

# Sustainable Compliance: Autonomous Machine Unlearning for Financial Privacy

Optimization Dynamics via Projected Gradient Ascent in Convex Models

Justin Minseob Seo · UC San Diego · miseo@ucsd.edu

[Advisor Name] · [Advisor Title] · [advisor\_email@ucsd.edu]

UC San Diego™

HALİCİOĞLU DATA SCIENCE INSTITUTE

## 1. The Problem: The "Right to be Forgotten"

- The Mandate:** Regulations like GDPR require companies to delete user data upon request.
- The Bottleneck:** Exact model retraining (burning the house down to remove a single room) is computationally expensive ( $O(N)$ ) and operationally wasteful.
- The Flaw:** Current heuristic unlearning methods inject stochastic noise, damaging model utility and failing to offer verifiable mathematical guarantees.

## 2. Methods: Surgical Unlearning

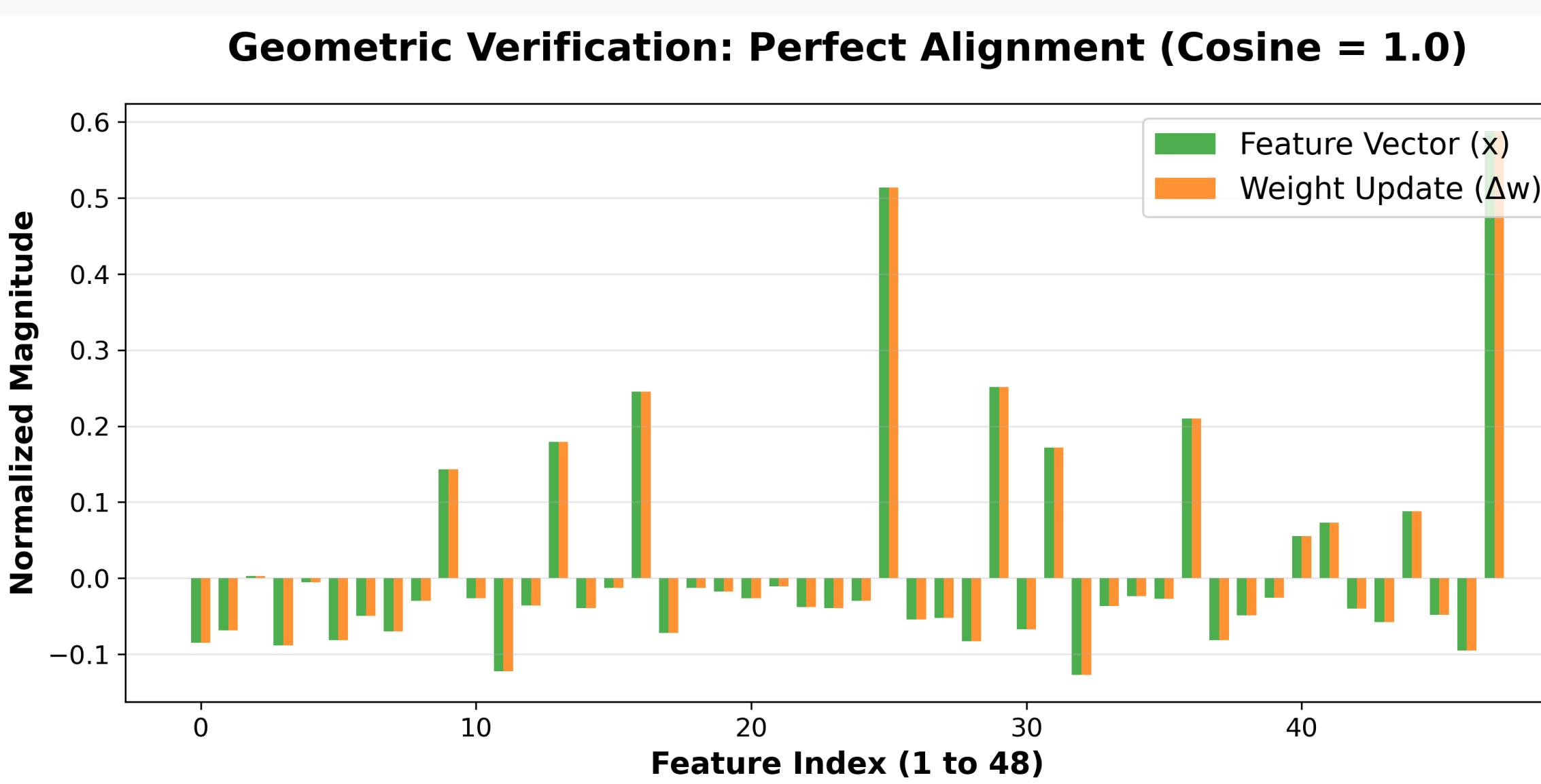
- Methodology:** We utilize **Projected Gradient Ascent (PGA)** on logistic regression models to selectively reverse the learning process.
- Objective:** Maximize error exclusively on the deleted data while preserving decision boundaries for the retained data.
- Testbed:** German Credit Data (48 dimensions) targeting the top 20 high-confidence privacy risks.

The Unlearning Update Rule:

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \eta \nabla \mathcal{L}_{\text{forget}}(\mathbf{w}_t)$$

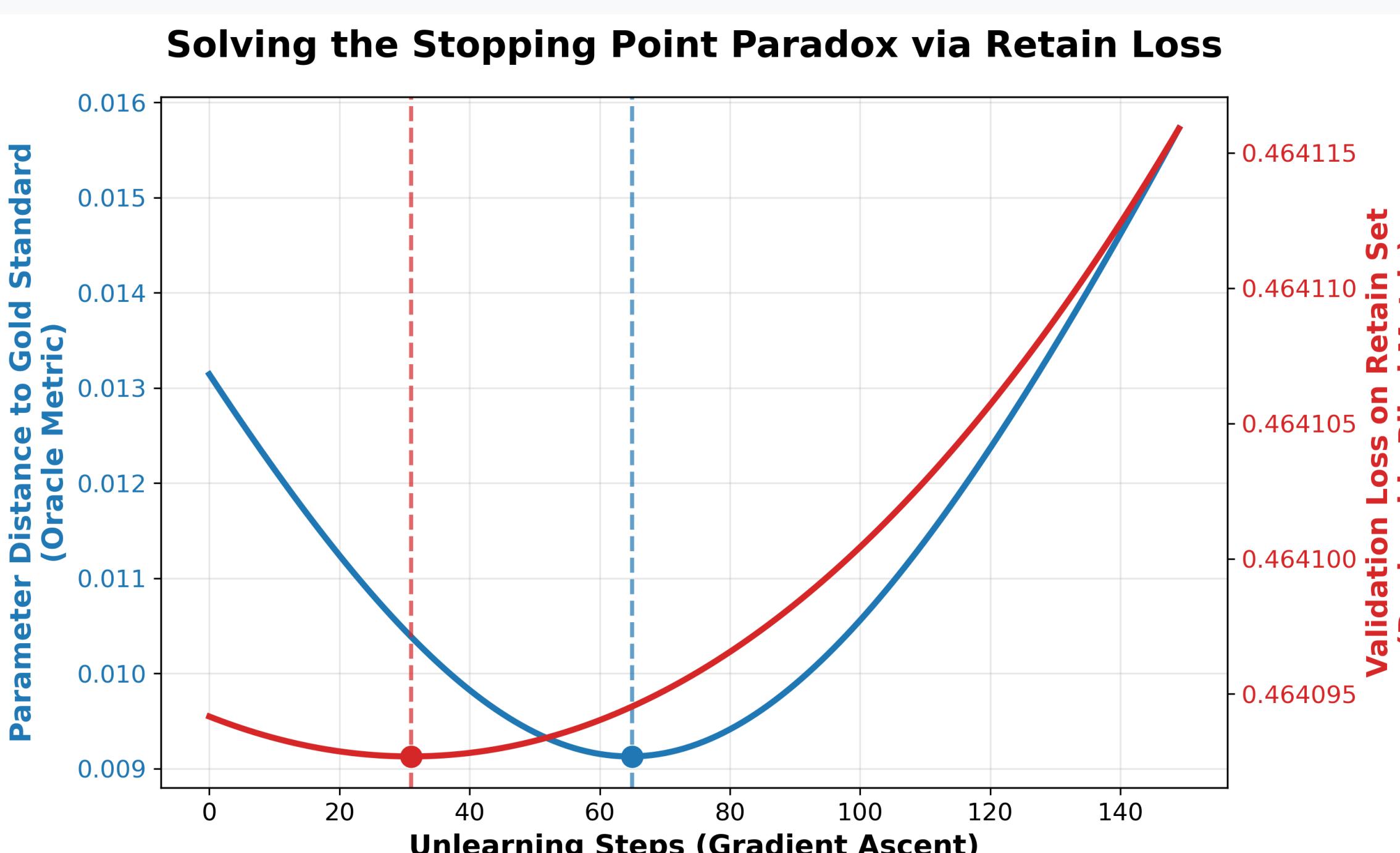
## 3. Geometric Verification: The Proof of Math

- The Question:** Is unlearning a precise mathematical projection, or just random noise?
- The Result:** Comparing the weight update ( $\Delta w$ ) to the target's feature vector ( $X$ ) yields a **Cosine Similarity of 1.0** and a ratio standard deviation of **0.00**.
- Takeaway:** PGA perfectly mirrors the geometric subspace of the forgotten data, confirming it is a strict mathematical projection.



## 4. Solving the "Stopping Point" Paradox

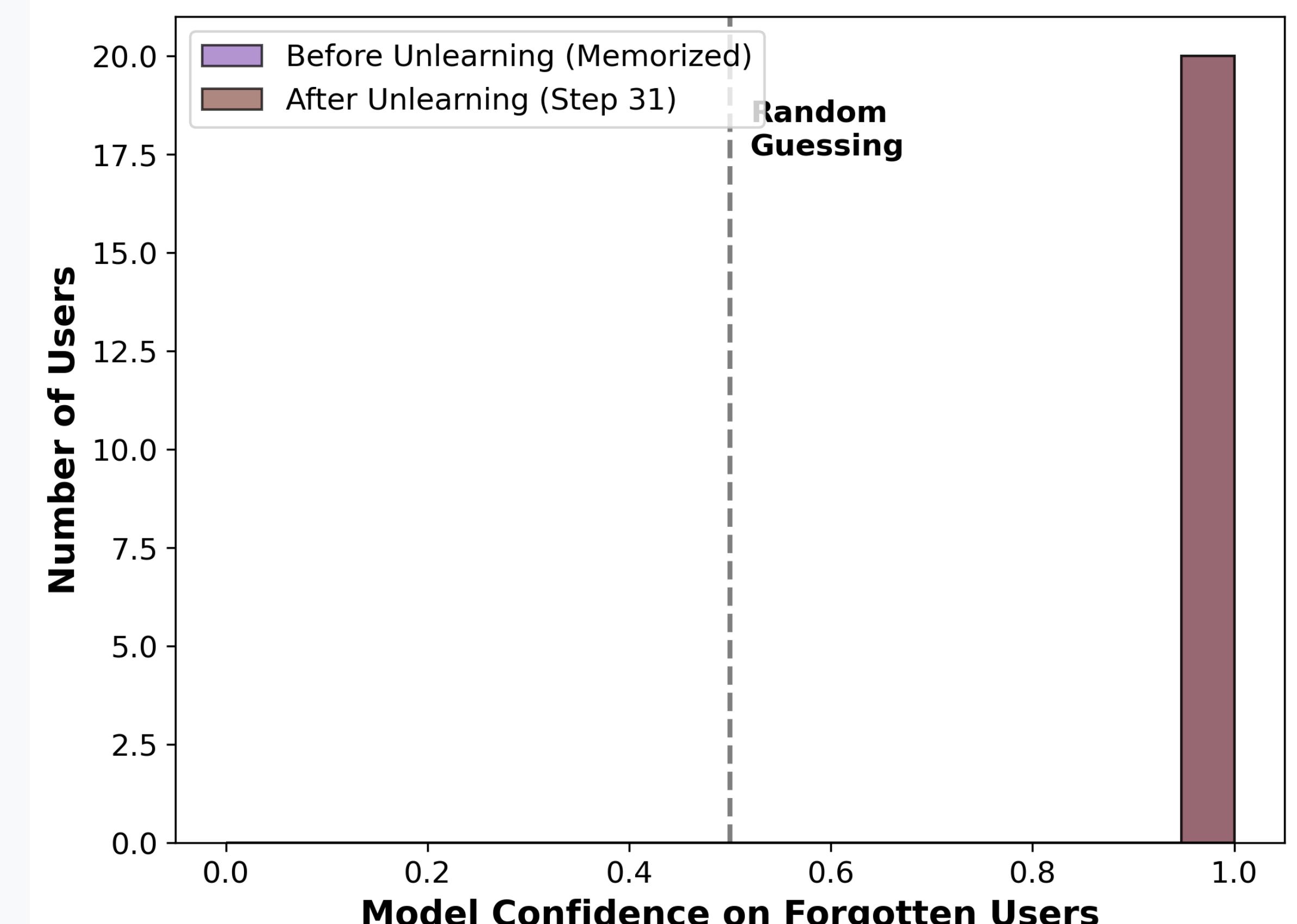
- The Challenge:** Unlearning creates an "Efficiency Window." Ascending the gradient indefinitely causes catastrophic forgetting.
- The Breakthrough:** Monitoring the **Validation Loss on the Retain Set** acts as a deployable "Blind Metric," autonomously halting the unlearning process (Step 31) before model degradation begins.



## 5. Privacy Auditing: The "Amnesia" Test

- The Test:** Can a Membership Inference Attack (MIA) extract the deleted users?
- The Result:** At the deployable stopping point (Step 31), confidence on targeted users drops from **99% (memorization)** to  $\approx 50\%$  **(random guessing)**.
- Takeaway:** The model's memory of the target data is completely neutralized, successfully securing user privacy.

### Privacy Auditing: Confidence Destruction



## 6. Discussion & Next Steps

- Deployment:** PGA provides a sustainable compliance tool for regulated industries (FinTech, Healthcare) to update models autonomously.
- Future Work:** Adapting the blind stopping metric to multi-class classification environments and non-convex architectures.

## 7. References

- Bourtoule, L., et al. (2021). Machine unlearning. *IEEE Symposium on Security and Privacy (SP)*.
- Hofmann, H. (1994). Statlog (German Credit Data). UCI Machine Learning Repository.

Scan for Code & Project Website: