International Journal of Artificial Intelligence (AI) in Business and Management
Volume 2 • Issue 1 • January-December 2026

Unified Federated AI Framework for Credit Scoring:
For Privacy, Fairness, and Scalability

Arjun Kaarat
https://orcid.org/0009-0001-0786-5723
Walmart Inc., USA

Venkata Jaipal Reddy Batthula
https://orcid.org/0009-0003-3957-3965
Southwestern College, USA

Richard S. Segall
https://orcid.org/0000-0001-7627-2609
Arkansas State University, USA
Received: October 20th, 2025 | **Accepted:** January 5th, 2026

ABSTRACT

This research introduces a unified federated AI framework that jointly achieves privacy, fairness, and accuracy in credit scoring. The system enables collaboration across 300 institutions without sharing raw data, combining three components: Dirichlet Process Gaussian Mixture Model (DP-GMM)-based differential privacy, Exponentiated Gradient fairness constraints, and 8-bit model quantization. Evaluated from small-scale (German Credit, n = 1,000) to production-scale (LendingClub, n = 500,000), the framework achieves 96.94% accuracy with a 0.069% demographic parity gap, meeting stringent regulatory thresholds. Parameter sweeps quantify privacy–fairness trade-offs and are supported by theoretical guarantees, providing actionable guidance for deployment. Key contributions include the production-ready integration of privacy, fairness, and federated learning with formal guarantees; multi-dataset scalability across heterogeneous institutions; an 8× reduction in communication via quantization; and a compliance blueprint addressing the General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and fair lending standards. Results position federated AI as a practical and responsible solution for credit decisioning, delivering collaborative intelligence with strong privacy protection, algorithmic fairness, and regulatory compliance.

KEYWORDS

Algorithmic Fairness, Credit Scoring, Demographic Parity, Differential Privacy, Federated AI

INTRODUCTION

Credit scoring algorithms face a fundamental tension between privacy protection, algorithmic fairness, and predictive accuracy in regulated financial services. Centralized AI models improve performance but increase breach risk and can amplify historical biases, while overly restrictive privacy or fairness interventions often reduce utility. This work develops a federated learning framework that allows financial institutions to collaboratively train credit models without sharing raw

DOI: 10.4018/IJAIBM.398859

customer data, while enforcing fairness constraints and maintaining competitive accuracy. Across five credit datasets and 300 simulated institutions, the framework achieves 96.94% accuracy, a 0.069% demographic parity gap (93.1% safety margin below regulatory thresholds), $\varepsilon = 1.0$ differential privacy, and 8× communication reduction, positioning it as a practical option for GDPR, CCPA, and fair-lending-compliant deployment.

AI-powered credit scoring faces a core challenge: achieving privacy protection, algorithmic fairness, and predictive accuracy simultaneously, creating competing regulatory and business pressures. Centralized AI systems expose financial data to breach risks—highlighted by the 2017 Equifax incident affecting 147 million consumers—while perpetuating bias against protected groups via discriminatory training data (Agarwal et al., 2018; Blattner & Nelson, 2021). Regulatory scrutiny has intensified: the CFPB states algorithmic complexity does not exempt institutions from fair lending laws and issued updated guidance in 2025 targeting discriminatory AI outcomes (CFPB, 2023; CFPB, 2024). The EU Court of Justice ruled automated AI credit scoring can violate GDPR Article 22 without adequate safeguards (CJEU, 2024). With Gartner projecting worldwide AI spending will reach $1.5T in 2025, scalable frameworks addressing privacy and fairness are urgently needed.

This research addresses three core challenges in responsible credit scoring. First, privacy preservation requires formal ε-differential privacy guarantees while maintaining predictive performance across heterogeneous institutional data (Dwork et al., 2006; McMahan et al., 2017). Second, fairness enforcement demands algorithmic constraints on demographic parity without sacrificing model accuracy (Agarwal et al., 2018). Third, scalable communication efficiency requires protocols that support hundreds of institutions with diverse data characteristics. Existing approaches address these objectives independently or in pairs, creating deployment barriers: privacy-focused methods degrade accuracy; fairness-constrained approaches lack privacy guarantees; communication-efficient techniques do not jointly optimize privacy and fairness.

To overcome these barriers, this research introduces a unified federated AI framework that achieves all three objectives. Primary contributions include: (1) a unified framework integrating differential privacy via DP-GMM synthetic augmentation, fairness constraints via Exponentiated Gradient optimization, and federated learning with formal convergence guarantees; (2) empirical validation scaling from proof-of-concept (German Credit, n=1,000) to enterprise deployment (LendingClub, n=500,000) across 300 simulated institutions; (3) theoretical foundations establishing privacy and convergence proofs under simultaneous fairness constraints on non-IID distributed data; and (4) a production-ready system achieving sub-10-minute training cycles on standard hardware (8-core CPU, 55GB RAM) with quantifiable regulatory compliance (GDPR, CCPA, fair lending standards).

The framework shows that privacy, fairness, and competitive performance can be achieved simultaneously through distributed architectures without requiring fundamental trade-offs among these objectives. Evaluation across five datasets yields 96.94% accuracy, a 0.069% demographic parity gap (a 93.1% safety margin below regulatory thresholds), while maintaining $\varepsilon=1.0$ differential privacy and achieving 8× communication efficiency through model quantization. This work positions federated AI as a practical paradigm for credit decisioning in regulated financial services.

Prior work addresses individual dimensions of the privacy–fairness–accuracy–efficiency challenge, but not all simultaneously. McMahan et al. (2017) introduced the FedAvg algorithm for decentralized training but provides no privacy or fairness guarantees. Kairouz et al. (2021) survey federated learning approaches that incorporate privacy via differential privacy on gradients (DP-SGD), but acknowledge DP-SGD degrades utility and does not address fairness. Calmon et al. (2019) propose fairness constraints for centralized machine learning, but these approaches require raw data centralization and do not address privacy or communication efficiency in federated settings. Table 1 compares these representative approaches to our work across twelve critical dimensions.

These technical achievements have economic relevance for financial institutions. For a mid-sized lender ($500M annual originations, 50K applications), potential value drivers include: (1) enhanced