centralized system that processes signals derived from personal data across the organization [14]–[16]. Access to the scoring model's outputs and to the audit log should be restricted and logged in the same way access to the underlying personal data would be, since a record's expiry score and action history are themselves informative about that record.

The hard-constraint design in Eq. 2 is itself a safety mechanism: it ensures that no combination of weights or threshold miscalibration in the softer expiry score can cause deletion of data still covered by a legal hold or an active consent basis. This asymmetry is intentional. A false retention, keeping data slightly longer than strictly necessary while a review occurs, is a compliance and cost risk that can be corrected later, whereas a false deletion of data still under legal hold can be irreversible and legally consequential, particularly where the deletion has also triggered machine unlearning against a deployed model. For this reason, the execution layer should default to archival with restricted access rather than immediate deletion whenever a hard-constraint check cannot be evaluated with high confidence, and irreversible actions should require a documented review step for record categories above a configurable sensitivity threshold. Explainability of each expiry decision, in the sense of being able to show which specific signals in Table I drove a given score, is also necessary both for data-subject transparency and for the audit and recalibration step described in Section III-H, since an opaque scoring model would make it difficult to diagnose why a false deletion or false retention occurred.

VI. LIMITATIONS

This work has several limitations that should be considered when interpreting its contribution. First, and most importantly, no

experiments were conducted. The architecture, scoring formulation, and algorithm are proposed and conceptually motivated by literature rather than empirically validated, and their real-world classification accuracy, false-deletion rate, and unlearning verification performance remain unknown. Second, as discussed in Section III-C, no existing public dataset jointly provides the sensitivity, purpose, consent, and legal-retention metadata needed for a full evaluation, so even a retrospective evaluation would first require a dedicated synthetic data construction effort or a data-sharing agreement with an organization willing to provide de-identified retention metadata. Third, the weights in Eq. 1 and the threshold in Eq. 2 are proposed as category-specific and calibrated rather than fixed, but this paper does not specify a calibration procedure beyond the general audit-driven recalibration described in Algorithm 1, and mis-calibration risk, in either direction, is a genuine operational concern for any deployment. Fourth, automated PII detection accuracy is known from the literature to degrade across domains and document types, as the cross-domain results reported for the ensemble system in Section IV-C illustrate, so discovery-layer errors could propagate into incorrect scoring or missed records regardless of how well the scoring layer itself performs

[9]. Fifth, machine unlearning research itself acknowledges that approximate unlearning methods offer weaker removal guarantees than exact retraining, and that verifying unlearning success is an open research problem in its own right, which means the execution layer's unlearning guarantees are only as strong as the underlying unlearning method selected for a given model [6], [7]. Finally, this paper has not addressed the organizational and change-management challenges of deploying an automated deletion system, including the risk that stakeholders may distrust or override an automated deletion recommendation, which is a practical adoption question distinct from the technical feasibility questions this paper addresses.

VII. FUTURE WORK

Several directions follow directly from the limitations above. The most immediate is the construction of a labeled evaluation dataset that combines an existing PII detection benchmark with a clearly synthetic retention-metadata layer, enabling the first genuine empirical evaluation of the scoring and decision logic proposed in Section III. Methodologically, future work should compare candidate expiry scoring models, such as gradient-boosted trees against simpler logistic or survival-analysis formulations, directly on such data, and should empirically test the sensitivity of the false-deletion and over-retention rates to different weight and threshold calibrations. On the unlearning side, future work should evaluate whether sharded training strategies such as SISA can be adopted retroactively for models not originally trained with unlearning in mind, and should compare exact and approximate unlearning cost and verification strength for realistic, non-uniform deletion request patterns [6]–[8]. On the privacy side, future work should quantify the accuracy or utility cost of performing classification and scoring under differential privacy and federated learning constraints relative to a centralized baseline, since this cost is currently only qualitatively motivated in Section III-G rather than measured. Finally, field studies within organizations that already operate data lifecycle management tooling would be valuable for assessing whether an AI-driven expiry score changes reviewer behavior and deletion outcomes relative to the static, manually authored schedules such tools currently rely on.

VIII. CONCLUSION

This paper proposed an AI-Powered Personal Data Expiry Management System, a framework that extends existing data lifecycle management practice, which has concentrated heavily on collection, storage, and classification [3]–[5], [9]–[11], into a continuous, AI-driven approach to the retention and destruction stages of the data lifecycle. Motivated by the storage limitation principle in data protection law [1], [2], by evidence that data loses operational and model-training value over time [12], [13], and by the observation that database-level deletion alone does not remove a record's influence from a trained machine learning model [6]–[8], we argued that existing tools generally apply static retention schedules rather than a continuously updated, evidence-based expiry estimate, and that database deletion and model unlearning are rarely integrated into a single governed workflow. We proposed a six-layer PDEMS architecture, an expiry scoring function combined with a hard-constrained decision rule that legal holds and active consent can never be overridden by, and an operating algorithm that carries an expiry decision through to model unlearning and audit. Because no dataset currently spans the metadata this framework requires and no experiments were conducted, this contribution should be read as a scoped, literature-grounded framework and research agenda rather than a validated system. We have accordingly distinguished throughout between established findings, our proposed methodology, and the conceptual, non-experimental discussion in Section IV, and have outlined the privacy, ethical, and empirical work, summarized in Sections V through VII, required before such a system could be responsibly built and deployed.

IX. DISCLOSURE AND CONFLICT OF INTEREST

The authors declare no known conflict of interest. No external funding is reported for this work. [Funding source, ethics approval, and data availability statements to be completed by the authors/institution prior to submission.]

REFERENCES

- [1] European Parliament and Council of the European Union, Regulation (EU) 2016/679 (General Data Protection Regulation), Official Journal of the European Union, 2016.
- [2] P. Voigt and A. Von dem Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Cham, Switzerland: Springer, 2017.
- [3] J. Koo, G. Kang, and Y.-G. Kim, "Security and privacy in big data life cycle: A survey and open challenges," *Sustainability*, vol. 12, no. 24, art. 10571, 2020.
- [4] S. Bose and D. Marijan, "A survey on privacy of health data lifecycle: A taxonomy, review, and future directions," *arXiv preprint arXiv:2311.05404*, 2023.
- [5] F. Rafiq, M. J. Awan, A. Yasin, H. Nobanee, A. M. Zain, and S. A. Bahaj, "Privacy prevention of big data applications: A systematic literature review," *SAGE Open*, vol. 12, no. 2, 2022.
- [6] T. T. Nguyen, T. T. Huynh, Z. Ren, P. L. Nguyen, A. W.-C. Liew, H. Yin, and Q. V. H. Nguyen, "A survey of machine unlearning," *ACM Transactions on Intelligent Systems and Technology*, 2025 (arXiv preprint arXiv:2209.02299).
- [7] H. Xu, T. Zhu, L. Zhang, W. Zhou, and P. S. Yu, "Machine unlearning: A survey," *ACM Computing Surveys*, vol. 56, no. 1, pp. 1–36, 2023.
- [8] L. Bourtole, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, and N. Papernot, "Machine unlearning," in *2021 IEEE Symposium on Security and Privacy (SP)*, 2021, pp. 141–159.
- [9] H. A. Alshamrani and M. Alnahari, "Token-level PII detection with symbolic, sequential, and transformer-based ensemble models," *International Journal of Advanced Computer Science and Applications*, vol. 17, no. 6, pp. 549–560, 2026.
- [10] P. Kulkarni and N. K. Cauvery, "Personally identifiable information (PII) detection in the unstructured large text corpus using natural language processing and unsupervised learning technique," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, pp. 508–516, 2021.
- [11] P. Kulkarni, N. K. Cauvery, and R. Hemavathy, "An advanced semantic feature-based cross-domain PII detection, de-identification, and re-identification model using ensemble learning," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 12, pp. 763–772, 2024.
- [12] N. Ardalani, J. Hestness, M. Iansiti, and E. Valavi, "Time and the value of data," *Harvard Business School Working Paper* 21-016, 2022.
- [13] C. Costa, A. Konstantinidis, A. Charalampous, D. Zeinalipour-Yazti, and M. F. Mokbel, "Continuous decaying of telco big data with data postdiction," *GeoInformatica*, vol. 23, no. 4, pp. 533–557, 2019.
- [14] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloq. Automata, Languages and Programming (ICALP)*, Springer LNCS 4052, 2006, pp. 1–12.
- [15] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, PMLR 54, 2017, pp. 1273–1282.
- [16] L. Sweeney, "k-Anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.