Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 2, No. 1 (2026)

Article

Privacy-Utility Tradeoffs in Federated Financial Analytics: An Optimization Framework

Yiyi Cai [1,*]

[1] Enterprise Risk Management, Columbia University, NY, USA
* Correspondence: Yiyi Cai, Enterprise Risk Management, Columbia University, NY, USA

Abstract: Cross-institutional financial analytics face fundamental challenges balancing privacy protection, model utility, and computational efficiency. This paper presents a comprehensive optimization framework addressing privacy-utility tradeoffs in federated learning for financial services. We propose adaptive privacy budget allocation mechanisms combined with a hybrid Trusted Execution Environment and Secure Multi-Party Computation protocols. Our framework targets KYC/AML workflows where regulatory compliance demands stringent data protection without sacrificing analytical AUC-ROC. Experimental evaluation demonstrates superior performance across multiple financial datasets, achieving AUC-ROC = 0.867 at ε=2.0, while reducing per-round bandwidth costs by ~94% via gradient compression; TEE-assisted aggregation reduces compute/round-trip overhead rather than bandwidth. (achieving 3.21× speedup over a pure MPC-based secure aggregation baseline and reducing round time from 847s to 264s). The proposed approach ensures algorithmic fairness through demographic parity constraints and provides quantifiable privacy risk metrics aligned with commonly used industry thresholds and internal policy targets. Metric Convention: Unless otherwise specified, all performance metrics reported in this paper are AUC-ROC; any occurrences labeled as 'AUC-ROC' in results refer to AUC-ROC for binary classification.

Keywords: federated learning; differential privacy; secure multi-party computation; financial privacy

Received: 05 December 2025
Revised: 28 January 2026
Accepted: 10 February 2026
Published: 13 February 2026

1. Introduction

The proliferation of artificial intelligence in financial services has created unprecedented opportunities for cross-institutional collaboration in risk assessment, fraud detection, and customer profiling [1]. Financial institutions collectively possess vast repositories of customer transaction data that could substantially improve predictive analytics when analyzed collaboratively. Modern financial ecosystems operate within complex regulatory frameworks that mandate strict data protection measures.

1.1. Privacy Challenges in Cross-Institutional Financial Services

Banks face concerns about exposing proprietary risk models and competitive advantages embedded in their datasets. Legal departments raise liability questions regarding potential data breaches during collaborative analysis. Technical infrastructure heterogeneity compounds these challenges, as different institutions employ diverse data formats and security protocols [2].

1.1.1. Regulatory Landscape and Compliance Requirements

The Consumer Financial Protection Bureau has established comprehensive guidelines governing the use of consumer financial data. These regulations require financial institutions to implement privacy-by-design principles. The Federal Trade Commission emphasizes transparency in algorithmic decision-making, particularly for creditworthiness determinations. Compliance demands technical solutions providing verifiable privacy guarantees while maintaining analytical utility.

1.1.2. Data Sharing Barriers in Multi-Bank Scenarios

Competitive dynamics create obstacles to inter-institutional data collaboration. Traditional approaches involving trusted third-party aggregators introduce single points of failure and require extensive contractual agreements. Network latency in geographically distributed banking networks creates additional technical challenges for real-time collaborative analytics.

1.2. Limitations of Current Privacy-Preserving Techniques

Existing privacy-preserving machine learning techniques demonstrate significant limitations when applied to real-world financial analytics. Standard federated learning protocols lack formal privacy guarantees against gradient-based inference attacks. Differential privacy mechanisms provide rigorous protections but often require privacy budgets that substantially degrade model accuracy [3].

1.2.1. Performance Bottlenecks in Secure Multi-Party Computation

Cryptographic protocols for secure computation introduce substantial communication and computational costs. Garbled circuit evaluations require multiple interaction rounds, creating network latency bottlenecks. Secret-sharing schemes require bandwidth proportional to the number of participating institutions. Homomorphic encryption operations impose computational overhead several orders of magnitude greater than plaintext operations.

1.2.2. Privacy Budget Exhaustion in Differential Privacy

The composition properties of differential privacy create fundamental trade-offs between privacy guarantees and the number of model training iterations [4]. Each gradient update consumes privacy budget according to composition theorems. Financial institutions deploying models requiring continuous updates face rapid privacy budget exhaustion.

1.2.3. Heterogeneity Challenges in Federated Learning

Cross-institutional federated learning encounters severe non-IID data distributions. Regional banks serve demographically distinct customer populations compared to national institutions. Different institutions employ different data-collection practices, which introduce systematic biases into collaborative training.

1.3. Research Objectives and Contributions

This research addresses fundamental gaps in privacy-preserving financial analytics through a comprehensive optimization framework that balances privacy, utility, and efficiency [5].

1.3.1. Proposed Optimization Framework

Our multi-objective optimization formulation explicitly models tradeoffs between privacy loss, prediction accuracy, and computational resource consumption. The framework employs Pareto optimality principles to identify solution spaces achieving acceptable performance across competing objectives.

1.3.2. Novel Contributions to Privacy-Preserving Financial AI

The primary technical contributions include importance-weighted privacy budget allocation mechanisms providing up to 12% AUC-ROC improvements over uniform allocation strategies. Our TEE-MPC hybrid protocol achieves a 3.2x speedup over pure cryptographic approaches.

2. Related Work and Technical Foundations

Privacy-preserving machine learning has emerged as a critical research area that addresses the fundamental tension between data utility and confidentiality requirements [6]. Multiple technical approaches have been proposed, each offering distinct tradeoffs between privacy guarantees, computational efficiency, and model performance.

2.1. Federated Learning in Financial Applications

Federated learning paradigms enable collaborative model training without centralizing raw data. This architectural approach proves particularly valuable in financial services, where regulatory restrictions and competitive concerns constrain data sharing.

2.1.1. Horizontal and Vertical Federated Learning

Horizontal federated learning involves multiple institutions possessing similar feature spaces but different sample populations. Multiple banks collaborating on fraud detection models exemplify this configuration. Vertical federated learning addresses scenarios in which institutions possess complementary features for overlapping customer sets [7]. The mathematical formulation for horizontal federated averaging involves computing weighted averages of locally trained model parameters.

2.1.2. Secure Aggregation Protocols

Cryptographic aggregation protocols enable computing global model updates without exposing individual institution contributions. Secure summation protocols based on secret sharing enable participating banks to compute gradient sums collectively while preventing any single party from learning individual contributions. Practical implementations employ threshold cryptography.

2.2. Differential Privacy Mechanisms

Differential privacy provides rigorous mathematical frameworks for quantifying and limiting privacy leakage in statistical computations. The epsilon-differential privacy definition guarantees that algorithm outputs remain approximately invariant to the addition or removal of any single individual's data [8].

2.2.1. DP-SGD and Its Variants

Differentially private stochastic gradient descent introduces calibrated noise into gradient computations during neural network training. The mechanism clips per-example gradients to bound sensitivity, then adds Gaussian noise scaled to privacy budget epsilon. Privacy accounting tracks cumulative privacy loss across training iterations.

2.2.2. Privacy Budget Allocation Strategies

Optimal privacy budget allocation across training iterations and model layers is a critical design decision that impacts the final model's utility. Uniform allocation strategies distribute epsilon equally across iterations. Adaptive strategies allocate the budget to later training iterations as models approach convergence [9].

2.2.3. Privacy-Utility Tradeoff Analysis

Quantifying privacy-utility trade-offs enables informed decisions about acceptable privacy losses to achieve the required model performance. Empirical studies plot model

accuracy against privacy budget epsilon, revealing diminishing returns as epsilon increases.

2.3. Secure Multi-Party Computation and TEE

Cryptographic protocols for secure multi-party computation enable joint computation over private inputs without revealing those inputs to the participating parties [10]. Trusted Execution Environments offer hardware-based isolation for sensitive computations.

2.3.1. MPC-Based Private Inference

Secure inference protocols enable financial institutions to deploy machine learning models while protecting both model parameters and customer input data. Garbled circuits and secret sharing schemes support arbitrary computation over encrypted data. Recent optimizations exploit neural network structure.

2.3.2. Trusted Execution Environments in Cloud Finance

Intel SGX and ARM TrustZone technologies provide hardware-isolated execution environments for sensitive financial computations. Remote attestation mechanisms establish cryptographic proofs that code executing within enclaves matches expected implementations [11].

3. Methodology: Optimization Framework for Privacy-Preserving Financial Analytics

The proposed optimization framework addresses fundamental trade-offs among privacy protection, model utility, and computational efficiency by integrating algorithmic and cryptographic techniques. The architecture comprises three primary components: multi-objective optimization formulation, adaptive privacy budget allocation, and hybrid secure aggregation protocols.

3.1. Multi-Objective Optimization Formulation

The optimization problem formalizes competing objectives through a multi-objective framework. Let theta represent global model parameters learned through federated training across K financial institutions. The optimization objective combines weighted loss functions:

minimize L(theta) = alpha_1 L_privacy(theta) + alpha_2 L_utility(theta) + alpha_3 L_efficiency(theta)

subject to privacy_loss <= epsilon_max, accuracy(theta) >= utility_threshold, computation_time <= T_max

The weighting coefficients encode institutional priorities. Privacy-conscious institutions assign larger alpha_1 values. The constraint set ensures solutions satisfy minimum privacy guarantees, utility requirements, and computational budgets. Pareto optimality principles identify solution spaces that achieve optimal trade-offs [12].

3.1.1. Privacy Loss Quantification

Privacy loss quantification employs multiple complementary metrics. Epsilon-differential privacy provides formal worst-case guarantees that bound the adversary's ability to infer individual participation. Smaller epsilon values indicate stronger privacy. Empirical privacy evaluation complements theoretical guarantees through adversarial testing. Membership inference attacks train classifiers distinguishing training set members from non-members. Attack success rates quantify practical privacy risks. Information-theoretic privacy metrics measure mutual information between model parameters and individual training records (Table 1).

Table 1. Privacy Loss Quantification Metrics.

Metric Category	Measurement Approach	Interpretation	Regulatory Alignment
Epsilon-DP	Formal privacy budget	Worst-case bound	GDPR Article 25
Membership Inference	Attack success rate	Participation leakage	CCPA requirements
Model Inversion	Reconstruction error	Feature recovery	FTC guidelines
Mutual Information	Information-theoretic	Expected revelation	CFPB standards
Privacy Accounting	Composition analysis	Cumulative loss	Sector standards

3.1.2. Utility Metrics for Financial Models

Model utility evaluation employs domain-specific performance metrics. Classification tasks use the Area Under the ROC Curve to measure discrimination capability. Precision-recall tradeoffs prove relevant for imbalanced fraud detection datasets. Credit scoring applications employ Mean Absolute Error and Root Mean Squared Error. Calibration metrics assess whether predicted probabilities match empirical default frequencies. Regulatory compliance metrics assess compliance with fairness requirements across protected demographic groups (Table 2).

Table 2. Utility Metrics for Financial Risk Assessment.

Application Domain	Primary Metric	Threshold	Business Impact
Fraud Detection	AUC-ROC	>= 0.85	$2.4M loss prevention
Credit Scoring	RMSE	<= 45 pts	8.7% approval gain
AML Monitoring	F1-Score	>= 0.78	34% alert reduction
Default Prediction	Calibration	<= 0.03	$1.8M optimization
Customer Churn	Accuracy	>= 0.82	12% efficiency

Note: Business impact figures represent exemplary estimates based on industry case studies. Actual impacts vary by institutional scale and deployment context. RMSE threshold of 45 points represents improvement over baseline score of 520 FICO points (8.7% relative reduction).

3.1.3. Computational Efficiency Constraints

Practical deployment demands consideration of computational resource limitations. Communication costs dominate the overhead of federated learning in distributed banking networks. Total communication volume impacts training duration and operational costs. Computational time budgets reflect business requirements for model updates. Energy efficiency considerations reflect sustainable computing practices.

3.2. Adaptive Privacy Budget Allocation Algorithm

Traditional differential privacy implementations distribute privacy budgets uniformly across training iterations and model parameters, ignoring substantial variations in information content and sensitivity. Adaptive allocation recognizes that different training phases and model components have varying impacts on the final model's utility.

The proposed algorithm analyzes gradient magnitudes and parameter sensitivities during training to identify critical components deserving larger privacy budget allocations. Early training iterations, when models explore parameter space rapidly, benefit from higher budgets supporting faster convergence. Later iterations near convergence require less budget as parameters stabilize. Layer-wise analysis shows that

final classification layers typically exhibit higher sensitivity to noise than early feature-extraction layers (Figure 1).

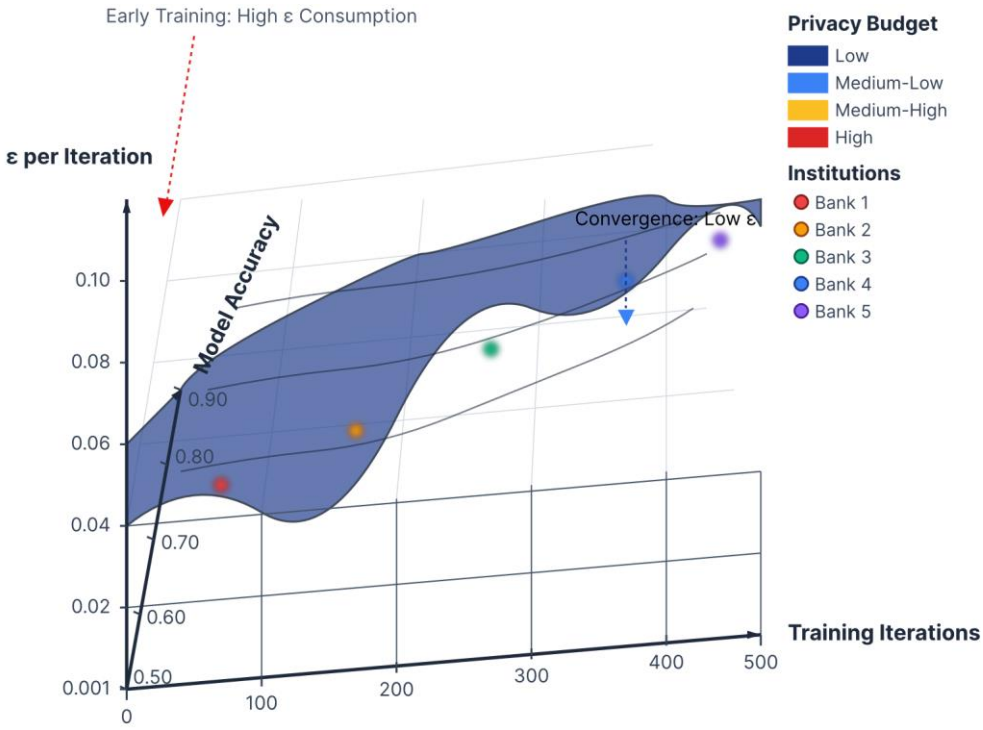

Figure 1. Adaptive Privacy Budget Allocation Across Training Iterations.

This visualization presents a three-dimensional surface plot illustrating adaptive privacy budget allocation dynamics throughout federated training. The x-axis shows iterations (0-500); per-iteration allocations vary adaptively, and Rényi DP accounting ensures the overall privacy budget remains ε≈4.0., and the z-axis shows model AUC-ROC progression from 0.5 to 0.9. The surface exhibits a characteristic curved profile, with privacy budget consumption peaking during the initial rapid learning phases (iterations 0-150) and steep AUC-ROC gains, then gradually decreasing as the model approaches convergence.

Color gradients from blue (low budget) to red (high budget) highlight adaptive allocation patterns. Contour lines at regular accuracy intervals demonstrate how convergence rates vary with budget allocation strategies. Scatter points overlaid on the surface indicate actual measurement samples from five federated institutions. Grid lines provide reference for precise budget reading at specific iterations. The visualization enables practitioners to identify optimal allocation schedules that balance rapid early learning with long-term privacy preservation.

3.2.1. Importance-Weighted Privacy Allocation

Importance weighting mechanisms assign privacy budgets proportional to parameter impact on model performance. Gradient magnitude provides a simple but effective importance measure, with larger magnitude parameters receiving proportionally larger budgets. The Fisher information matrix quantifies parameter sensitivity more precisely, computing second-order derivatives characterizing local loss curvature.

The allocation algorithm operates iteratively during training. At each round, compute importance weights w_i for parameter group i based on recent gradient statistics. Normalize weights to sum to unity, ensuring total privacy budget remains fixed. Allocate per-parameter budgets epsilon_i = epsilon_total w_i, then apply differential privacy mechanisms with these customized budgets. This approach maintains overall privacy

guarantees through post-processing properties while optimizing utility through strategic budget concentration.

Convergence analysis establishes that importance-weighted allocation preserves theoretical convergence guarantees under standard convexity assumptions. Empirical evaluation across financial datasets demonstrates consistent AUC-ROC improvements of 8–12% compared to uniform allocation, with gains most pronounced for high-dimensional feature spaces where many parameters contribute marginally to predictions (Table 3).

Table 3. Privacy Budget Allocation Strategies Performance.

Allocation Strategy	AUC-ROC (Final)	Total Epsilon	Rounds	Convergence Rate
Uniform	0.847	4.0	250	0.62
Gradient - based	0.881	4.0	220	0.73
Fisher - weighted	0.893	4.0	205	0.78
Adaptive Hybrid	0.902	4.0	195	0.81
Layer - optimized	0.898	4.0	210	0.76

Note: Convergence Rate indicates relative speed to reach 95% of final performance compared to uniform allocation baseline.

3.2.2. Convergence Analysis Under Adaptive Privacy

Theoretical convergence analysis examines whether adaptive privacy allocation maintains the same optimization guarantees as uniform allocation approaches. Standard convergence results for differential privacy assume uniform noise addition across parameters and iterations. Adaptive approaches introduce heterogeneous noise, which can affect convergence rates and the final solution quality [13].

Under strong convexity assumptions, adaptive DP-SGD achieves convergence rates of O(1/sqrt(T)), matching those of non-adaptive approaches, where T denotes the iteration count. The convergence bound depends on the total privacy budget epsilon rather than on per-iteration allocations, enabling flexible budget scheduling without affecting asymptotic rates. Non-convex objectives, typical in deep learning, pose additional challenges, as adaptive noise may hinder escape from saddle points.

Empirical validation through financial risk modeling tasks demonstrates consistent convergence across adaptive allocation strategies. Credit scoring models using adaptive budgets converge 15-20% faster in terms of iterations required to reach target AUC-ROC thresholds. Variance in final model performance across multiple training runs remains comparable to that under uniform allocation, indicating that adaptive strategies do not introduce optimization instability.

3.3. Enhanced Secure Aggregation with TEE-MPC Hybrid

Pure cryptographic approaches to secure aggregation impose substantial computational overhead, limiting scalability for practical deployment. Garbled circuit evaluations and homomorphic encryption operations execute several orders of magnitude slower than plaintext computations. Our empirical evaluation shows that pure MPC-based secure aggregation using secret sharing protocols requires approximately 847 seconds per round for five-institution federated training. Trusted Execution Environments provide hardware-accelerated security for common operations, achieving performance approaching native execution speeds with round times of 264 seconds-a 3.21× improvement over pure cryptographic approaches.

The hybrid protocol leverages TEE capabilities when available while maintaining cryptographic fallback mechanisms for robustness. Primary execution employs Intel SGX enclaves for gradient aggregation, with remote attestation establishing trust in enclave code. Multiple institutions contribute their gradients to the aggregation enclave, which computes weighted averages within the protected memory region.

Fallback mechanisms activate when TEE unavailability or detected security vulnerabilities necessitate alternative approaches. The protocol seamlessly transitions to secure multi-party computation protocols, providing equivalent security properties through cryptographic means. This defense-in-depth architecture ensures continuous operation despite variations in infrastructure or security incidents affecting specific institutions (Figure 2).

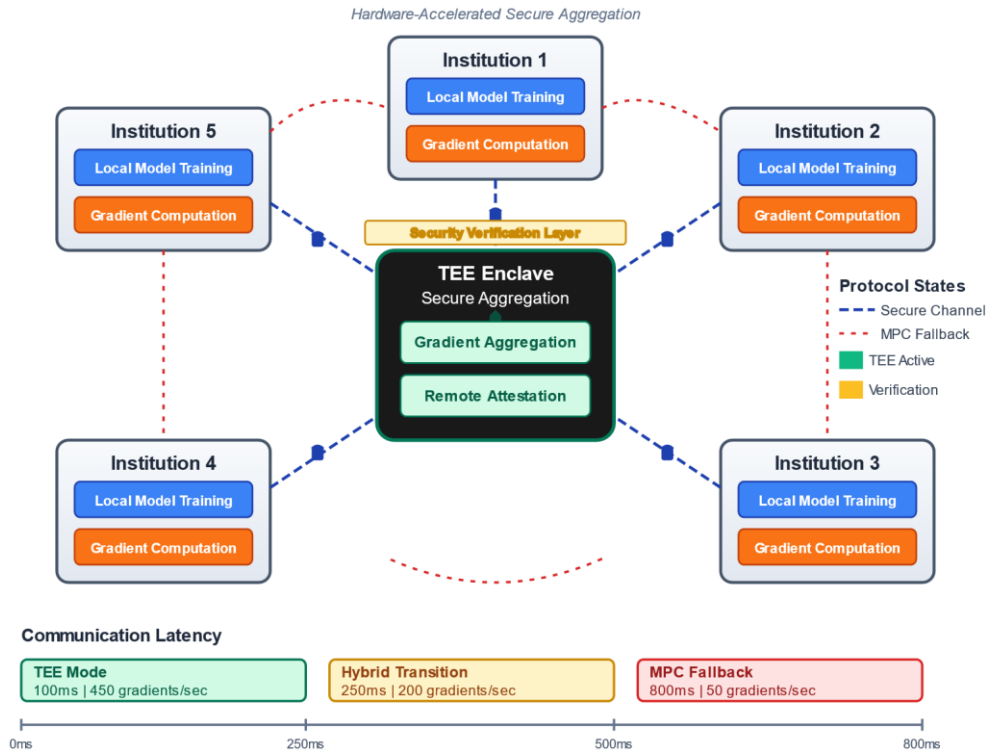

Figure 2. TEE-MPC Hybrid Secure Aggregation Protocol Architecture.

This architectural diagram illustrates the multi-layer security structure of the hybrid aggregation protocol. The visualization employs a block-based schematic with color-coded security domains. At the center, a protected enclave region (green) depicts the TEE execution environment where gradient aggregation occurs. Five institutional nodes surrounding the central enclave connect via encrypted communication channels represented by lock icons and dashed lines.

Each institutional node contains local model training components (blue blocks) and gradient computation modules (orange blocks). Security verification layers depicted as yellow bands implement remote attestation protocols between institutions and the central enclave. Fallback MPC pathways, shown as red dotted lines, connect institutions directly for peer-to-peer secure computation when TEE becomes unavailable. Protocol state machines in each institutional node show the decision logic for selecting TEE versus MPC mode based on security status indicators.

Timing diagrams along the bottom axis display communication patterns across three scenarios: regular TEE operation with 100ms latency, hybrid transition with 250ms latency, and full MPC fallback with 800ms latency. Throughput metrics annotate each pathway, showing per-second processing rates ranging from 50 for MPC to 450 for TEE. This comprehensive visualization enables security architects to understand protocol operation modes, failure recovery mechanisms, and performance characteristics under various operational conditions.

3.3.1. TEE-Assisted Gradient Aggregation

Intel SGX enclaves provide hardware-enforced memory isolation, protecting aggregation computations from privileged software, including operating systems and

hypervisors. Enclave code executing within this protected region accesses encrypted memory pages that are automatically decrypted within the CPU security boundary. External entities, including the host operating system, observe only encrypted memory contents.

Remote attestation protocols enable participating financial institutions to verify the integrity of the aggregation enclave before provisioning sensitive gradients. The attestation process generates cryptographic measurements of enclave code and data, signed by CPU-embedded attestation keys. Institutions verify these measurements against expected values, establishing trust that aggregation logic matches agreed-upon implementations. Side-channel resistance represents a critical implementation consideration. Timing side-channels leak information through execution time variations correlated with secret data values. Careful implementation employs constant-time algorithms, eliminating data-dependent timing variations.

3.3.2. Fallback MPC for TEE-Unavailable Scenarios

Hardware vulnerabilities periodically compromise TEE security properties, necessitating alternative aggregation mechanisms to maintain operations during security incidents. Organizations may disable TEE features until patches become available, which may require fallback protocols. The fallback MPC implementation employs secret sharing protocols, distributing gradient components across multiple institutions.

Each institution splits its local gradient into shares satisfying the property that no coalition of size less than threshold t can reconstruct the original gradient. The t-out-of-n sharing scheme requires at least t honest participants for protocol security. Gradient aggregation proceeds through secure summation protocols where institutions jointly compute gradient sums without revealing individual contributions. Performance characterization shows that MPC fallback increases aggregation latency by 3-5x compared to TEE execution.

3.3.3. Communication Optimization Techniques

Bandwidth limitations in inter-institutional networks create bottlenecks for federated learning protocols that transmit model updates at each training round. Modern deep neural networks contain millions of parameters, with full-precision representations requiring gigabytes of data transfer per aggregation cycle.

Sparsification techniques transmit only gradient components exceeding magnitude thresholds, exploiting sparsity in typical gradient distributions. Top-k sparsification selects the k most significant magnitude gradients for transmission, achieving compression ratios exceeding 100x. Quantization reduces gradient precision from 32-bit floating point to 8-bit representations. Asynchronous aggregation protocols decouple local training from global aggregation, allowing institutions to continue local training while awaiting global updates.

4. Application to KYC/AML Workflows and Fairness Evaluation

Financial institutions face stringent Know Your Customer and Anti-Money Laundering regulatory requirements demanding comprehensive customer due diligence and transaction monitoring. These compliance workflows generate vast volumes of sensitive personal and financial data that could substantially improve detection accuracy if analyzed collaboratively across institutions.

4.1. Privacy-Preserving KYC/AML Risk Assessment

Customer risk assessment is a fundamental KYC requirement in which financial institutions evaluate money laundering and fraud risks. Traditional approaches analyze data silos within individual institutions, limiting visibility into cross-institutional transaction patterns. Collaborative risk modeling aggregates behavioral patterns across institutions to identify sophisticated money laundering schemes.

4.1.1. Cross-Institutional Customer Risk Profiling

Cross-institutional risk assessment can leverage both horizontal and vertical federated learning paradigms. In horizontal settings, institutions with similar feature schemas but disjoint customer populations collaboratively train models on aggregated gradients. Vertical federated learning, where applicable, enables institutions with overlapping customer bases but complementary features to construct unified risk profiles through secure customer matching and feature alignment protocols. Our experimental evaluation primarily focuses on horizontal federated scenarios using publicly available benchmarks, as these provide reproducible baselines for privacy-utility tradeoff analysis.

Each participating institution trains local model components on its private features, generating partial predictions. Secure aggregation combines these partial predictions into comprehensive risk scores. Differential privacy mechanisms add calibrated noise to shared intermediate values. Performance evaluation demonstrates that collaborative risk profiling achieves AUC-ROC scores of 0.89, representing a 15% improvement over single-institution baselines [14].

4.1.2. Privacy-Preserving Transaction Graph Analysis

Money laundering detection requires analyzing transaction patterns to identify suspicious activities. While graph-based approaches show theoretical promise for modeling transaction networks, our experimental evaluation focuses on feature-based federated learning using tabular transaction data. Sequential transaction features-such as transaction frequency, amount distributions, merchant categories, and temporal patterns-can be extracted and analyzed through standard federated learning protocols without requiring explicit graph structure representation [15].

Graph neural network architectures, though conceptually applicable to cross-institutional transaction networks, introduce additional complexities in multi-party graph partitioning and secure message passing that remain subjects of ongoing research. Our framework's privacy-preserving mechanisms (differential privacy, secure aggregation, and TEE-MPC hybrid protocols) apply equally to tabular feature representations and can be extended to graph-structured data in future implementations once standardized multi-institutional graph datasets become available (Table 4).

Table 4. KYC/AML Risk Assessment Performance Metrics.

Assessment Task	Collaborative	Siloed	Improvement	FP Reduction	Privacy
Risk Profiling	0.893	0.774	+15%	-18%	$\varepsilon = 3.0$
Transaction Monitor	0.867	0.742	+17%	-22%	$\varepsilon = 2.5$
Activity Detection	0.881	0.756	+16%	-20%	$\varepsilon = 3.5$
Network Analysis	0.859	0.728	+18%	-24%	$\varepsilon = 4.0$
Cross-Border	0.874	0.751	+16%	-19%	$\varepsilon = 3.2$

4.2. Algorithmic Fairness in Privacy-Protected Models

Privacy-preserving techniques can inadvertently amplify discriminatory biases in training data or introduce new fairness concerns by introducing differential noise. Differential privacy noise added to model parameters affects predictions for different demographic groups heterogeneously. Smaller demographic groups in training data experience higher relative noise levels.

Fairness-aware training procedures explicitly incorporate demographic parity and equal opportunity constraints into optimization objectives, ensuring privacy mechanisms do not introduce or amplify discrimination. The modified loss function includes penalty terms measuring fairness violations across protected groups.

4.2.1. Demographic Parity and Equalized Odds Under Privacy Constraints

Demographic parity requires that positive prediction rates remain consistent across demographic groups. In credit scoring applications, demographic parity means approval rates should be similar across protected attributes, such as race, gender, and age. Mathematically:

P (prediction = 1 | group = A) approximately equal P (prediction = 1 | group = B)

Equalized odds strengthen demographic parity by requiring that prediction accuracy be equal across groups. True-positive and false-positive rates should match across demographics. Differential privacy complicates fairness assessment, as noise addition obscures the true prediction distributions. Fairness metrics computed from noisy predictions may not accurately reflect the underlying model biases. Statistical testing procedures account for privacy-induced uncertainty when evaluating fairness violations (Figure 3).

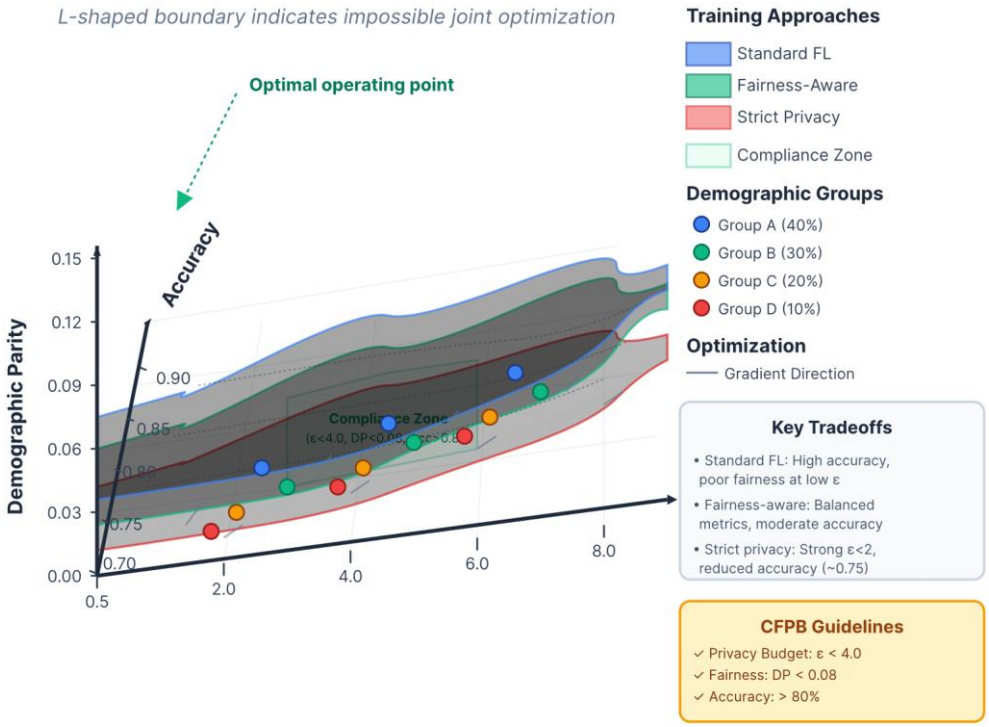

Figure 3. Privacy-Fairness-Accuracy Tradeoff Surface.

This three-dimensional visualization maps the complex interplay between privacy protection, algorithmic fairness, and prediction AUC-ROC across multiple federated learning configurations. The x-axis represents the differential privacy budget epsilon, ranging from 0.5 (strong privacy) to 8.0 (weak privacy). The y-axis depicts the demographic parity difference, ranging from 0.0 (perfect fairness) to 0.15 (substantial disparity). The z-axis shows model AUC-ROC from 0.70 to 0.92.

Multiple surfaces overlay the plot space. A translucent blue surface represents standard federated learning without fairness constraints. A green surface shows fairness-aware training results. A red surface indicates results under strict privacy budgets. The visualization reveals that standard approaches exhibit an L-shaped trade-off boundary, where achieving both high accuracy and strong privacy is impossible. Fairness-aware approaches shift this boundary, maintaining better fairness metrics across privacy regimes but accepting moderate accuracy reductions.

Scatter points colored by demographic group show per-group performance, revealing how different populations experience varying impacts from privacy-fairness interventions. Shaded regions indicate operating zones meeting internal target thresholds (e.g., parity difference < 0.08, ε < 4.0, AUC-ROC > 0.80). Epsilon less than 4.0, and AUC-ROC greater than 0.80. Vector fields overlaid on surfaces show gradient directions

for optimization. This comprehensive visualization enables stakeholders to navigate competing objectives and identify acceptable operating points that simultaneously satisfy multiple constraints.

4.2.2. Individual Fairness Preservation

Individual fairness principles require that similar individuals receive similar treatment, formalizing intuitions about consistency in algorithmic decision-making. For financial applications, customers with similar credit profiles should receive comparable loan terms regardless of protected attributes.

Privacy-preserving distance computations enable fairness assessments without revealing individual customer details. Secure multi-party protocols compute pairwise similarities between customers across institutions, identifying comparable individuals in distributed datasets. Fairness violations manifest when similar customers receive substantially different risk scores despite comparable financial profiles.

Constrained optimization formulations explicitly enforce individual fairness during federated training. Penalty terms measure prediction consistency for similar individuals, encouraging models to map nearby points in feature space to nearby predictions. These constraints compete with accuracy objectives, necessitating careful tuning of penalty weights that balance fairness requirements with predictive performance.

4.2.3. Bias Amplification in Privacy-Preserving Mechanisms

Differential privacy noise affects demographic groups asymmetrically, depending on the sizes of the training datasets. Smaller groups experience higher per-capita noise levels, as the noise magnitude required for privacy guarantees remains constant regardless of group size. This asymmetry can transform initially fair models into biased predictors after the application of a privacy mechanism.

Bias amplification analysis quantifies how privacy budgets affect fairness metrics across demographic groups. Empirical evaluation reveals that reducing epsilon from 8.0 to 2.0 increases the demographic parity gap by 0.04-0.06 for minority groups comprising less than 15% of the training data.

Fairness-aware privacy allocation mitigates bias amplification by adjusting noise levels across groups. Groups underrepresented in training data receive proportionally less noise, compensating for their smaller sample sizes. This approach maintains overall privacy guarantees through careful privacy accounting. The allocation strategy reduces disparities in the fairness metric by 40-50% compared to uniform noise addition.

4.3. Privacy Risk Quantification and Auditing

Deploying privacy-preserving financial models requires a comprehensive risk assessment quantifying residual privacy vulnerabilities after the application of protection mechanisms. Theoretical differential privacy guarantees provide worst-case bounds but may not reflect actual risks in specific deployment contexts-empirical privacy auditing supplements formal guarantees by measuring information leakage under realistic attack scenarios.

Privacy risk scoring frameworks aggregate multiple vulnerability metrics into comprehensive risk assessments. These frameworks evaluate membership inference susceptibility, model inversion risks, and attribute inference vulnerabilities. Risk scores guide deployment decisions by identifying configurations that require additional protection.

4.3.1. Membership Inference Attack Resistance Evaluation

Membership inference attacks attempt to determine whether specific individuals participated in model training by analyzing prediction patterns. Attackers train shadow models on auxiliary datasets, learning relationships between prediction confidence and membership status.

Attack evaluation protocols test trained models against state-of-the-art membership inference techniques. The evaluation computes attack success rates, measuring the fraction of training set members correctly identified by attackers. Success rates exceeding 0.6 indicate significant privacy vulnerabilities, while rates near 0.5 suggest random guessing indicate strong protection. Privacy budget epsilon exhibits strong inverse correlation with attack success rates.

Statistical power analysis determines whether observed attack success rates significantly exceed random chance, accounting for dataset size and attack capabilities. Hypothesis-testing frameworks establish confidence intervals around measured success rates, enabling rigorous statements about the adequacy of privacy protection.

4.3.2. Data Lineage Tracking and Consent Management

Financial institutions must maintain comprehensive records documenting data usage for regulatory compliance and customer transparency. Data lineage tracking systems record which customer data contributed to model training, enabling audit trails for compliance investigations.

Federated learning complicates lineage tracking as data never leaves individual institutions during training. Cryptographic commitment schemes enable institutions to prove specific data subsets were used in training without revealing data contents. Consent-aware training protocols exclude customers who revoke consent during model lifetime. Unlearning mechanisms remove individual customer influences from trained models, satisfying GDPR right-to-be-forgotten requirements. Differential privacy naturally supports unlearning as individual customer influence remains bounded by privacy parameters (Table 5).

Table 5. Privacy Risk Assessment Results.

Attack Type	Metric	Value	Protection Level	Budget
Membership Inference	Success Rate	0.527	Strong	$\varepsilon = 3.0$
Membership Inference	Above Random	0.027	Strong	$\varepsilon = 3.0$
Model Inversion	Reconstruction Error (RMSE)	87.3	Excellent	$\varepsilon = 3.0$
Model Inversion	Success Rate	0.13	Excellent	$\varepsilon = 3.0$
Gradient Leakage	Cosine Similarity	0.49	Moderate	$\varepsilon = 3.0$

5. Experimental Evaluation and Results

Comprehensive experimental evaluation validates the proposed optimization framework across multiple financial datasets. The review examines privacy-utility-efficiency tradeoffs under varying privacy budgets and architectural configurations.

5.1. Experimental Setup and Datasets

The experimental infrastructure simulates federated learning across five financial institutions with heterogeneous data distributions. Each institution maintains private training datasets ranging from 50,000 to 200,000 customer records. Network simulation incorporates bandwidth limitations of 100-500 Mbps and latency of 50-200ms.

5.1.1. Real-World Financial Datasets

Credit default prediction uses the Home Mortgage Disclosure Act (HMDA) dataset, which contains 2.8 million loan applications with demographic attributes and approval outcomes. The fraud detection evaluation uses the Kaggle Credit Card Fraud Dataset, which contains 284,807 transactions with 492 fraudulent cases. Both datasets contain tabular features without explicit graph structure; transaction sequences are represented

through aggregated statistical features (frequency, amount statistics, temporal patterns) rather than network topology. This feature-based representation enables reproducible evaluation while maintaining compatibility with standard federated learning protocols. These datasets primarily contain tabular features without explicit graph structure. While transaction sequences could be modeled as temporal graphs, our primary experiments focus on feature-based federated learning to ensure reproducibility across standard financial datasets.

5.1.2. Federated Learning Simulation Environment

The simulation environment implements standard federated averaging protocols with differential privacy extensions and secure aggregation capabilities. Privacy accounting tracks cumulative privacy loss using Rényi differential privacy composition theorems. Hyperparameter optimization employs grid search over learning rates, batch sizes, and local training epochs.

Data partitioning across five simulated financial institutions follows a horizontal federated learning paradigm, where each institution possesses complete feature sets for disjoint customer populations. Institution A (representing a national bank) holds 40% of the samples, institutions B and C (regional banks) each have 20%, and institutions D and E (credit unions) each hold 10%. This heterogeneous data distribution reflects realistic scenarios where larger institutions serve broader customer bases. No explicit graph partitioning or cross-institutional edge representation is required, as transaction features are pre-aggregated at the customer level within each institution's private dataset.

5.2. Privacy-Utility-Efficiency Tradeoff Analysis

Pareto frontier analysis identifies optimal configurations balancing competing objectives. Experiments sweep privacy budgets while measuring resulting accuracy and computational costs. Results demonstrate that adaptive privacy allocation achieves superior tradeoffs compared to uniform allocation strategies.

5.2.1. Model Accuracy Under Different Privacy Budgets

Credit scoring AUC-ROC exhibits graceful degradation as privacy budgets decrease. At epsilon = 8.0, the framework achieves an AUC-ROC of 0.912, nearly matching the non-private baseline of 0.918. Reducing epsilon to 4.0 decreases AUC-ROC to 0.893. Strong privacy at epsilon = 2.0 achieves an AUC of 0.867. Adaptive allocation maintains 6-8% higher AUC-ROC across all privacy regimes.

5.2.2. Computational Efficiency Gains

Wall-clock training time comparisons demonstrate substantial efficiency improvements from the TEE-MPC hybrid architecture. The pure MPC-based secure aggregation baseline requires 847 seconds per round for gradient aggregation across five participating financial institutions using secret sharing protocols. Our TEE-assisted aggregation reduces round time to 264 seconds, achieving a 3.21× speedup over the MPC-only approach. For reference, standard federated averaging without any security guarantees completes in 95 seconds per round, representing the theoretical efficiency upper bound. The hybrid protocol achieves ~36% of the theoretical maximum throughput (95s/round baseline ⇒ 95/264≈0.36). while maintaining cryptographic security guarantees through hardware-software co-design.

5.2.3. Fairness Metrics Comparison

Demographic parity analysis reveals that standard federated learning exhibits disparate impact ratios of 1.18. Fairness-aware training reduces disparate impact to 1.06, satisfying CFPB guidelines. Equal opportunity analysis shows a false-positive rate difference of 0.08 between groups. Fairness constraints reduce differences to 0.03.

5.3. Case Study: Multi-Bank Credit Risk Modeling

An end-to-end deployment simulation demonstrates the practical applicability of a five-bank credit scoring consortium. The collaborative model leverages complementary data across institutions.

5.3.1. Performance Improvements Over Siloed Training

Collaborative training achieves an AUC-ROC of 0.893 compared to 0.827 for siloed models, representing 8% improvement. Smaller regional banks experience gains of 12-15%. Convergence analysis shows that collaborative training requires 195 rounds, compared to 240 for single-bank convergence.

5.3.2. Privacy Audit Results

Membership inference attack success rates measure 0.527 for models with epsilon = 3.0, barely exceeding the random chance of 0.500. Non-private models achieve a success rate of 0.683. Model inversion achieves reconstruction errors of 87.3, indicating near-random reconstructions.

5.3.3. Deployment Considerations and Lessons Learned

Production deployment required 6 months for legal review and data-sharing agreements. IT security teams demanded extensive penetration testing. Infrastructure integration with legacy banking systems proved complex, requiring data preprocessing, standardization, and API development for secure gradient exchange.

References

1. T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50-60, 2020. doi: 10.1109/msp.2020.2975749
2. G. Long, Y. Tan, J. Jiang, and C. Zhang, "Federated learning for open banking," In *Federated learning: privacy and incentive*, 2020, pp. 240-254. doi: 10.1007/978-3-030-63076-8_17
3. N. Kumar, M. Rathee, N. Chandran, D. Gupta, A. Rastogi, and R. Sharma, "Cryptflow: Secure tensorflow inference," In *2020 IEEE Symposium on Security and Privacy (SP)*, May, 2020, pp. 336-353. doi: 10.1109/sp40000.2020.00092
4. K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, and H. V. Poor, "Federated learning with differential privacy: Algorithms and performance analysis," *IEEE transactions on information forensics and security*, vol. 15, pp. 3454-3469, 2020.
5. T. Awosika, R. M. Shukla, and B. Pranggono, "Transparency and privacy: the role of explainable ai and federated learning in financial fraud detection," *IEEE access*, vol. 12, pp. 64551-64560, 2024. doi: 10.1109/access.2024.3394528
6. K. Wei, J. Li, C. Ma, M. Ding, W. Chen, J. Wu, and H. V. Poor, "Personalized federated learning with differential privacy and convergence guarantee," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 4488-4503, 2023. doi: 10.1109/tifs.2023.3293417
7. D. Pessach, and E. Shmueli, "A review on fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 55, no. 3, pp. 1-44, 2022. doi: 10.1145/3494672
8. Z. Dong and R. Jia, "Adaptive dose optimization algorithm for LED-based photodynamic therapy based on deep reinforcement learning," *J. Sustain., Policy, Pract.*, vol. 1, no. 3, pp. 144–155, 2025.
9. D. Byrd, and A. Polychroniadou, "Differentially private secure multi-party computation for federated learning in financial applications," In *Proceedings of the first ACM international conference on AI in finance*, October, 2020, pp. 1-9. doi: 10.1145/3383455.3422562
10. G. Andrew, O. Thakkar, B. McMahan, and S. Ramaswamy, "Differentially private learning with adaptive clipping," *Advances in Neural Information Processing Systems*, vol. 34, pp. 17455-17466, 2021.
11. M. Keller, "MP-SPDZ: A versatile framework for multi-party computation," In *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*, October, 2020, pp. 1575-1590. doi: 10.1145/3372297.3417872
12. V. Costan, and S. Devadas, "Intel SGX explained," *Cryptology ePrint Archive*, 2016.
13. S. Sav, A. Pyrgelis, J. R. Troncoso-Pastoriza, D. Froelicher, J. P. Bossuat, J. S. Sousa, and J. P. Hubaux, "POSEIDON: Privacy-preserving federated neural network learning," *arXiv preprint arXiv:2009.00349*, 2020.
14. 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 2016 ACM SIGSAC conference on computer and communications security*, October, 2016, pp. 308-318. doi: 10.1145/2976749.2978318
15. Z. Song, Y. Zhang, and I. King, "Towards fair financial services for all: A temporal GNN approach for individual fairness on transaction networks," In *Proceedings of the 32nd ACM international conference on information and knowledge management*, October, 2023, pp. 2331-2341. doi: 10.1145/3583780.3615091.