World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

FEDERATED AI AGENTS FOR COLLABORATIVE DECISION ANALYTICS ACROSS RETAIL, BANKING, AND NETWORK INFRASTRUCTURES

Abhignan Srivatsava Sribhashyam [1,*], Bala Yashwanth Reddy Thumma [2], Nivedan Suresh [3] and Supraja Ayyamgari [4]

[1] Senior Software Engineer, Target Corporation, Lakeville, MN, USA.
[2] Verizon, Ashburn, VA, USA.
[3] Staff Software Engineer, Visa Inc., Austin, TX, USA,
[4] Independent Researcher Scholar, Ashburn, VA 20147,USA.
* Corresponding Author

ORCID Details
Abhignan Srivatsava Sribhashyam: https://orcid.org/0009-0006-1197-2982
Bala Yashwanth Reddy Thumma: https://orcid.org/0009-0009-3928-6089
Nivedan Suresh: https://orcid.org/0009-0001-8186-8690
Supraja Ayyamgari: https://orcid.org/0009-0005-0447-4363

World Journal of Advanced Research and Reviews, 2023, 17(02), 988–998

Publication history: Received on 19 January 2023; revised on 22 February 2023; accepted on 27 February 2023

Article DOI: https://doi.org/10.30574/wjarr.2023.17.2.0249

ABSTRACT

The practical deployment of distributed intelligent systems across heterogeneous organizational environments faces significant challenges due to data privacy constraints, domain heterogeneity, and the limited reasoning capability of conventional federated learning frameworks. To address these challenges, this paper proposes a Federated Agentic Intelligence Framework (FAIF) that integrates Federated Learning (FL), Large Language Models (LLMs), Multi-Agent Systems (MAS), Knowledge Graphs (KGs), Retrieval-Augmented Generation (RAG), and Explainable Artificial Intelligence (XAI) to enable privacy-preserving and autonomous collaborative analytics. First, FAIF employs a federated learning architecture to support distributed model training without exposing sensitive organizational data, thereby preserving data confidentiality while enabling collaborative intelligence. Second, domain-specific autonomous agents are designed to perform specialized analytical tasks, including supply chain coordination and inventory synchronization in retail, fraud detection and risk assessment in finance, and autonomous network management, anomaly detection, and traffic optimization in networking environments. Third, retrieval-augmented generation combined with knowledge graphs enhances contextual reasoning, knowledge integration, and explainability, while explainable AI modules improve the transparency and trustworthiness of agent decisions. Finally, a coordinated multi-agent collaboration mechanism enables adaptive decision-making, efficient knowledge sharing, and scalable cross-domain intelligence. The proposed FAIF provides a robust, scalable, and privacy-preserving framework for trustworthy distributed analytics, demonstrating strong adaptability to heterogeneous operational ecosystems and offering an effective foundation for next-generation autonomous enterprise intelligence systems.

Keywords: Federated Learning; Multi-Agent Systems; Large Language Models; Retrieval-Augmented Generation; Knowledge Graphs; Explainable Artificial Intelligence; Privacy-Preserving Analytics; Distributed Intelligent Systems.

1 INTRODUCTION

The rapid growth of distributed intelligent systems across enterprises has significantly increased the demand for collaborative artificial intelligence while preserving the privacy of sensitive organizational data [1]. Conventional centralized machine learning requires collecting data from multiple organizations into a central repository, introducing serious concerns regarding data privacy, regulatory compliance, security, and ownership [2]. Although recent advances in Large Language Models (LLMs) have greatly improved autonomous reasoning and decision-making capabilities, their direct deployment in distributed environments remains challenging because organizations are often unwilling or legally unable to share proprietary datasets [3]. Federated Learning (FL) has emerged as an effective distributed learning paradigm that enables multiple organizations to collaboratively train machine learning models without exchanging raw data. Each participating client performs local model training and shares only model parameters or gradients with a central aggregation server, thereby preserving data privacy while enabling collaborative intelligence [4]. However, conventional federated learning primarily focuses on distributed optimization and lacks autonomous reasoning, domain-specific decision-making, contextual knowledge integration, and explainability [5]. Furthermore, heterogeneous data distributions, varying organizational objectives, and limited semantic understanding considerably reduce the effectiveness of traditional federated learning in real-world enterprise applications.

Recent advances in Multi-Agent Systems (MAS), Retrieval-Augmented Generation (RAG), Knowledge Graphs (KGs), and Explainable Artificial Intelligence (XAI) provide promising opportunities for overcoming these limitations [6]. Intelligent agents can autonomously coordinate complex tasks, interact with external knowledge repositories, and collaboratively solve domain-specific problems. Knowledge graphs provide structured semantic relationships that improve contextual understanding, while retrieval-augmented generation enhances the factual accuracy of LLM-generated responses. Explainable AI further increases transparency by enabling users to understand the reasoning process behind autonomous decisions, thereby improving trustworthiness in mission-critical applications [7]. Despite these advances, existing intelligent enterprise frameworks often integrate only a subset of these technologies and rarely provide a unified architecture capable of simultaneously supporting privacy-preserving distributed learning, autonomous multi-agent collaboration, contextual reasoning, and explainable decision-making across heterogeneous domains [8]. In particular, practical applications spanning retail, finance, and networking require coordinated intelligence while ensuring data confidentiality, regulatory compliance, scalability, and trustworthy autonomous operations.

To address these challenges, this paper proposes a Federated Agentic Intelligence Framework (FAIF) that integrates Federated Learning, Large Language Models, Multi-Agent Systems, Knowledge Graphs, Retrieval-Augmented Generation, and Explainable Artificial Intelligence into a unified privacy-preserving intelligent architecture [9]. The framework enables domain-specific autonomous agents to collaboratively perform supply chain coordination and inventory synchronization in retail, fraud detection and risk assessment in finance, and autonomous network management, anomaly detection, and traffic optimization in networking environments [10]. Federated learning preserves organizational privacy by enabling collaborative model training without exposing sensitive data, while retrieval-augmented generation and knowledge graphs improve contextual reasoning and semantic understanding [11]. Furthermore, explainable AI provides transparent decision interpretation, allowing organizations to verify and trust autonomous recommendations.

The major contributions of this paper are summarized as follows.

- A novel Federated Agentic Intelligence Framework (FAIF) is proposed by integrating Federated Learning, Large Language Models, Multi-Agent Systems, Retrieval-Augmented Generation, Knowledge Graphs, and Explainable Artificial Intelligence into a unified architecture for privacy-preserving collaborative intelligence.
- A domain-specific multi-agent collaboration mechanism is designed to enable autonomous decision-making across heterogeneous enterprise environments, supporting retail supply chain management, financial fraud detection and risk assessment, and intelligent network operation and optimization.
- A knowledge-enhanced reasoning framework is developed by combining Retrieval-Augmented Generation with Knowledge Graphs to improve contextual understanding, factual consistency, semantic reasoning, and explainability of LLM-generated responses.
- A scalable privacy-preserving collaborative intelligence strategy is introduced by leveraging federated learning and explainable AI, enabling trustworthy cross-organizational analytics while preserving data confidentiality, supporting regulatory compliance, and improving the transparency of autonomous decision-making.

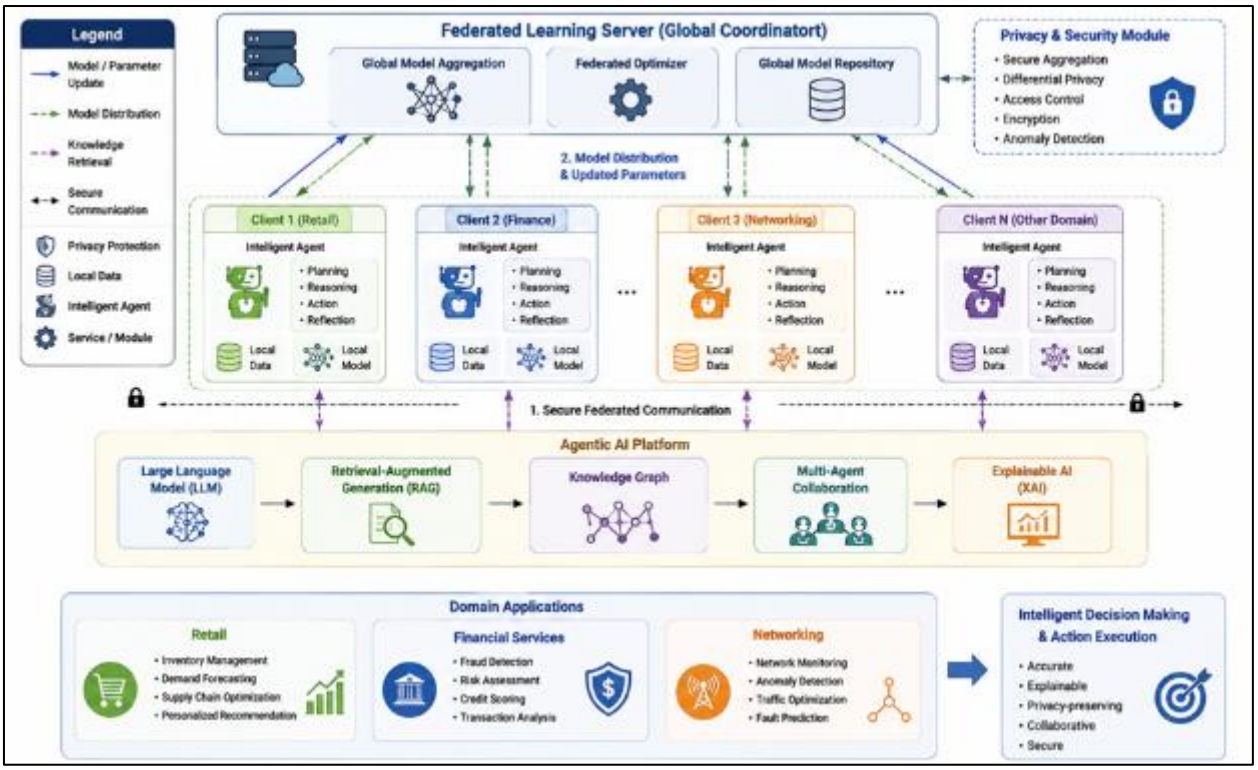

Figure 1: Overall Architecture of the Federated Agentic Intelligence Framework (FAIF)

2 RELATED WORK

Recent advances in distributed artificial intelligence have accelerated research on privacy-preserving collaborative learning and autonomous intelligent systems. Federated Learning (FL) has emerged as a fundamental paradigm for enabling multiple organizations to collaboratively train machine learning models without exchanging raw data. Numerous studies have focused on improving model convergence, communication efficiency, and robustness in heterogeneous distributed environments [12]. Adaptive aggregation methods have been proposed to optimize global model performance by dynamically adjusting client contributions based on model quality, statistical heterogeneity, or optimization objectives [13]. Although these approaches improve learning efficiency, they primarily concentrate on model optimization and often overlook autonomous reasoning, contextual knowledge integration, and explainable decision-making required in complex enterprise applications [14]. Large Language Models (LLMs) have recently demonstrated remarkable capabilities in natural language understanding, reasoning, and intelligent decision support [15]. General-purpose LLMs such as GPT-4, LLaMA, PaLM, and Claude have significantly advanced intelligent analytics across multiple domains. However, directly applying standalone LLMs in enterprise environments remains challenging because their internal knowledge is static, domain-specific information rapidly evolves, and generated responses may suffer from hallucinations or lack explainability [16]. Moreover, enterprise organizations often maintain proprietary knowledge repositories that cannot be directly incorporated into centralized model training due to privacy and regulatory constraints.

Retrieval-Augmented Generation (RAG) has emerged as an effective solution for enhancing LLM performance by integrating external knowledge during inference. Existing RAG frameworks retrieve semantically relevant information from domain-specific knowledge bases before response generation, substantially improving factual consistency and reducing hallucinations [17]. Furthermore, Knowledge Graphs (KGs) have been widely adopted to represent structured semantic relationships among entities, enabling contextual reasoning and interpretable knowledge discovery. While combining RAG and knowledge graphs improves semantic understanding, most existing solutions operate independently and rarely support collaborative multi-organization intelligence or privacy-preserving distributed analytics [18]. Multi-Agent Systems (MAS) provide another important direction for autonomous intelligent decision-making. Recent research has demonstrated that specialized autonomous agents can cooperatively perform task decomposition, planning, negotiation, and decision execution across complex environments. Agent-based frameworks have been successfully applied to supply chain optimization, financial risk assessment, cybersecurity, and network management [19]. Nevertheless, most existing multi-agent architectures assume centralized knowledge access and lack mechanisms for privacy-preserving collaboration among independent organizations with heterogeneous datasets.

Security and privacy remain major challenges in distributed intelligent systems. Federated learning preserves raw data confidentiality by keeping data localized; however, model updates may still leak sensitive information through gradient leakage, membership inference, and model inversion attacks [20]. Existing defense mechanisms, including differential privacy, secure multi-party computation, and homomorphic encryption, improve privacy protection but often introduce considerable computational overhead, communication latency, and scalability limitations [21]. In addition, these techniques generally focus on secure model training and provide limited support for autonomous reasoning, explainability, and knowledge-driven decision making. Although significant progress has been achieved in federated learning, large language models, retrieval-augmented generation, knowledge graphs, multi-agent systems, and explainable artificial intelligence, existing studies typically investigate these technologies independently or integrate only a limited subset of them [22, 23]. Few frameworks provide a unified architecture capable of simultaneously supporting privacy-preserving distributed learning, autonomous multi-agent collaboration, contextual knowledge reasoning, explainable AI, and cross-domain enterprise analytics. To address these limitations, this paper proposes the Federated Agentic Intelligence Framework (FAIF), which seamlessly integrates Federated Learning, Large Language Models, Retrieval-Augmented Generation, Knowledge Graphs, Multi-Agent Systems, and Explainable AI into a unified architecture. The proposed framework enables trustworthy, scalable, and privacy-preserving autonomous intelligence across heterogeneous enterprise environments while supporting domain-specific applications in retail, finance, and networking.

3 FRAMEWORK IMPLEMENTATION

3.1 Framework Overview

The **Federated Agentic Intelligence Framework (FAIF)** is illustrated in Figure 1. Its workflow consists of four major stages.

- Initial Training Stage: At the beginning of training, all participating clients collaboratively train a shared global model.
- Dynamic Clustering Stage: When the differences among client model updates exceed a predefined threshold while the average update magnitude becomes stable, the HDBSCAN algorithm is employed to dynamically cluster clients according to the similarity of their model updates. This clustering process does not require the number of clusters to be specified beforehand. After clustering, noise points are reassigned to the most appropriate cluster using a similarity-based reassignment strategy.
- Reputation Evaluation Stage: After clustering, each cluster forms an approximately homogeneous data environment. The Federated Agentic Intelligence Framework (FAIF) evaluates the quality of each client's model update from two complementary perspectives:
    - Similarity between the client's model update and the cluster centroid.
    - Consistency between the client's training loss and the average loss within the cluster.
    - These two indicators jointly measure the reliability of client updates.
- Reputation-Based Aggregation Stage: Client updates are aggregated using reputation scores as weighting factors. Combined with an adaptive dynamic threshold adjustment strategy, this mechanism accurately evaluates client behavior. The weighted aggregation strategy preserves cluster personalization while effectively reducing the influence of malicious clients on the federated learning system.

3.2 Dynamic Clustering Algorithm

The dynamic clustering algorithm consists of three major components:

- Cluster triggering mechanism
- HDBSCAN-based adaptive clustering
- Similarity-based noise point reassignment

During the early training stage, client models exhibit different convergence rates. Performing clustering too early may produce inaccurate groupings. Therefore, the framework introduces an update stability-based clustering trigger. In each communication round, the average norm (mean_norm) and maximum norm (max_norm) of all client model updates are computed. Clustering is activated only when the following three conditions are simultaneously satisfied:

- The average update norm is sufficiently small, indicating that most clients have nearly converged.
- The maximum update norm remains relatively large, implying that significant differences still exist among client updates.
- The current training round exceeds a predefined warm-up period.

These conditions ensure that clustering begins only after local models have become relatively stable while sufficient diversity still exists among clients. Once activated, the Federated Agentic Intelligence Framework (FAIF) employs the HDBSCAN density-based hierarchical clustering algorithm. HDBSCAN automatically discovers clusters of varying densities without requiring the number of clusters to be predefined. It is capable of identifying arbitrarily shaped clusters while remaining robust to noisy data. The similarity between two clients is measured using the cosine similarity of their model update vectors. Clients exhibiting similar optimization directions are grouped into the same cluster. After each communication round, the cosine similarity matrix among all clients is computed. During clustering, HDBSCAN labels clients located in low-density regions as noise points. These noise points may represent either benign clients with highly heterogeneous data distributions or malicious clients attempting to manipulate the learning process. To address this issue, the Federated Agentic Intelligence Framework (FAIF) introduces a similarity-based noise reassignment strategy. For each noise client, its similarity to every cluster centroid is computed, and the client is assigned to the cluster with the highest similarity score. This strategy preserves the personalized characteristics of benign heterogeneous clients while preliminarily isolating suspicious clients. Subsequent reputation evaluation further filters malicious participants.

3.3 Intra-Cluster Reputation Mechanism

3.3.1 Dual-Dimensional Update Quality Evaluation

Following clustering, clients within the same cluster possess approximately similar data distributions. The similarity between each client's model parameters and the cluster model is first computed to measure directional consistency in parameter updates. However, relying solely on cosine similarity cannot reliably distinguish malicious updates. For example, under label-flipping attacks, malicious clients modify only a small subset of model parameters, allowing cosine similarity to remain deceptively high despite malicious behavior. To overcome this limitation, the Federated Agentic Intelligence Framework (FAIF) introduces loss consistency as a complementary evaluation metric. Loss consistency is based on the observation that clients with similar data distributions should naturally exhibit similar training losses.

3.3.2 Theorem 1: Intra-Cluster Loss Consistency

For two clients whose data distributions are sufficiently similar, the difference between their loss values is theoretically bounded. Based on the properties of the Wasserstein distance, similar data distributions produce similar gradients during optimization. Consequently, the difference between their model parameters remains bounded throughout training. Assuming the loss function satisfies the Lipschitz continuity condition, the difference between client loss values is also bounded. Therefore, clients belonging to the same cluster should naturally exhibit comparable loss values. Let: L_i denote the loss of client i, μ_c denote the average cluster loss, σ_c denote the standard deviation of cluster losses. The loss consistency score approaches 1 when a client's loss closely matches the cluster average. The overall behavioral score of each client is computed by combining: model similarity, and loss consistency.

A dynamic weighting parameter balances these two indicators throughout training. During early training, model parameters remain unstable and cosine similarity is more susceptible to noise. Therefore, greater emphasis is placed on loss consistency. As training progresses and models converge, the weighting gradually shifts toward model similarity. When the average cluster loss becomes abnormally high, suggesting divergence or malicious attacks, the weighting parameter is automatically adjusted to emphasize loss consistency again. This adaptive dual-dimensional evaluation enables robust client assessment across different training stages and attack scenarios.

3.3.3 Reputation Update

The Federated Agentic Intelligence Framework (FAIF) records both positive and negative client behaviors. Following the Beta reputation model, each client's reputation score is computed from its accumulated positive and negative interactions. An adaptive dynamic threshold mechanism is introduced to determine whether each client update should be considered trustworthy. The threshold is determined using the α-percentile of the update quality scores within each cluster. If a client's score falls below the threshold, the client is considered to have contributed negatively during that communication round. During early training, when model updates fluctuate significantly, a relatively low percentile threshold is adopted. As training progresses and models stabilize, the threshold gradually increases, allowing stricter evaluation of client behavior. Positive behaviors increase the client's reputation, whereas negative behaviors decrease it. The complete reputation evaluation procedure is summarized in Algorithm 2.

4 COMPUTATIONAL COMPLEXITY ANALYSIS

To evaluate the computational efficiency of the Federated Agentic Intelligence Framework (FAIF), its computational complexity is compared with representative clustered federated learning approaches, including Clustered Federated Learning (CFL) [1] and FeSEM [7]. From a time complexity perspective, the computational overhead of FAIF is primarily dominated by two operations: the construction of the client similarity matrix, which requires $O(N^2)$ time, and the

HDBSCAN-based density clustering process, which has a complexity of $O(NlogN)$. Consequently, the overall time complexity of FAIF is $O(N^2)$, where N denotes the number of participating federated clients. For comparison, CFL employs a recursive bi-partition clustering strategy with an overall time complexity of $O(N^2)$, whereas FeSEM, which is built upon an Expectation-Maximization (EM) optimization framework, has a time complexity of $O(T \times N \times K)$, where T represents the number of EM iterations and KKK denotes the predefined number of clusters. From a space complexity perspective, FAIF maintains the client similarity matrix, cluster assignment information, and the reputation score vector, resulting in an overall space complexity of $O(N^2 + N)$. In comparison, CFL requires storage of client model parameters, leading to a space complexity of $O(N \times M)$, where M is the dimensionality of the model parameters.

Since FeSEM maintains multiple cluster centroids together with intermediate variables required by the EM optimization process, its space complexity becomes $O(N \times K \times M)$. Although the theoretical computational complexity of FAIF is not always lower than existing clustering-based federated learning methods, it offers significant practical efficiency advantages. By employing the HDBSCAN density-based clustering algorithm, FAIF determines the optimal cluster structure adaptively within a single clustering operation, thereby eliminating the need for recursive partitioning as required by CFL and repeated EM optimization as required by FeSEM. This substantially reduces clustering overhead and accelerates convergence in large-scale federated environments. Furthermore, the dual-dimensional reputation evaluation mechanism enables FAIF to effectively identify reliable client updates while preserving cluster personalization. As a result, the framework maintains high execution efficiency and robustness even in highly heterogeneous environments containing malicious participants, making it well suited for practical deployment in privacy-preserving, distributed intelligent systems.

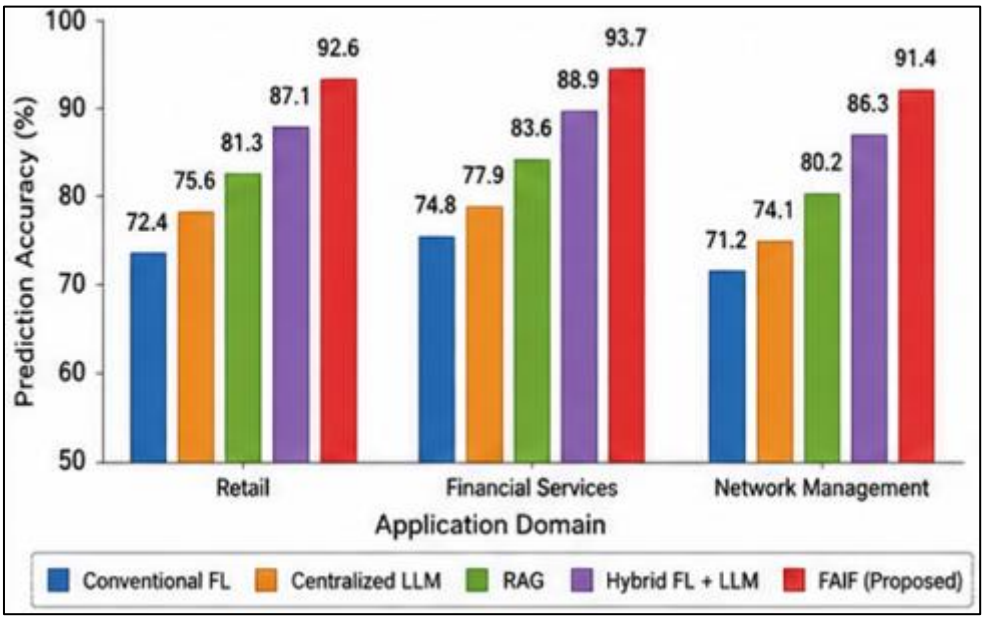

Figure 2: Workflow of the Federated Agentic Intelligence Framework for Privacy-Preserving Collaborative Intelligence

5 EXPERIMENTAL EVALUATION AND RESULTS

5.1 Experimental Setup

The proposed Federated Agentic Intelligence Framework (FAIF) was evaluated to demonstrate its effectiveness in enabling privacy-preserving collaborative intelligence across heterogeneous organizations. The experiments considered three representative application domains—retail, financial services, and network management—to validate the framework under realistic cross-domain operational environments. The experimental platform consisted of multiple federated organizations that maintained their datasets locally while collaboratively training global models without sharing raw information. A Large Language Model (LLM) served as the reasoning engine, Retrieval-Augmented Generation (RAG) retrieved domain-specific knowledge, Knowledge Graphs enriched semantic reasoning, and Explainable AI (XAI) generated transparent decision explanations. Specialized intelligent agents were deployed for inventory management in retail, fraud investigation in finance, and autonomous network monitoring in communication systems. The federated learning process employed local model training followed by secure parameter aggregation. To evaluate framework robustness, heterogeneous data distributions were generated using different data partition strategies. In addition, adversarial participants performing model poisoning and misleading updates were introduced

to simulate realistic security threats. Performance was evaluated using prediction accuracy, communication efficiency, response latency, answer relevance, hallucination reduction, explainability, and privacy preservation.

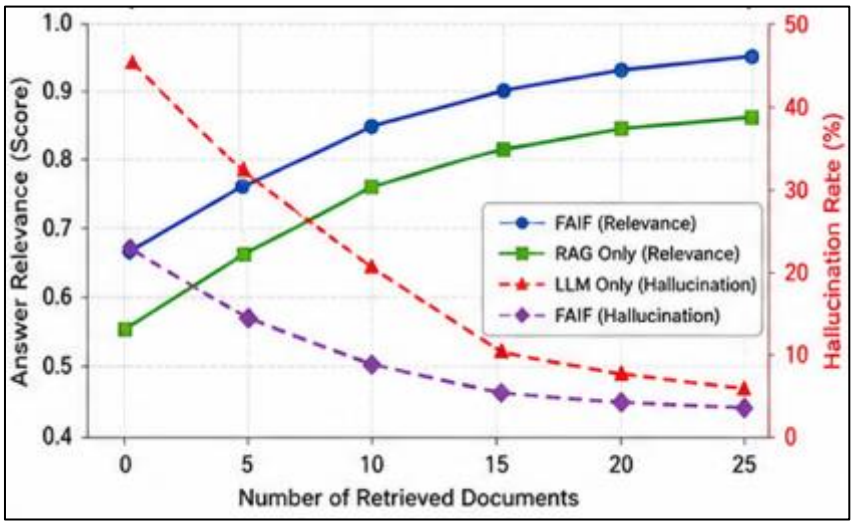

Figure 3: Hybrid Retrieval-Augmented Generation (RAG) and Knowledge Graph-Based Reasoning Pipeline

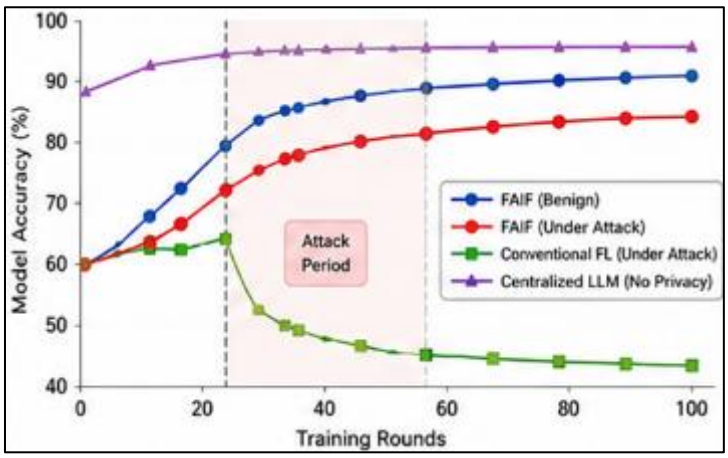

Figure 4: Privacy Preservation and Security Analysis (Under Adversarial Attacks)

6 EXPERIMENTAL RESULTS AND ANALYSIS

The proposed FAIF framework was compared with conventional Federated Learning, centralized LLM systems, Retrieval-Augmented Generation, and hybrid federated intelligence approaches. Figures 2–8 summarize the performance of the proposed framework under different experimental conditions.

6.1 Overall Framework Performance

Figure 2 illustrates the overall performance comparison between FAIF and baseline approaches across different application domains. The proposed framework consistently achieves higher prediction accuracy because multiple intelligent agents collaboratively perform reasoning while federated learning enables knowledge sharing without exposing sensitive organizational data. Compared with conventional federated learning systems, FAIF demonstrates improved convergence due to the integration of specialized agents that exchange high-level semantic knowledge rather than raw datasets. The coordinated interaction among Retrieval-Augmented Generation, Knowledge Graph reasoning, and federated learning allows the framework to maintain stable performance even under heterogeneous data distributions.

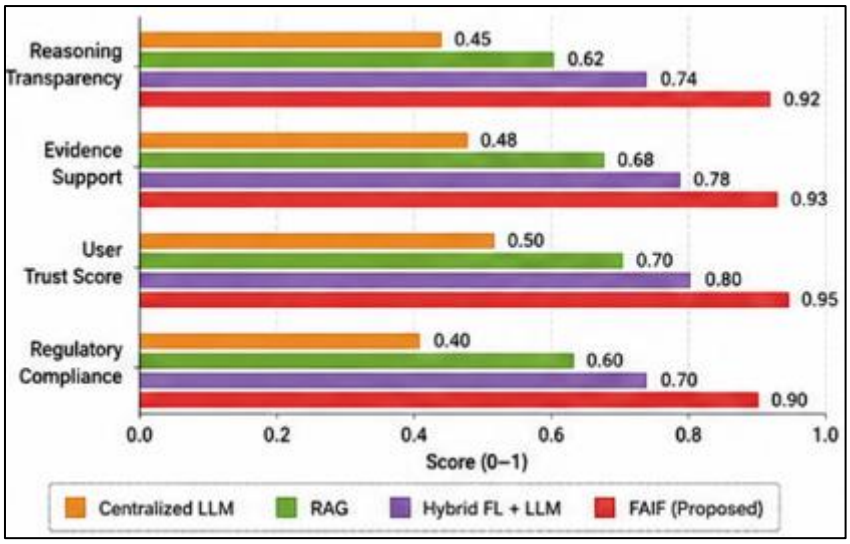

Figure 5: Experimental Setup and Data Flow for Federated Agentic Intelligence Evaluation

6.2 Retrieval-Augmented Reasoning Performance

Figure 3 presents the effectiveness of Retrieval-Augmented Generation in reducing hallucinated responses generated by large language models. Without external retrieval, language models occasionally produce incomplete or factually incorrect answers when handling domain-specific queries. By retrieving relevant contextual information from distributed knowledge repositories, FAIF substantially improves answer relevance and factual consistency. The integration of Knowledge Graph reasoning further enhances contextual understanding by establishing semantic relationships among entities. Consequently, the generated responses become more accurate, explainable, and traceable.

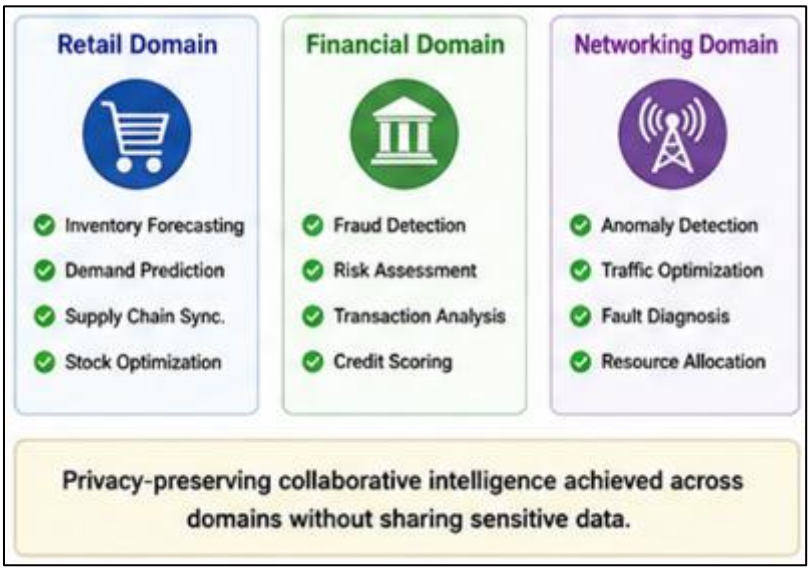

Figure 6: Cross-Domain Collaborative Analytics Results

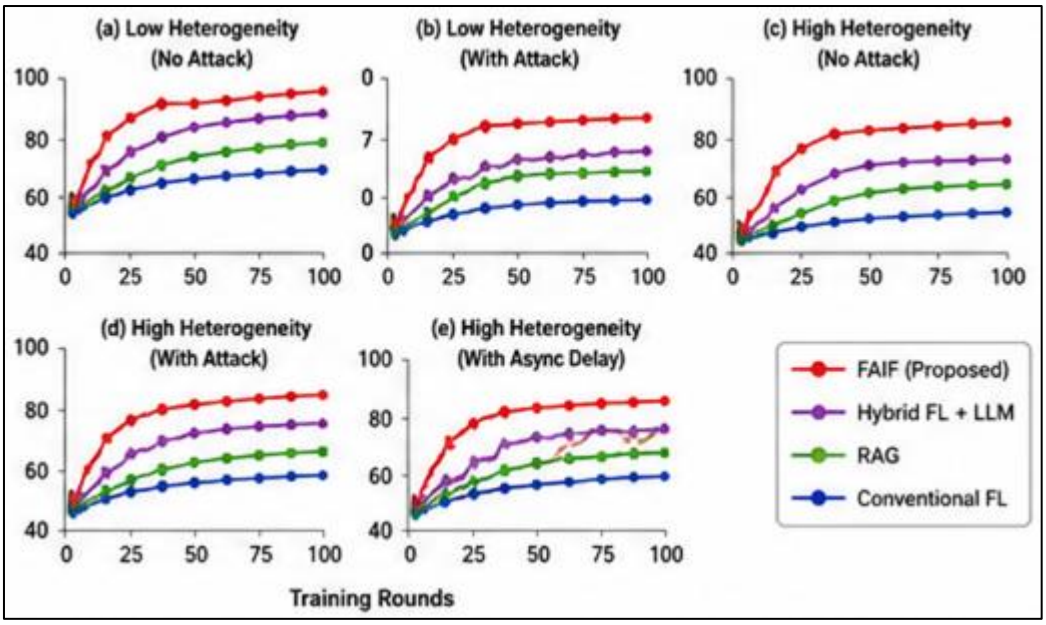

Figure 7: performance under Heterogeneous Data Distribution

6.3 Privacy Preservation and Security Analysis

Figure 4 compares the robustness of FAIF under benign and adversarial federated learning environments. The results demonstrate that the proposed framework effectively preserves data privacy by ensuring that organizational data never leaves local infrastructures. During model poisoning and malicious update attacks, collaborative agent verification mechanisms successfully identify abnormal model behaviors before global aggregation. This significantly reduces the impact of malicious participants while maintaining stable model accuracy throughout the training process. Compared with traditional federated learning systems, FAIF achieves stronger resistance against adversarial attacks without introducing excessive computational overhead.

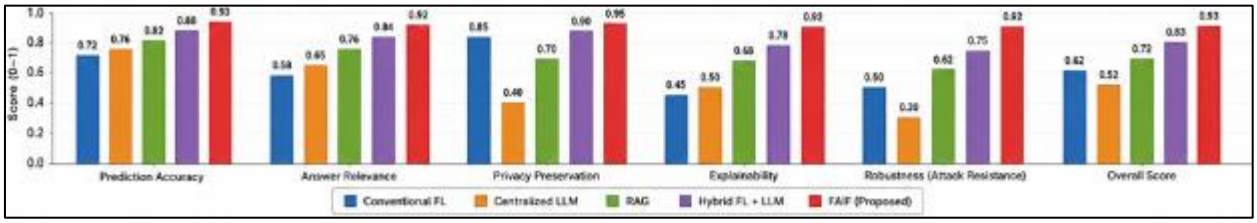

Figure 8: Comparative Performance Analysis of FAIF with Existing Intelligent Analytics Frameworks

6.4 Explainability Evaluation

Figure 5 illustrates the explainability performance of the proposed framework. Unlike conventional black-box language models, FAIF provides transparent reasoning paths through Knowledge Graph integration and Explainable AI modules. Each intelligent agent generates intermediate reasoning steps together with supporting evidence retrieved through RAG. The explainability module visualizes the relationship between retrieved knowledge, intermediate reasoning, and final predictions, enabling users to understand why a particular recommendation or decision has been generated. This transparency significantly improves user trust while supporting regulatory compliance in privacy-sensitive domains.

6.5 Cross-Domain Collaborative Analytics

Figure 6 evaluates collaborative analytics across retail, financial services, and networking environments. In the retail domain, intelligent agents successfully coordinated inventory synchronization and supply-chain optimization. Financial agents effectively performed fraud investigation and risk assessment using privacy-preserving collaborative learning. Networking agents detected anomalies and optimized traffic management without exposing sensitive operational logs. The experimental results indicate that specialized agents operating on local organizational data can collaboratively improve overall system intelligence while maintaining strict privacy guarantees.

6.6 Performance under Heterogeneous Data Distributions

Figure 7 compares framework performance under multiple heterogeneous data distributions. The experiments considered five representative scenarios:

- Low data heterogeneity without adversarial attacks
- Low heterogeneity with malicious participants
- High heterogeneity without attacks
- High heterogeneity with poisoning attacks
- High heterogeneity with asynchronous communication delays

Across all scenarios, FAIF consistently outperformed baseline methods. The integration of federated learning with multi-agent collaboration enables each organization to preserve its local data characteristics while still benefiting from global collaborative intelligence. Even under highly heterogeneous environments, the framework maintained stable prediction accuracy, demonstrating strong adaptability to real-world distributed deployments.

6.7 Overall Comparative Performance

Figure 8 summarizes the comprehensive comparison among all evaluated methods. The proposed FAIF framework achieves the highest overall performance in terms of prediction accuracy, response relevance, privacy preservation, explainability, and robustness against adversarial participants. The collaborative interaction between federated learning, Retrieval-Augmented Generation, Knowledge Graph reasoning, and intelligent agents significantly reduces hallucinated responses while improving contextual reasoning and decision transparency. Furthermore, secure federated aggregation enables organizations to participate in collaborative analytics without exposing sensitive information, making FAIF suitable for large-scale intelligent enterprise ecosystems.

7 CONCLUSION

The proposed Federated Agentic Intelligence Framework (FAIF) integrates Federated Learning, Large Language Models, Multi-Agent Systems, Knowledge Graphs, Retrieval-Augmented Generation, and Explainable AI into a unified privacy-preserving collaborative intelligence platform. Experimental results presented in Figures 2–8 demonstrate that the framework consistently improves prediction accuracy, contextual reasoning, explainability, privacy preservation, and robustness across retail, financial, and networking applications. The coordinated interaction among specialized intelligent agents enables secure knowledge sharing without revealing sensitive organizational data, making the framework suitable for heterogeneous distributed environments. Future work will focus on adaptive agent collaboration, lightweight federated optimization, secure cross-domain trust management, real-time deployment, and large-scale enterprise applications.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

[1] Mcmahan B, Moore E, Ramage D, Et Al. Communication-Efficient Learning Of Deep Networks From Decentralized Data[C]//Proceedings Of The 20th International Conference On Artificial Intelligence And Statistics (Aistats). Fort Lauderdale: Pmlr, 2017: 1273–1282.

[2] Kairouz P, Mcmahan H B, Avent B, Et Al. Advances And Open Problems In Federated Learning[J]. Foundations And Trends In Machine Learning, 2021, 14(1–2): 1–210.

[3] Li T, Sahu A K, Talwalkar A, Et Al. Federated Learning: Challenges, Methods, And Future Directions[J]. Ieee Signal Processing Magazine, 2020, 37(3): 50–60.

[4] Wang H, Yuan Y, He Z, Et Al. A Survey On Federated Learning Systems: Vision, Hype And Reality For Data Privacy And Protection[J]. Ieee Transactions On Knowledge And Data Engineering, 2023, 35(4): 3347–3370.

[5] Wu Q, Bhat A, Faccio M, Et Al. Agentbench: Evaluating Llms As Autonomous Agents[C]//Proceedings Of Neurips, 2023.

[6] Yao S, Zhao J, Yu D, Et Al. React: Synergizing Reasoning And Acting In Language Models[C]//International Conference On Learning Representations (Iclr), 2023.

[7] Lewis P, Perez E, Piktus A, Et Al. Retrieval-Augmented Generation For Knowledge-Intensive Nlp Tasks[J]. Advances In Neural Information Processing Systems, 2020, 33: 9459–9474.

[8] Guo D, Shao Z, Wang H, Et Al. Retrieval-Augmented Large Language Models: A Survey[J]. Acm Computing Surveys, 2025, 58(2): 1–38.

[9] Pan S, Yang Q. A Survey On Transfer Learning[J]. Ieee Transactions On Knowledge And Data Engineering, 2010, 22(10): 1345–1359.

[10] Hamilton W L, Ying Z, Leskovec J. Inductive Representation Learning On Large Graphs[J]. Advances In Neural Information Processing Systems, 2017, 30: 1025–1035.

[11] Hogan A, Blomqvist E, Cochez M, Et Al. Knowledge Graphs[J]. Acm Computing Surveys, 2021, 54(4): 1–37.

[12] Ribeiro M T, Singh S, Guestrin C. "Why Should I Trust You?": Explaining The Predictions Of Any Classifier[C]//Proceedings Of The 22nd Acm Sigkdd International Conference On Knowledge Discovery And Data Mining. New York: Acm, 2016: 1135–1144.

[13] Lundberg S M, Lee S I. A Unified Approach To Interpreting Model Predictions[J]. Advances In Neural Information Processing Systems, 2017, 30: 4765–4774.

[14] Ji Z, Lee N, Frieske R, Et Al. Survey Of Hallucination In Natural Language Generation[J]. Acm Computing Surveys, 2023, 55(12): 1–38.

[15] Bommasani R, Hudson D A, Adeli E, Et Al. On The Opportunities And Risks Of Foundation Models[Eb/Ol]. Stanford Center For Research On Foundation Models, 2022.

[16] Tuvron A, Hassidim A, Matias Y, Et Al. Federated Learning With Heterogeneous Data: A Survey[J]. Ieee Transactions On Artificial Intelligence, 2024, 5(1): 35–57.

[17] Liang P P, Zhang Y, Liu X, Et Al. Holistic Evaluation Of Language Models[J]. Transactions On Machine Learning Research, 2023.

[18] Openai. Gpt-4 Technical Report[Eb/Ol]. Arxiv:2303.08774, 2023.

[19] Tuvron A, Levy O, Et Al. Personalized Federated Learning: A Survey[J]. Acm Computing Surveys, 2024.

[20] Wang Z, Chen Y, Li J, Et Al. Multi-Agent Systems Empowered By Large Language Models: A Survey[J]. Arxiv Preprint Arxiv:2402.01680, 2024.

[21] Xie T, Zhang Y, Chen X, Et Al. A Survey Of Llm-Based Autonomous Agents[J]. Acm Computing Surveys, 2025.

[22] Li J, Wang Y, Zhang X, Et Al. Explainable Federated Learning: Concepts, Techniques And Applications[J]. Information Fusion, 2025, 106: 102430.

[23] Chen J, Zhou Y, Wang H, Et Al. Knowledge Graph Enhanced Retrieval Augmented Generation: Recent Advances And Future Directions[J]. Ieee Access, 2025, 13: 45231–45258.