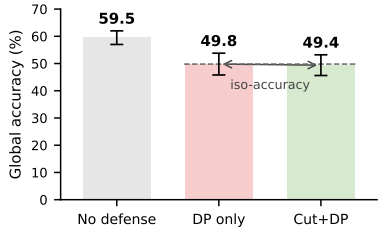
More Alignment Is Not Always More Useful in Federated Learning

Francesco Simoni[1,2], Andrej Jovanović[2], Nicholas D. Lane[2]
[1]University of Bologna [2]University of Cambridge

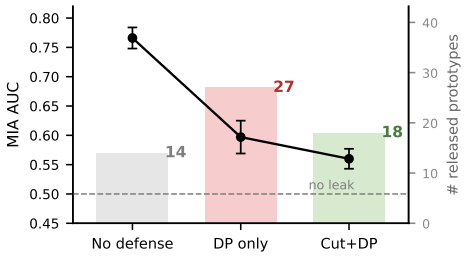
Federated Learning (FL) trains models across mobile and edge devices while keeping data local [1]. Real deployments are highly heterogeneous: clients differ in users, domains, sensing conditions, and class distributions [2], causing local representations to drift, biasing the global model toward dominant clients. Prototype-based FL mitigates this by sharing compact per-class prototypes as semantic anchors [3] to realign clients to the global objective. Prototypes can be constructed statically [3], [4] or *dynamically* [5], where their count varies across rounds. However, irrespective of the mechanism, prototypes are derived from training data and are vulnerable to membership-inference attacks [6]. Prior work typically adds Differential Privacy (DP) [7] to each released prototype [4], [6], but none has studied its effect in the dynamic clustering regime.

In this work, we study how DP affects prototype alignment in the dynamic clustering regime [5]: specifically, under what conditions does alignment remain beneficial once privacy constraints are imposed? We make two contributions. First, we apply Gaussian (ε, δ)-DP noise to every released prototype [6]. Second, we observe that many prototypes carry negligible signal under DP noise, motivating a simple, broadly applicable pre-release *cutting* mechanism that discards uninformative prototypes before noise is added. Experimentally, we evaluate four parameter-free cutting strategies across Digit-Five, Office-10, and DomainNet [5], tracking MIA AUC and TPR@1%FPR [6], global and worst-domain accuracy, and released pool size.

We find that DP consistently inflates the number of prototypes relative to the non-DP baseline; while this reduces privacy risk (reflected in lower MIA AUC), it simultaneously introduces many noisy and redundant prototypes that degrade alignment quality. Our cutting mechanism proves complementary to DP: it further reduces MIA AUC at no cost to accuracy on these benchmarks. This confirms that a substantial fraction of DP-protected prototypes are redundant and safely removable. Crucially, structure of the cut is important: random removal of prototypes impairs accuracy, suggesting that the *selection* of prototypes to discard plays a key role in maintaining model performance. Together, these results show that DP, alignment strength, and prototype count must be tuned jointly, and that a data-dependent pre-release cut is an effective, broadly applicable lever for the privacy–utility trade-off in prototype-based FL.



(a) DP noise drops accuracy, but cutting preserves it at iso-accuracy.



(b) Cut counteracts DP prototype inflation and reduces privacy risk.

Fig. 1: Office-10 evaluation metrics under DP and cutting.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

[2] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in *Proceedings of Machine Learning and Systems (MLSys)*, 2020.

[3] Y. Tan, G. Long, L. Liu, T. Zhou, Q. Lu, J. Jiang, and C. Zhang, "Fedproto: Federated prototype learning across heterogeneous clients," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

[4] H. Q. Le, L. X. Nguyen, Y. Qiao, S. T. Kim, E.-N. Huh, and C. S. Hong, "FedDAP: Domain-aware prototype learning for federated learning under domain shift," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[5] L. Wang, J. Bian, L. Zhang, C. Chen, and J. Xu, "Taming cross-domain representation variance in federated prototype learning with heterogeneous data domains," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

[6] Y. Wang, Q. Zhang, X. Li, H. Zhang, Y. Sun, W. Qiu, H. Zhang, Y. Tong, and Z. Zheng, "Taming noise-induced prototype degradation for privacy-preserving personalized federated fine-tuning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

[7] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2016.