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

The Convergence of Big Data and AI Through Learning-Based Methods for Business Intelligence

Prasanna Rajbhandari
https://orcid.org/0009-0009-3394-2137
Arkansas State University, USA

Richard S. Segall
https://orcid.org/0000-0001-7627-2609
Arkansas State University, USA
Received: October 20th, 2025 | **Accepted:** January 19th, 2026

ABSTRACT

The combination of big data and artificial intelligence (AI) is redefining how organizations learn from information by enabling systems that autonomously discover patterns, adapt to change, and generate valuable insights. This article presents overview of three AI applications areas for big data and then discusses one of these in depth using three potential transformative AI methods for learning-based methods: (1) federated learning for privacy-preserving customer behavior analysis, (2) self-supervised learning for detecting anomalies and fraud without labeled data, and (3) contrastive learning for creating robust representations that enhance personalization and recommendations. Together, these methods show how advanced learning paradigms extract actionable intelligence from distributed, unlabeled, and complex data while meeting ethical and regulatory standards.

KEYWORDS

Artificial Intelligence, Big Data, Digital Twins, Federated Learning, Graph Neural Networks, Self-Supervised Learning, Swarm Intelligence

1. INTRODUCTION

Customer analytics increasingly relies on data distributed across organizational, geographic, and regulatory boundaries. Centralized machine learning approaches require aggregating sensitive customer data into a single repository, raising significant concerns related to privacy leakage, regulatory non-compliance, and governance transparency. These challenges are particularly pronounced in domains such as finance, healthcare, and digital commerce, where customer data is both high-value and tightly regulated.

Federated learning (FL) offers a decentralized alternative by enabling collaborative model training across multiple clients while keeping raw data localized. Despite its promise, practical deployment of federated learning in enterprise environments remains limited by three persistent challenges: (i) the lack of auditable training processes suitable for regulatory oversight, (ii) degradation in model utility under differential privacy constraints, and (iii) communication inefficiencies that hinder scalability.

This paper addresses these challenges by focusing on federated learning as the primary learning paradigm and proposing an auditable, privacy-aware training framework for customer analytics.

DOI: 10.4018/IJAIBM.400274

While self-supervised and contrastive learning techniques are discussed, they are incorporated solely as supporting mechanisms to improve federated convergence and robustness under limited labeled data. The scope of this work is intentionally narrowed to provide empirical depth and quantitative validation of federated learning performance under real-world constraints.

1.1 Research Problem Definition and Objectives

Although extensive literature explores AI-driven business intelligence, most current studies address isolated capabilities such as predictive accuracy or cloud-scale computation without establishing a unified, learning-based framework that harmonizes privacy preservation, label efficiency, and representation optimization.

Accordingly, this work defines three interlinked research problems that collectively form the foundation of *learning-based AI–Big Data convergence for business and management*:

P1: An Auditable Federated Learning Framework for Privacy-Preserving Customer Analytics: Empirical Evaluation on Transactional Data

Empirical motivation: Modern enterprises (banks, retailers, healthcare systems) hold fragmented, privacy-sensitive customer data that cannot be centralized because of the EU AI Act (2025) and California ADMT regulations.

Scholarly gap: Classical *Federated Learning* (McMahan et al., 2017) enables distributed training but lacks verifiable governance; privacy budgets are often untracked, deletion requests unverifiable, and bandwidth overheads impractical.

Research question:

Research Question 1: How can federated models in business analytics achieve *auditable*, *differentially private*, and *bandwidth-efficient* training while maintaining competitive utility compared with centralized baselines?

P2: Self-Supervised Learning for Fraud and Anomaly Detection (FRAUD-SSL) - Label Efficiency and Drift Adaptation

Empirical motivation: In domains such as e-commerce or finance, less than 1% of transactions are labeled as fraud, while new patterns appear daily. Manual labeling is infeasible, and drift quickly degrades supervised models.

Scholarly gap: Existing fraud-detection methods (Nguyen et al., 2021; Chen et al., 2022) rely on supervised or static graph learning, ignoring temporal evolution, privacy, and explainability under scarce labels.

Research question:

Research Question 2: How can a self-supervised architecture integrate *sequence-aware* and *temporal-graph* objectives to maintain high precision under drift and limited labels, while preserving data privacy and analyst interpretability?

P3: Contrastive Representation Optimization (CRO) - Transferable and Efficient Embeddings

Empirical motivation: Business intelligence systems increasingly reuse pretrained embeddings for recommendation, personalization, and multimodal analytics, yet these embeddings often fail to generalize or calibrate across domains.

Scholarly gap: Contrastive learning frameworks (SimCLR, MoCo, CLIP) emphasize instance discrimination but ignore *representation geometry optimization* the balance between alignment, uniformity, calibration, and compute efficiency.

Research question: