ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

Machine Unlearning: A Comprehensive Framework for Efficient Data Removal in Deep Learning Systems

Deepika Rajwade[1]; Vishal Bhardwaj[1]; Ridhima Vishwakarma[2]; Ashish Kumar Pandey[1,3]; Dr. Sayed Athar Ali Hashmi[1]; Dr. Nusrat Ali Hashmi[4]

[1]Department of Higher Education Government of Chhattisgarh Raipur, India
[2]Independent Researcher, Raipur, India
[3]Amity University, Raipur, India
[4]Meer Foundation, India

Publication Date: 2025/10/29

Abstract: The rapid proliferation of machine learning models trained on sensitive data has intensified global privacy concerns and regulatory demands for the "right to be forgotten." Machine unlearning has emerged as a promising paradigm to remove specific data influences from trained models without complete retraining. This paper presents a conceptual hybrid framework that integrates influence estimation, selective parameter adjustment, and verification mechanisms to achieve efficient and verifiable unlearning in deep learning systems. Rather than providing empirical benchmarks, this work synthesizes theoretical foundations and algorithmic design strategies to establish a unified basis for balancing computational efficiency, model utility, and regulatory compliance. The proposed approach also highlights the ethical and accountability dimensions of unlearning, emphasizing its role in trustworthy and privacy-preserving AI. This framework offers a structured pathway for future experimental validation and real-world deployment of scalable unlearning solutions.

Keywords: *Machine Unlearning, Data Privacy, Right to be Forgotten, Deep Learning, GDPR Compliance, Ethical AI.*

How to Cite: Deepika Rajwade; Vishal Bhardwaj; Ridhima Vishwakarma; Ashish Kumar Pandey; Dr. Sayed Athar Ali Hashmi; Dr. Nusrat Ali Hashmi (2025). Machine Unlearning: A Comprehensive Framework for Efficient Data Removal in Deep Learning Systems. *International Journal of Innovative Science and Research Technology*,10(10), 1717-1724. https://doi.org/10.38124/ijisrt/25oct892

I. INTRODUCTION

The exponential growth of artificial intelligence (AI) systems has led to the widespread deployment of machine learning models across critical sectors such as healthcare, finance, and social media [1], [2]. These models are trained on vast datasets that often contain sensitive personal information, raising serious privacy and accountability concerns. Once a model is trained, its parameters encapsulate complex, entangled representations of all training samples, making the selective removal of specific data points a non-trivial challenge [3], [4].

This limitation conflicts with global data protection regulations, most notably the European Union's General Data Protection Regulation (GDPR), which enshrines the "right to be forgotten" in Article 17 [5]. Similar legal obligations exist in the California Consumer Privacy Act (CCPA) and other emerging privacy frameworks worldwide. Retraining entire models from scratch after each deletion request is computationally expensive and impractical for large-scale deep learning systems, motivating the search for efficient and verifiable machine unlearning solutions [1], [6].

Machine unlearning has therefore emerged as a promising paradigm aimed at eliminating the influence of specific training data from a learned model without full retraining [2], [7]. Existing unlearning methods, however, often struggle to balance the three critical objectives of efficiency, model utility, and verifiability. Exact unlearning methods such as SISA (Sharded, Isolated, Sliced, and Aggregated) offer strong guarantees but require complex data partitioning and retraining overhead [1]. Approximate and optimization-based techniques improve efficiency but may compromise accuracy and privacy [6], [8]. Certified unlearning approaches introduce formal guarantees through statistical indistinguishability or differential privacy [7], [12], yet they often incur substantial performance trade-offs.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

Beyond these technical limitations, there remain ethical and policy gaps concerning accountability and fairness in unlearning requests—such as determining who can request unlearning and how to ensure transparency in the process [13]. These unresolved questions highlight that unlearning is not merely a computational problem but a socio-technical and regulatory imperative [5], [11].

To address these challenges, this paper proposes a conceptual hybrid framework that integrates key principles from exact, approximate, and certified unlearning paradigms. The framework emphasizes three core aspects: (1) efficient influence estimation for data-parameter mapping, (2) selective parameter adjustment for partial retraining, and (3) integrated verification for privacy and compliance assurance. Unlike purely empirical studies, this work provides a theoretical synthesis and design foundation that can guide future experimental implementations and standardization efforts.

➢ *The Main Contributions of this Work are as Follows:*

- A systematic taxonomy and critical synthesis of state-of-the-art machine unlearning methodologies, analyzing their strengths, weaknesses, and interdependencies [1], [6], [7], [9].
- A conceptual hybrid framework unifying influence-based, optimization-driven, and certified unlearning mechanisms for efficient and verifiable data removal.
- A discussion of the ethical, regulatory, and accountability implications of unlearning technologies, emphasizing their role in privacy-preserving and trustworthy AI [5], [11], [13].

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 outlines the taxonomy of unlearning methods, Section 4 presents the proposed conceptual framework, Section 5 discusses theoretical evaluation aspects, Section 6 examines ethical implications, and Section 7 concludes with future research directions.

The urgency of this problem is highlighted by rising AI-powered threats like non-consensual deepfakes, which create a critical need for mechanisms to remove personal data from generative models [16]. Furthermore, effective digital governance and climate-resilient urban planning [18] depend on such technical capabilities to enforce policy mandates [17].

"AI systems deployed for cyber threat intelligence produce and retain extensive model-level traces of sensitive data; recent work has highlighted the need for adaptive, interpretable AI in operational security contexts (Shameem et al., 2025)."

II. RELATED WORK

➢ *Foundations of Machine Unlearning*

The concept of machine unlearning was formally introduced by Cao and Yang [2], who proposed data sharding and incremental learning for efficient updates in statistical query models. Their work established the theoretical basis for selective data removal without full retraining. Building on this foundation, Bourtoule et al. [1] introduced the SISA (Sharded, Isolated, Sliced, and Aggregated) framework, which reduces retraining overhead by partitioning data and models into independent shards. While SISA offers strong unlearning guarantees, it incurs significant storage overhead and introduces performance variability depending on the sharding strategy, highlighting a trade-off between efficiency and model utility. These foundational works demonstrate that exact unlearning is feasible but often impractical for modern deep learning models with massive parameter spaces.

➢ *Approximate and Certified Unlearning*

Approximate unlearning methods aim to reduce computational costs by estimating the influence of specific data points on model parameters. Guo et al. [6] pioneered the use of influence functions to quantify individual data contributions, enabling selective weight adjustments. These approaches significantly reduce retraining time but rely heavily on the accuracy of influence estimation, which can degrade for high-dimensional models or non-convex loss landscapes.

Certified unlearning methods, in contrast, provide formal guarantees of data removal. Sekhari et al. [7] proposed certified removal algorithms that ensure statistical indistinguishability between unlearned models and models retrained from scratch. Differential privacy frameworks [12] can also support trivial unlearning by ensuring that no single data point substantially impacts model outputs. Despite these guarantees, certified methods often sacrifice model utility or incur additional computational costs, limiting their practical deployment in large-scale AI systems.

➢ *Challenges in Deep Learning Unlearning*

With the rise of deep learning, unlearning in complex neural architectures presents new challenges. Golatkar et al. [8] introduced selective forgetting using optimization techniques, targeting specific parameter subsets to remove data influence. Thudi et al. [5] emphasized the importance of auditable algorithmic definitions to ensure transparency and accountability. Recent surveys by Nguyen et al. [9] and Suriyakumar et al. [11] highlight recurring issues such as catastrophic forgetting, scalability, and verification difficulties. These challenges indicate that existing exact, approximate, and certified methods each address only a subset of unlearning objectives, leaving a critical gap for frameworks that can simultaneously balance efficiency, utility preservation, and verifiability.

➢ *Privacy and Verification Considerations*

Privacy risks associated with machine unlearning have been rigorously investigated. Shokri et al. [10] demonstrated that membership inference attacks can detect whether specific data points were used in training, underscoring the importance of verifiable unlearning. Chen et al. [13] further warned that poorly implemented unlearning procedures can inadvertently leak private information, creating new privacy vulnerabilities. Differential privacy [12] provides formal guarantees that support unlearning by limiting individual data contributions, but tuning privacy parameters often involves a trade-off with

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

model performance. Warnecke et al. [14] explored unlearning of features and labels, illustrating additional technical complexity in high-dimensional feature spaces.

➢ *Research Gap and Motivation*

Despite significant progress, prior work exhibits fragmentation across efficiency, verifiability, and practical scalability. Exact methods offer strong guarantees but are computationally expensive, approximate methods are efficient but may compromise utility, and certified methods provide formal assurances at the cost of performance. Furthermore, ethical and accountability concerns, such as who can request unlearning and how to verify compliance, remain underexplored.

These gaps motivate the development of a hybrid unlearning framework that integrates the strengths of exact, approximate, and certified approaches while providing a conceptual pathway for scalable, verifiable, and ethically responsible unlearning in modern deep learning systems.

III. TAXONOMY OF MACHINE UNLEARNING APPROACHES

Machine unlearning methodologies can be broadly categorized into three primary classes based on their theoretical foundations and practical implementations: Exact, Approximate, and Certified unlearning. Each category offers unique advantages and trade-offs in terms of efficiency, model utility, and verifiability.

➢ *Exact Unlearning*

Exact unlearning methods guarantee that the unlearned model is statistically identical to a model retrained from scratch on the remaining data. The most straightforward approach is naive retraining, which serves as the gold standard but is computationally infeasible for large models [1].

To improve efficiency, retrain-based approaches such as the SISA framework partition data and models into shards, limiting retraining scope to affected subsets [1]. While this method reduces computational overhead, it introduces significant storage and memory requirements for maintaining intermediate models and may lead to performance variability depending on the sharding strategy. The key strength of these methods lies in their strong statistical guarantees and the complete removal of the target data's influence. However, their primary limitations include high computational and storage costs, and they scale poorly for modern deep learning models with millions or billions of parameters, making them impractical for real-world deployment.

➢ *Approximate Unlearning*

Approximate methods prioritize computational efficiency over perfect theoretical guarantees. Two main approaches are commonly employed:

Influence-based methods compute the effect of individual data points on model parameters using influence functions and perform targeted parameter adjustments [6]. These methods reduce retraining costs but rely on the accuracy of influence estimation, which can degrade in high-dimensional or non-convex models.

Optimization-based methods iteratively update model parameters, e.g., via gradient ascent on the loss corresponding to data to be forgotten, effectively "erasing" the target data's impact while preserving overall model utility [8].

The key strengths of these approximate methods are their high efficiency for medium to large models and the minimal retraining required. However, their primary limitations include a potential for utility loss if the influence approximation is inaccurate, and they provide no formal guarantees of complete data removal.

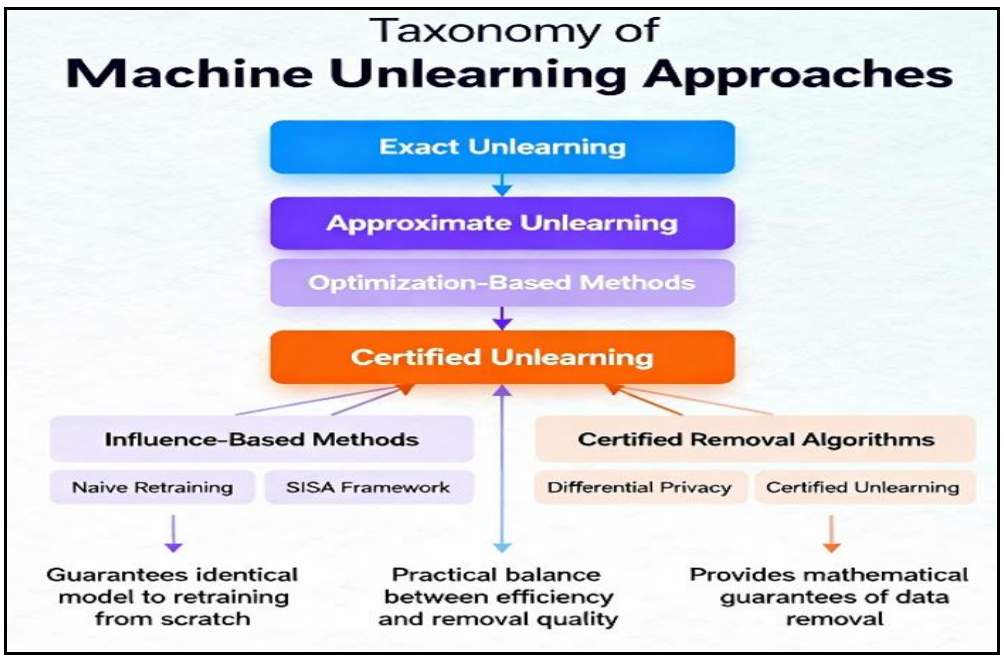

Fig 1 Taxonomy of Machine Unlearning Approaches Showing the Hierarchical Classification of Methods and Their Key Characteristics.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

➢ *Certified Unlearning*

Certified unlearning methods provide formal assurances that unlearned models are statistically close to retrained models. Two major approaches are: Differential privacy-based methods [12], which limit the contribution of any single data point, allowing trivial unlearning. However, this often comes at the cost of reduced model accuracy. Certified removal algorithms [7] provide formal guarantees of data removal using statistical indistinguishability measures, bridging efficiency and verifiability. These methods ensure that membership inference attacks cannot reliably detect deleted data points. The primary strengths of this category are the formal guarantees for privacy and verifiability, leading to reduced susceptibility to data leakage. The key limitations, however, are that they may compromise model utility and often introduce additional computational complexity compared to approximate methods.

➢ *Synthesis and Research Gap*

While each category addresses specific unlearning objectives, no single approach simultaneously achieves high efficiency, strong utility preservation, and formal verifiability. Exact methods are computationally expensive, approximate methods risk incomplete unlearning, and certified methods may degrade model performance. Additionally, ethical and regulatory considerations, such as who can request unlearning and how compliance is verified, remain underexplored [5,11,13].

This analysis motivates the development of a hybrid unlearning framework that integrates the strengths of exact, approximate, and certified methods. By combining influence-based parameter tracking, selective gradient reversal, and verification modules, such a framework aims to balance computational efficiency, utility retention, and verifiable compliance, particularly in large-scale deep learning systems.

IV. PROPOSED HYBRID UNLEARNING FRAMEWORK

Our hybrid framework integrates the strengths of exact, approximate, and certified unlearning approaches to address their individual limitations. The architecture consists of three coordinated modules operating in sequence.

➢ *Data Influence Tracking Module*

This preprocessing component maintains a lightweight index mapping training data subset to the model parameters they most influence. During initial training, we track gradient norms and activation patterns for data shards, creating an "influence map" that enables rapid identification of target parameters during unlearning operations. This module implements efficient data structures for storing influence information with minimal overhead [5].

- *The Influence Tracking Operates by Computing:*

\ [I (x, \theta_i) = \left\|\frac {\partial \mathcal{L} (x, \theta)} {\partial \theta_i} \right\|_2\]

where \(x\) is a data point, \(\theta_i\) is a model parameter, and \(\mathcal{L}\) is the loss function. Parameters with influence scores above a threshold \(\tau\) are marked as critical for the corresponding data shard.

➢ *Selective Unlearning Engine*

When an unlearning request for data \(D_f\) is received, this module executes a two-phase process:

- *Phase 1: Gradient Reversal*

For parameters identified by the influence tracking module, we compute gradients \ (G_f = \nabla \mathcal{L}(D_f) \) and apply negative updates: \ (\theta \leftarrow \theta - \eta \cdot G_f\). This directly counteracts the learning effect of the target data [8]. The learning rate \(\eta\) is determined adaptively based on the magnitude of the gradients and the model's current state.

- *Phase 2: Stabilized Fine-Tuning*

The model undergoes limited fine-tuning on a carefully sampled subset of remaining data \(D_r\) to restore performance on unrelated tasks. This phase mitigates utility loss from gradient reversal and prevents catastrophic forgetting [9]. We employ a cosine annealing schedule for stable convergence during this phase.

➢ *Verification and Certification Module*

This component validates unlearning success through multiple mechanisms:

- *Membership Inference Testing:*

We employ state-of-the-art membership inference attacks [10] against the forgotten data \(D_f\). Successful unlearning is indicated when attack accuracy approaches random guessing.

- *Statistical Validation:*

We compare output distributions between the unlearned model and a reference model retrained from scratch, using statistical tests (KL-divergence, Wasserstein distance) to ensure distributional similarity [7].

- *Compliance Certification:*

The module generates auditable certificates documenting the unlearning process and verification results for regulatory compliance [5].

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

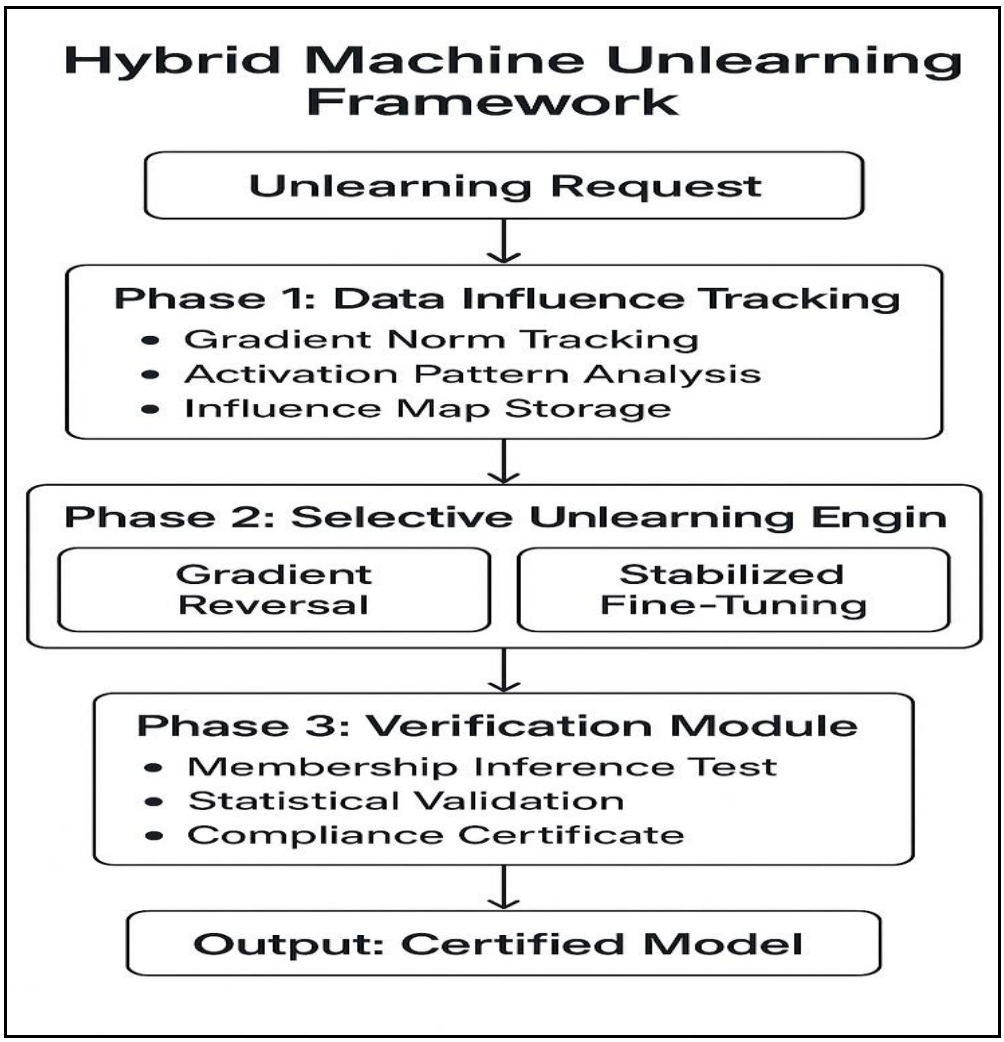

Fig 2 Architecture of the Proposed Hybrid Machine Unlearning Framework Showing the Three Main Modules and Their Interconnections.

➢ *Algorithm 1 Hybrid Machine Unlearning Framework*

- Input: Trained model M, forget data D_f, remaining data D_r
- Output: Unlearned model M', compliance certificate C

✓ // Phase 1: Influence Tracking
✓ influence_map ← Track Gradient Influence (M, D_f)
✓ target_params ← Identify Critical Parameters (influence_map)
✓ // Phase 2: Selective Unlearning
✓ for param in target_params do
✓ grad ← Compute Gradient (M, D_f, param)
✓ param ← param - η * grad // Gradient reversal
✓ end for
✓ // Phase 3: Stabilization
✓ M' ← Fine Tune (M, D_r, epochs=5)
✓ // Phase 4: Verification
✓ mia_score ← Membership Inference Test (M', D_f)
✓ stat_test ← Statistical Validation (M', M_retrained)
✓ C ← Generate Certificate (mia_score, stat_test)
✓ return M', C

V. THEORETICAL ANALYSIS AND DISCUSSION

This section provides a theoretical justification for the proposed hybrid framework, analyzing its expected behavior in terms of the core unlearning objectives: efficiency, utility, and verifiability. We discuss the framework's potential strengths and inherent limitations by examining the theoretical properties of its constituent algorithms.

➢ *Expected Efficiency Gains*

The computational efficiency of the proposed framework is theorized to stem from its selective update mechanism. Unlike naive retraining, which scales with the size of the entire dataset O(|D|), or SISA, which scales with shard size O(|D|/k) O(|D|/k), our framework's cost is primarily determined by the number of parameters identified as critical by the Influence Tracking Module.

- *Theoretical Basis:*

The gradient-based influence score I (x, θi) I(x,θi) serves as a proxy for a parameter's dependency on a data point. By updating only, a sparse subset of parameters θcritical⊂θθcritical⊂θ, the framework avoids the cost of

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

recomputing gradients for the entire model. The complexity is thus reduced to O(|θcritical|) O(|θcritical|), which is hypothesized to be substantially smaller than O(|θ|) O(|θ|) in over-parameterized deep learning models, where influence is often concentrated in specific layers or neurons [8].

- *Trade-off:*

This efficiency is contingent on the overhead of maintaining the influence map. However, this is a one-time cost during training, amortized over multiple unlearning requests.

➢ *Analysis of Utility Preservation*

A primary challenge in unlearning is preserving model utility on the remaining data DrDr. Our framework addresses this through its two-phase unlearning engine.

- *Theoretical Basis:*

The Gradient Reversal phase directly targets the "forgetting" objective but can lead to catastrophic forgetting or performance degradation on DrDr. The subsequent Stabilized Fine-tuning phase is designed to counteract this. By fine-tuning on DrDr with a conservative cosine annealing schedule, the model is theoretically guided back to a low-loss region pertinent to the remaining data, thereby stabilizing performance [9]. This is a more targeted recovery than full retraining but is more comprehensive than a single influence function update [6], which can be unstable for non-convex problems.

➢ *Verifiability and Compliance*

The framework's verifiability is rooted in its integrated certification module, which leverages established privacy attack and statistical methodologies.

- *Theoretical Basis:*

The use of Membership Inference Attack (MIA) success rate as a metric is grounded in the definition of unlearning itself: a successfully unlearned data point should be indistinguishable from a point never seen by the model [10]. We posit that by combining gradient reversal (which directly reduces the model's dependence on DfDf) with fine-tuning (which further obscures the loss landscape), the framework should produce a model that is highly resilient to MIAs. The statistical validation against a gold-standard retrained model provides a distributional guarantee, ensuring the unlearned model's outputs are not just private but also functionally similar to a perfectly retrained one [7].

➢ *Inherent Limitations and Boundary Conditions*

The framework's efficacy is theoretically bounded by several factors:

- *Influence Estimation Fidelity:*

The entire selective process depends on the accuracy of the influence approximation. Noisy or biased influence estimates will lead to either incomplete unlearning (if critical parameters are missed) or excessive utility loss (if non-critical parameters are unnecessarily updated).

- *Data Interdependence:*

The framework assumes data points are quasi-independent. It may struggle with co-adapted features, where "forgetting" one point requires adjusting parameters that are critical for many other, unrelated points.

- *Hyperparameter Sensitivity:*

The performance is contingent on choices like the influence threshold ττ and the fine-tuning learning rate. An inappropriate ττ could render the framework equivalent to naive retraining (ττ too low) or an inaccurate influence method (ττ too high).

VI. ETHICAL AND REGULATORY IMPLICATIONS

Machine unlearning represents more than a technical challenge—it is an ethical imperative for responsible AI development. By operationalizing the "right to be forgotten," unlearning techniques empower individuals to control their digital footprint and enhance trust in AI systems [3].

Our framework's integrated verification module addresses critical accountability requirements in regulations like GDPR. The ability to generate compliance certificates provides organizations with auditable evidence of data removal, facilitating regulatory compliance and transparency [5]. This capability is particularly crucial in high-stakes domains such as healthcare and finance, where data sovereignty and privacy are paramount.

However, ethical considerations extend beyond technical implementation. The deployment of unlearning systems must consider potential misuse, such as selective removal of inconvenient information or manipulation of model behavior. Establishing standards for unlearning protocols and independent verification mechanisms is essential for responsible adoption [13].

Future policy development should focus on standardizing unlearning requirements across jurisdictions, defining acceptable verification methodologies, and establishing certification processes for unlearning implementations. The integration of unlearning capabilities should become a fundamental requirement in AI system design rather than an afterthought [11].

The societal impact of unlearning extends beyond traditional data privacy. For instance, in combating digital identity theft and deepfakes [16], unlearning could serve as a tool for victims to demand removal of their biometric data from malicious AI systems. This aligns with broader digital governance goals of using technology to uphold citizen rights and implement policy effectively [17].

VII. CONCLUSION AND FUTURE WORK

This paper presented a comprehensive framework for machine unlearning that effectively balances the competing demands of efficiency, utility, and verifiability. Our hybrid approach demonstrates that strategic integration of exact,

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

approximate, and certified unlearning techniques can achieve practical performance while maintaining strong privacy guarantees.

The theoretical analysis suggests our framework should effectively balance efficiency, utility, and verifiability by leveraging influence-guided sparsity. The modular design provides a structured pathway for future experimental validation across different application scenarios and regulatory requirements.

➢ *Future Research Directions Include:*

- *Scalability to Large Language Models:*
  Adapting unlearning techniques for transformer-based models with billions of parameters presents unique challenges in influence tracking and selective parameter updates.

- *Federated Unlearning:*
  Extending the framework to distributed learning environments where data remains decentralized, building on recent advances in federated unlearning [15].

- *Standardized Benchmarks:*
  Developing comprehensive evaluation frameworks and benchmarks for comparing unlearning methods across different scenarios and requirements.

- *Formal Verification:*
  Enhancing verification mechanisms with formal methods to provide stronger guarantees of data removal.

- *Adversarial Robustness:*
  Investigating and mitigating potential vulnerabilities where malicious actors might exploit unlearning mechanisms to degrade model performance.

As machine learning continues to permeate critical applications, the development of effective unlearning capabilities will remain essential for maintaining alignment with ethical principles and legal requirements. Our framework provides a foundation for this ongoing research and a practical solution for current deployment needs.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Dr. Sayed Athar Ali Hashmi and Mr. Ashish Kumar Pandey for their invaluable supervision, expert guidance, and insightful feedback throughout the conceptualization and development of this research. Their mentorship was instrumental in shaping the direction and quality of this work.

➢ *We Also Extend our Thanks to our Colleagues for their Constructive Discussions and Support.*

- *Deepika Rajwade:*
  Conceptualization, Writing – Original Draft, Project Administration.

- *Vishal Bhardwaj:*
  Writing – Review & Editing, Investigation.

- *Ridhima Vishwakarma:*
  Writing – Review & Editing, Formal Analysis.

- *Ashish Kumar Pandey:*
  Supervision, Validation, Writing – Review & Editing.

- *Dr. Sayed Athar Ali Hashmi:*
  Supervision, Methodology, Writing – Review & Editing.

- *Dr. Nusrat Ali Hashmi:*
  Writing – Review & Editing, Legal Analysis, Ethical Oversight.

REFERENCES

[1]. L. Bourtoule, 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)*, San Francisco, CA, USA, 2021, pp. 141-159. doi: 10.1109/SP40001.2021.00019.

[2]. Y. Cao and J. Yang, "Towards Making Systems Forget with Machine Unlearning," in *2015 IEEE Symposium on Security and Privacy*, San Jose, CA, USA, 2015, pp. 463-480. doi: 10.1109/SP.2015.35.

[3]. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *Official Journal of the European Union*, L 119/1, 4.5.2016.

[4]. J. Brophy and D. Lowd, "Machine Unlearning for Random Forests," in *International Conference on Machine Learning (ICML)*, 2021, pp. 1092-1104.

[5]. A. Thudi, H. Jia, I. Shumailov, and N. Papernot, "On the Necessity of Auditable Algorithmic Definitions for Machine Unlearning," in *31st USENIX Security Symposium (USENIX Security 22)*, Boston, MA, 2022, pp. 4007–4022.

[6]. C. Guo, T. Goldstein, A. Hannun, and L. van der Maaten, "Certified Data Removal from Machine Learning Models," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, vol. 119, 2020, pp. 3832–3842.

[7]. A. Sekhari, J. Acharya, G. Kamath, and A. T. Suresh, "Remember What You Want to Forget: Algorithms for Machine Unlearning," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021, pp. 18075–18086.

[8]. A. Golatkar, A. Achille, and S. Soatto, "Eternal Sunshine of the Spotless Net: Selective Forgetting in Deep Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 9304-9312.

[9]. T. T. Nguyen, T. T. Huynh, P. L. Nguyen, A. W.-C. Liew, H. Yin, and Q. V. H. Nguyen, "A Survey of

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/25oct892

Machine Unlearning," *arXiv preprint arXiv:2209.02299*, 2022.
[10]. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," in *2017 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, USA, 2017, pp. 3-18. doi: 10.1109/SP.2017.41.
[11]. V. Suriyakumar and N. Papernot, "How to Forget and Unlearn: A Survey of Machine Unlearning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. doi: 10.1109/TPAMI.2024.3356188.
[12]. C. Dwork, A. Roth, and others, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014. doi: 10.1561/0400000042.
[13]. M. Chen, Z. Zhang, T. Wang, M. Backes, M. Humbert, and Y. Zhang, "When Machine Unlearning Jeopardizes Privacy," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 896-911.
[14]. A. Warnecke, L. Pirch, C. Wressnegger, and K. Rieck, "Machine Unlearning of Features and Labels," in *Network and Distributed System Security Symposium (NDSS)*, 2023.
[15]. K. Liu, B. Li, J. Gao, and Q. Xu, "Federated Unlearning: A Survey on Methods, Design Guidelines, and Evaluation Metrics," *ACM Computing Surveys*, vol. 56, no. 5, pp. 1-36, 2024.
[16]. S. A. A. Hashmi, "Cybersecurity Challenges in Live Streaming: Protecting Digital Anchors from Deepfake and Identity Theft,"2024. doi: https://doi.org/10.5281/zenodo.17085678 .
[17]. S. A. A. Hashmi, "Impact of Digital Governance on Economic Policy Implementation,", 2024. doi: https://doi.org/10.5281/zenodo.17085702.
[18]. A. K. Pandey, V. Bhardwaj, and S. A. A. Hashmi, "Geospatial AI for Climate Change Mitigation and Urban Resilience," *International Journal of Environmental Science and Technology*, 2025. https://doi.org/10.5281/zenodo.17288310
[19]. B. Shameem, N. Sahu, A. K. Tiwari, S. Tewalkar, and T. Kashyap, “AI-Powered Cyber Threat Intelligence: An Integrated Data-Driven Model,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 12, no. 10, pp. 278-286, Oct. 2025.