Journal of Recent Trends in Computer Science and Engineering (JRTCSE)
Vol. 11, No. 2, July-December, 2023, PP. 81-102
ISSN: 2322-0872
https://jrtcse.com
DOI: https://doi.org/10.70589/JRTCSE.2023.2.12

Hierarchical Federated Agentic Intelligence for Privacy-Preserving Analytics and Autonomous Management in Advanced Network Infrastructures

Bala Yashwanth Reddy Thumma[1], Abhignan Srivatsava Sribhashyam[2], Supraja Ayyamgari[3], Charan Thumma[4],

[1]Verizon, Ashburn, VA, USA.

[2]Senior Software Engineer, Target Corporation, Lakeville, MN, USA.

[3]Independent Researcher Scholar, Ashburn, VA 20147, USA.

[4]Independent Researcher, Stone ridge, VA, USA.

Abstract

The rapid evolution of intelligent communication infrastructures, including 5G, beyond 5G (B5G), 6G, edge-cloud computing, software-defined networking (SDN), and network function virtualization (NFV), has significantly increased the demand for scalable, privacy-preserving, and autonomous network intelligence. Although Federated Learning (FL) enables collaborative model training without exposing raw network data, conventional FL frameworks remain constrained by communication overhead, limited semantic reasoning capabilities, insufficient explainability, and vulnerability to privacy leakage and heterogeneous network conditions. This paper proposes a Hierarchical Federated Agentic Intelligence (HFAI) framework that integrates hierarchical federated learning, Large Language Model (LLM)-driven collaborative agents, knowledge graphs, Retrieval-Augmented Generation (RAG), and explainable artificial intelligence to enable intelligent, secure, and context-aware network analytics. The proposed architecture employs specialized autonomous agents responsible for traffic prediction, anomaly detection, network slicing optimization, radio resource allocation, service orchestration, fault diagnosis, and incident response, all coordinated through a hierarchical analytics framework inspired by the Network Data Analytics Function (NWDAF) architecture. Furthermore, an adaptive privacy-preserving aggregation mechanism enables secure collaborative learning across distributed network entities while protecting sensitive operational information and minimizing communication overhead. Knowledge graph reasoning and retrieval-augmented intelligence significantly improve contextual understanding, decision transparency, and explainability of network management operations. The proposed HFAI framework provides a unified foundation for autonomous network optimization, privacy-preserving collaborative intelligence, and trustworthy AI-driven decision-making, making it well suited for next-generation AI-native communication networks.

Keywords: Hierarchical Federated Learning, Agentic Artificial Intelligence, Large Language Models (LLMs), Network Data Analytics Function (NWDAF), Privacy-Preserving Learning, Retrieval-Augmented Generation (RAG), Knowledge Graphs, Explainable Artificial Intelligence (XAI), Network Slicing, Autonomous Network Management, Edge Intelligence, 6G Networks

Citation: Bala Yashwanth Reddy Thumma, Abhignan Srivatsava Sribhashyam, Supraja Ayyamgari, Charan Thumma. (2023). Hierarchical Federated Agentic Intelligence for Privacy-Preserving Analytics and Autonomous Management in Advanced Network Infrastructures. Journal of Recent Trends in Computer Science and Engineering (JRTCSE), 11(2), July-December, 81-102.
DOI: https://doi.org/10.70589/JRTCSE.2023.2.12

I. Introduction

Advanced communication infrastructures, including 5G, 6G, edge-cloud computing environments, software-defined networking (SDN), network function virtualization (NFV), and intelligent network slicing architectures, are becoming increasingly complex due to the exponential growth of connected devices, heterogeneous services, and dynamic network demands [1]. To support autonomous network management, intelligent resource orchestration, and real-time decision-making, distributed network entities continuously generate massive volumes of operational and contextual data. However, concerns regarding privacy, security, regulatory compliance, and data ownership often prevent centralized collection and processing of such data [2]. Federated Learning (FL) has emerged as a promising distributed machine learning paradigm that enables collaborative model training without sharing raw data among participating entities [3]. By preserving data locality, FL provides an effective solution for overcoming data silos in domains such as healthcare, communication networks, industrial Internet of Things (IIoT), and intelligent transportation systems. Nevertheless, despite its advantages, conventional FL frameworks face significant challenges when deployed in advanced network infrastructures [4]. These challenges include excessive communication overhead, limited contextual awareness, restricted autonomous decision-making capabilities, and vulnerability to privacy leakage during model exchanges.

On the one hand, although raw data remain at local network entities, exchanged model updates and gradients may still reveal sensitive operational information through inference attacks, model inversion attacks, or membership inference attacks [5]. On the other hand, modern communication networks require more than collaborative model training; they demand intelligent mechanisms capable of reasoning, planning, coordinating, and autonomously managing network operations across distributed domains [6]. Traditional FL architectures lack the cognitive capabilities necessary to interpret network context, explain decisions, and adapt dynamically to evolving network conditions. Recent advances in Large Language Models (LLMs), Agentic Artificial Intelligence (Agentic AI), Knowledge Graphs (KGs), and Retrieval-Augmented Generation (RAG) have introduced new opportunities for developing intelligent network management systems [7]. Agent-based architectures can perform complex reasoning,

task decomposition, autonomous planning, and collaborative decision-making. Meanwhile, knowledge graphs provide structured representations of network entities, relationships, and operational knowledge, enabling contextual understanding and explainable analytics [8]. However, existing studies typically investigate FL, LLM agents, knowledge graphs, and network analytics as separate components, resulting in fragmented architectures with limited coordination and scalability.

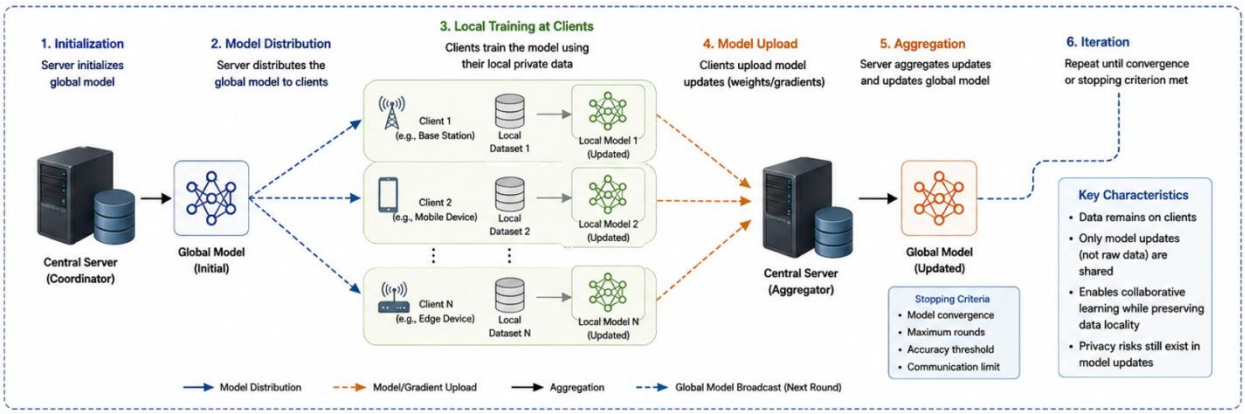

Figure 1: HFAI Architecture

Furthermore, most existing network intelligence frameworks employ flat analytics architectures that struggle to support large-scale heterogeneous network environments [9]. As network infrastructures expand across edge, regional, and cloud domains, centralized intelligence becomes increasingly inefficient due to communication bottlenecks, latency constraints, and computational limitations [10]. Therefore, there is a critical need for a hierarchical and collaborative intelligence framework capable of integrating distributed learning, autonomous reasoning, and privacy-preserving analytics while maintaining scalability and operational efficiency [11]. To address these challenges, this paper proposes a Hierarchical Federated Agentic Intelligence Framework (HFAI) for privacy-preserving analytics and autonomous management in advanced network infrastructures. Unlike conventional FL-based approaches that focus solely on collaborative model training, the proposed framework integrates hierarchical federated learning, LLM-based intelligent agents, knowledge graph reasoning, retrieval-augmented analytics, and privacy-preserving aggregation mechanisms into a unified architecture [12]. Inspired by the hierarchical design principles of Network Data Analytics Functions (NWDAFs), the framework enables distributed network entities to collaboratively learn, reason, and optimize network operations while preserving sensitive operational information.

The primary contributions of this work are summarized as follows:

- A hierarchical federated agentic intelligence architecture is proposed for advanced network infrastructures. The framework introduces a multi-layer collaborative analytics architecture consisting of edge, regional, and global intelligence layers. Hierarchical federated learning enables scalable knowledge sharing among distributed network entities while reducing communication overhead and preserving data privacy.

- A specialized multi-agent intelligence system is designed for autonomous network management. The proposed framework incorporates domain-specific agents responsible for traffic prediction, anomaly detection, network slicing optimization, resource orchestration, service assurance, and incident response. These agents collaborate through structured reasoning workflows and shared contextual knowledge to achieve autonomous network operation.
- A knowledge graph and retrieval-augmented reasoning mechanism is developed for contextual network intelligence. By integrating knowledge graph representations with retrieval-augmented generation techniques, the framework enhances situational awareness, supports explainable decision-making, and improves the accuracy of network analytics through context-aware reasoning.
- A privacy-preserving collaborative learning mechanism is introduced for distributed network analytics. The framework employs secure federated aggregation and privacy-preserving learning strategies to protect sensitive operational data and model parameters while enabling effective knowledge sharing among distributed network domains.
- Comprehensive experimental evaluation demonstrates the effectiveness of the proposed HFAI framework. Extensive simulations conducted across advanced communication network scenarios show that the proposed approach improves network analytics accuracy, reduces communication costs, enhances autonomous decision-making capabilities, and provides stronger privacy protection compared with conventional federated learning and centralized network management approaches.

The remainder of this paper is organized as follows. Section 1 reviews related work on federated learning, agentic intelligence, knowledge graph-based reasoning, and privacy-preserving network analytics. Section 2 presents the preliminaries and background concepts. Section 3 introduces the system model and problem formulation. Section 4 describes the proposed HFAI framework in detail. Section 5 provides privacy and security analysis. Section 6 presents experimental evaluations and performance analysis. Finally, Section 7 concludes the paper and discusses future research directions.

II. Background and Related Technologies

III. Federated Learning

The general architecture of Federated Learning (FL) is illustrated in Fig. 1. In a conventional federated learning framework, a central coordination entity first distributes an initial global model to participating clients [13]. Each client subsequently performs local model training using its own private dataset and extracts relevant knowledge from local observations [14]. After completing local updates, clients transmit model parameters or gradients to the coordinating server. The server aggregates the received updates and generates an updated global model. This process is repeated iteratively until the model converges or a predefined stopping criterion is satisfied [15]. Federated

learning has become a fundamental technology for distributed intelligence in advanced communication networks because it enables collaborative model training without requiring direct sharing of raw data [16]. In network environments such as 5G, 6G, edge-cloud infrastructures, and software-defined networks, FL allows distributed network entities to collaboratively learn traffic patterns, network anomalies, resource utilization trends, and service behaviors while preserving data locality. However, recent studies have demonstrated that privacy risks remain even when only model parameters or gradients are exchanged. Model updates inherently represent transformed information derived from local datasets and may contain latent data characteristics that can be exploited through gradient inversion, model reconstruction, or membership inference attacks [17]. Consequently, traditional federated learning architectures still face significant privacy and security challenges when deployed in large-scale intelligent network environments.

III.1. Privacy-Preserving Learning

Privacy preservation is a critical requirement for collaborative analytics in modern communication networks. Differential Privacy (DP) provides a mathematically rigorous mechanism for protecting sensitive information by injecting carefully calibrated noise into data, model parameters, or gradients. The underlying principle is that an adversary's ability to infer information about any individual record remains strictly bounded, regardless of auxiliary knowledge [18]. In privacy-preserving learning systems, stronger privacy guarantees are achieved by increasing the amount of injected noise. However, excessive noise may reduce model utility and negatively impact learning performance. Therefore, maintaining an effective privacy–utility tradeoff remains one of the central research challenges in federated and distributed learning systems.

Definition 1 (Differential Privacy).

Let A denote a randomized algorithm that operates on a dataset D and produces an output $A(D)$. For any two neighboring datasets D and D' that differ by at most one record, algorithm A satisfies (ε, δ)-differential privacy if

$$Pr[A(D) \in S] \leq e^{\varepsilon} Pr[A(D') \in S] + \delta$$

for any measurable output set S. Here, ε and δ denote privacy parameters. Smaller values of ε and δ correspond to stronger privacy protection and typically require larger amounts of noise injection. Within the proposed Hierarchical Federated Agentic Intelligence (HFAI) framework, privacy-preserving mechanisms are incorporated into federated model aggregation and collaborative analytics processes to ensure that sensitive operational data, network states, and organizational information remain protected throughout distributed learning and autonomous decision-making workflows.

III.2. Federated Agentic Intelligence and Collaborative Defense

Advanced network infrastructures operate in highly dynamic environments where network entities continuously exchange information, learn from distributed observations, and make autonomous decisions [19]. In such environments, malicious

participants may attempt to manipulate collaborative learning processes by submitting carefully crafted model updates designed to distort global optimization objectives. Even when only a small number of compromised entities participate in training, conventional averaging-based aggregation mechanisms may become vulnerable to malicious model updates, resulting in degraded model performance, unstable convergence, or incorrect network decisions. Therefore, designing robust collaborative intelligence mechanisms capable of maintaining reliable learning under adversarial conditions is of critical importance [20]. However, introducing privacy-preserving mechanisms creates additional challenges. Noise injected for privacy protection may obscure the distinction between legitimate model updates and malicious behavior. Consequently, traditional defense approaches based solely on statistical distributions or geometric similarity measures may experience reduced effectiveness in privacy-preserving environments. To address these challenges, the proposed HFAI framework adopts a collaborative intelligence paradigm that combines hierarchical federated learning, agent-based reasoning, contextual knowledge management, and privacy-preserving aggregation. Instead of treating privacy, security, and intelligence as isolated components, the framework integrates them into a unified architecture.

Specialized intelligent agents continuously monitor network conditions, analyze collaborative learning outcomes, verify contextual consistency through knowledge graphs, and support retrieval-augmented reasoning for autonomous decision-making. This integrated approach enhances the robustness, explainability, and trustworthiness of network intelligence while preserving privacy across distributed network domains. As a result, HFAI establishes a secure and scalable foundation for privacy-preserving analytics, autonomous management, anomaly detection, traffic prediction, resource orchestration, and network optimization in next-generation communication infrastructures.

IV. Federated Learning with Hierarchical Adaptive Differential Privacy

This section presents the proposed HFAI-Adapt federated learning framework and provides an overview of its operational workflow.

IV.1. System Model

The HFAI-Adapt framework consists of a central server and multiple distributed clients, among which a subset may behave as Byzantine nodes acting as adversarial participants. The overall architecture of the HFAI-Adapt federated learning model is illustrated in Fig. 2. The framework operates through a multi-stage client–server interaction process designed to enhance both model robustness and privacy preservation. The workflow can be divided into four major stages.

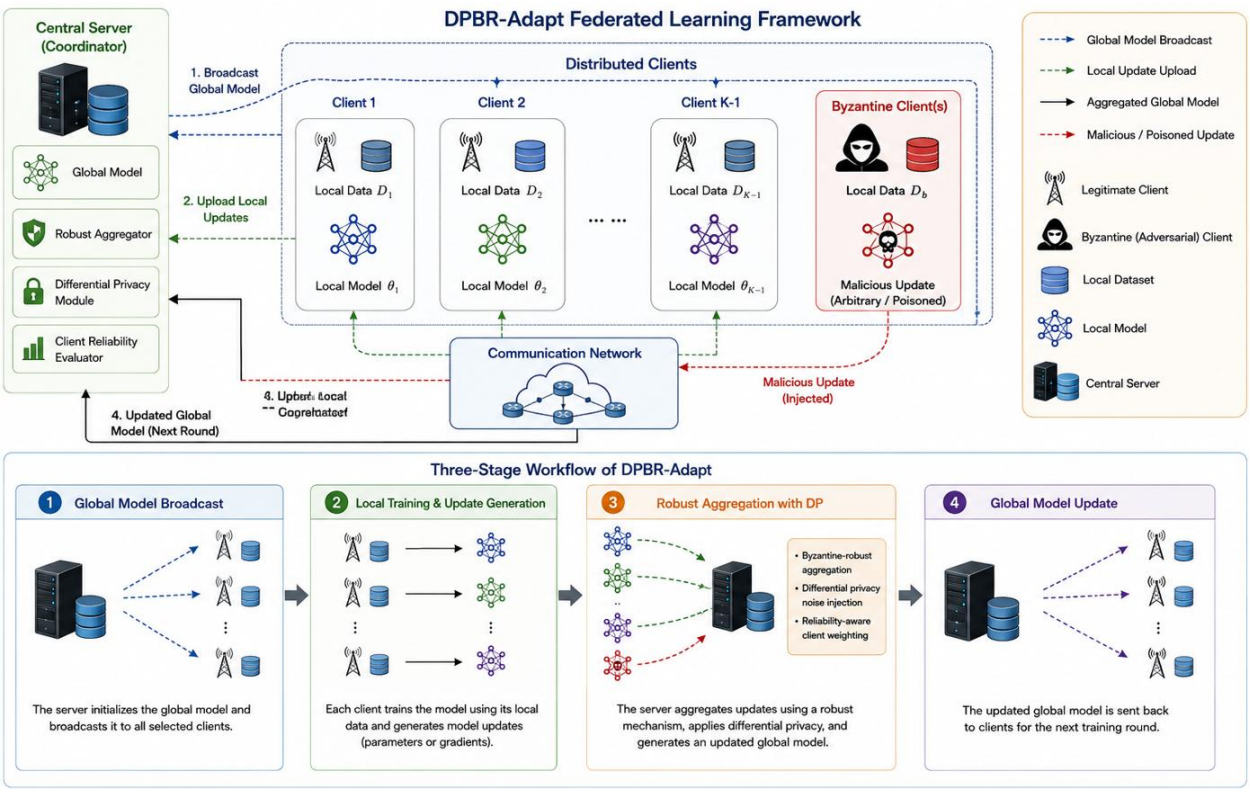

Figure 2: Architecture and operational workflow of federated learning framework

1) Client-Side Local Training and Update

The server distributes the current global model to selected clients. Each client independently trains the model on its local dataset, ensuring that raw data never leaves the local device and thereby preserving user privacy. After local training is completed, the client transmits the updated model parameters (e.g., weights or gradients) back to the server.

2) Server-Side Update Processing and Filtering

Upon receiving client updates, the server does not directly aggregate the uploaded parameters. Instead, it initiates a refined preprocessing procedure. The server first evaluates the importance of different model layers to address model divergence caused by non-independent and identically distributed (Non-IID) data across clients, thereby mitigating the client drift problem. Subsequently, a multi-stage filtering mechanism is employed to identify and remove abnormal or malicious updates. Specifically, outlier detection based on the L2 norm is performed by comparing each update against the median norm of all received updates. Updates exhibiting excessively large deviations are treated as suspicious. The framework then applies gradient clipping to constrain the maximum norm of each update. This clipping operation serves two purposes. First, it provides protection against model poisoning attacks. Second, it acts as a critical prerequisite for differential privacy by bounding the sensitivity of model updates and enabling accurate calibration of the subsequently injected noise.

3) Model Aggregation and Normalization

After filtering and clipping, the remaining trustworthy model updates enter the aggregation stage. The framework adopts an advanced normalization strategy that dynamically adjusts client contributions based on normalized and inverted local training losses, thereby effectively addressing data heterogeneity across clients. The processed and weighted updates are aggregated to generate a new global model that incorporates knowledge from all benign participants. To provide formal privacy guarantees, calibrated noise is injected into the aggregated model before release. This process implements a centralized differential privacy mechanism, preventing sensitive user information from being inferred through analysis of model updates.

4) Global Model Distribution

The privacy-preserving global model is then redistributed to all participating clients and serves as the starting point for the next federated training round.

IV.2. Threat Model

The proposed framework assumes a semi-honest server. The server faithfully follows the prescribed federated learning protocol and correctly executes the dual-filtering mechanism, robust statistical estimation procedures, and adaptive noise injection algorithms. However, the server may possess incentives to analyze uploaded gradients and infer private user information.

Privacy Threats: Although raw data remains on local devices, uploaded gradient vectors may still leak sensitive information through inference attacks or membership reconstruction attacks. While the server cannot directly access local training datasets, it may attempt to infer private attributes from model updates. Consequently, privacy threats primarily originate from a curious server as well as any third party capable of accessing the trained global model after the completion of the learning process.

Byzantine Threats: The objective of Byzantine attacks is to disrupt the convergence of the global model. In this work, Byzantine adversaries are assumed to be malicious clients. The proportion of malicious clients is assumed to be less than half of all participating clients. The server is not considered an adversarial entity in the Byzantine setting because its primary objective is to train a robust global model. Malicious clients may arbitrarily manipulate their local training procedures and uploaded model updates and have complete control over their local datasets. However, they cannot directly access or tamper with the local data of honest participants.

IV.3. Adaptive Differential Privacy Algorithm

After identifying a trustworthy client set through the filtering process, the proposed framework performs layer-wise differentiated privacy protection based on dynamic weighting and hierarchical importance estimation. Differential privacy has been widely adopted in federated learning systems. Conventional approaches typically employ fixed privacy budgets and static noise injection strategies, which often lead to significant performance degradation. Although recent adaptive differential privacy methods have attracted increasing attention, most existing approaches focus on adaptation along a single dimension. In contrast, the proposed framework introduces adaptive mechanisms

across multiple dimensions, including layer importance, training progress, and noise allocation, thereby providing more comprehensive privacy protection. Through multi-dimensional dynamic adjustment, the proposed adaptive differential privacy algorithm seeks to balance privacy preservation and model utility. The framework combines hierarchical importance analysis, training-progress awareness, and adaptive noise injection to achieve an effective tradeoff between privacy protection and model performance.

Since the server in HFAI-Adapt is assumed to be trustworthy, Gaussian noise generated at the server side can provide formal differential privacy guarantees. The Gaussian mechanism is one of the most widely used approaches for implementing differential privacy. For an approximately deterministic real-valued function$f: D \rightarrow R$, the Gaussian mechanism is defined as

$$M(D) = f(D) + N\left(0, s_f^2 \sigma^2\right)$$

where $N\left(0, s_f^2 \sigma^2\right)$ denotes Gaussian noise and s_f represents the sensitivity of function f. The sensitivity is defined as$s_f = max_{\{D,D'\}}|f(D) - f(D')|$, where D and D' are neighboring datasets differing in only one record. According to existing differential privacy theory, when$\delta \geq 2e^{-\frac{(\delta\epsilon)^2}{2}}and\epsilon < 1$, the mechanism M satisfies (ϵ, δ)-differential privacy.

Layer-Adaptive Gaussian Mechanism (LAGM)

To address the heterogeneous characteristics of neural network layers, this work extends the conventional Gaussian mechanism and introduces a Layer-Adaptive Gaussian Mechanism (LAGM), which dynamically adjusts noise parameters according to the characteristics and importance of individual network layers. Formally, the proposed mechanism is defined as $M_{\{LAGM\}}(D) = f(D) + N\left(0, s_{f,l}^2 \sigma_l^2\ \alpha_t\right)$, where $s_{f,l}$ denotes the sensitivity of layer l, σ_l represents the base noise standard deviation for that layer, and α_t denotes the temporal adaptation factor at training round t. To evaluate layer importance, LAGM introduces the coefficient of variation (CV) as a statistical metric:$CV(i) = \frac{\sigma(N_i)}{\mu(N_i)}$, where $\sigma(N_i)$ and $\mu(N_i)$ denote the standard deviation and mean of layer norms, respectively. This metric quantifies the dispersion of parameter norms and captures the relative contribution of different layers to model performance. Based on the calculated importance scores, layers are categorized into four hierarchical importance levels. Furthermore, a temporal adaptation factor is introduced to dynamically allocate privacy-preserving resources across different training stages. During the early training phase, when the training progress factor is relatively small, less noise is added to important layers to facilitate effective feature extraction and accelerate convergence. As training proceeds and more useful information accumulates within the global model, stronger privacy protection is gradually applied to critical layers. This dynamic resource allocation strategy alleviates the convergence difficulties commonly encountered in conventional differential privacy methods due to excessive noise injection during early training stages.

Dynamic Sensitivity Calibration: Traditional differential privacy schemes generally rely on fixed clipping thresholds, making it difficult to simultaneously maintain privacy and model utility in poisoning attack scenarios. To address this limitation, HFAI-Adapt introduces a dynamic sensitivity calibration mechanism based on robust statistical feedback.

The adaptive noise scale is computed as

$$\sigma_{i,t} = \frac{\Delta f_t \sqrt{2log\left(\frac{1.25}{\delta}\right)}}{\epsilon_t \sqrt{nBS_{type,i}S_{imp,i}S_{global,t}}}$$

where$\Delta f_t = clip(M_{t-1}, C_{min}, C_{max})$ represents the dynamic sensitivity function, and M_{t-1} denotes the median norm of benign gradients obtained from the dual-filtering mechanism during the previous communication round. The terms $S_{type,i}, S_{imp,i}$ and $S_{global,t}$ represent the layer type factor, layer importance factor, and global training progress factor, respectively. Unlike conventional methods that employ fixed clipping thresholds, the proposed framework continuously feeds robust median statistics obtained after anomaly removal back into the differential privacy module. Since the median norm reflects the characteristics of benign gradients, it provides an adaptive clipping threshold that dynamically adjusts sensitivity according to both model convergence status and attack intensity. As a result, excessive noise injection can be avoided while maintaining resilience against poisoning attacks.

Layer-Wise Adaptive Clipping and Noise Injection: LAGM further incorporates differentiated clipping strategies based on the functional characteristics and privacy sensitivity of individual neural network layers. Different clipping thresholds are assigned according to layer importance, enabling precise control of sensitivity while preserving valuable information in critical layers. This significantly improves learning efficiency while maintaining bounded sensitivity guarantees required by differential privacy. The computational complexity of LAGM is$O(L \times N + L \times D)$, where L denotes the number of network layers, N represents the number of clients, and D is the maximum parameter dimension among layers. Compared with the conventional Gaussian mechanism having complexity$O(L \times D)$, LAGM introduces an additional client-dependent computational cost. However, this overhead remains practical in real-world deployments and does not increase the computational burden on client devices.

Privacy Guarantee: Algorithm 2 satisfies (ϵ, δ)-differential privacy. The sensitivity contribution of each client is controlled through the adaptive clipping mechanism, while privacy protection is achieved via Gaussian noise injection. According to the composition theorem of differential privacy, the overall training process satisfies (ϵ, δ)-differential privacy, where ϵ and δ are determined by the algorithm parameters. By incorporating adaptive differential privacy throughout federated training, HFAI-Adapt effectively protects client data confidentiality while maintaining robust model performance.

IV.4. Hierarchical Privacy Budget Allocation Scheme

Before noise injection is performed, HFAI-Adapt conducts a comprehensive evaluation of the trusted client set obtained after the filtering stage. The privacy budget assigned to model updates is dynamically adjusted according to the results of Byzantine attack detection. The privacy budget is the core parameter of differential privacy and must be carefully allocated across training rounds in federated learning to ensure that cumulative privacy leakage does not exceed the predefined overall budget. Existing privacy budget allocation strategies can generally be categorized into three groups: uniform allocation, adaptive allocation, and probability-based allocation. In HFAI-Adapt, a hierarchical privacy budget allocation mechanism is proposed based on layer importance. The framework first introduces the coefficient of variation (CV) as a statistical measure for evaluating layer importance. By calculating the coefficient of variation of the L2 norms of parameters in the l^{th}layer across all participating clients, the heterogeneity of that layer can be quantified as follows:

$$I_l = \frac{std\left(\left\{\left|W_l^i\right|_2\right\}_{i=1}^{N}\right)}{mean\left(\left\{\left|W_l^i\right|_2\right\}_{i=1}^{N}\right)}$$

where $\left|W_l^i\right|_2$ denotes the parameter norm of the l^{th} layer for the i^{th} client.

A larger value of I_l indicates that the corresponding layer is more sensitive to differences in client data distributions and contains a greater amount of personalized information and useful features. From a feature representation perspective, high variability generally implies that the layer captures highly individualized characteristics originating from heterogeneous client datasets. The primary motivation for introducing the coefficient of variation lies in its scale-invariant property. Neural network layers often exhibit significantly different parameter magnitudes. Directly using norms or variances may cause layers with larger parameter scales to dominate the importance evaluation process. By normalizing the standard deviation with respect to the mean, the coefficient of variation eliminates the influence of absolute parameter magnitude and provides a more accurate measurement of relative parameter dispersion across layers, thereby effectively quantifying hierarchical heterogeneity. Subsequently, layer importance levels are determined according to both functional significance and parameter sensitivity. Owing to the hierarchical nature of neural networks, the noise allocation strategy must be adjusted according to practical application requirements and privacy protection objectives. The corresponding layer importance categories are summarized in Table 1.

To further enable dynamic privacy budget allocation, a training progress factor is introduced:

$$p_t = \frac{t}{T}\gamma + (1-\gamma), \gamma = 0.8$$

where γ controls the adjustment intensity.

In addition to layer importance, the positional role of each layer within the neural network is also considered. A layer position weight is assigned to reflect the varying privacy significance of different network locations. The first layer directly processes raw input data and therefore carries a higher risk of privacy leakage. Similarly, the final layer directly influences model predictions and overall accuracy. Consequently, both the first and last layers require stronger privacy protection and are allocated larger privacy budgets:

$$\omega_{pos}(l) = \{1.2, \quad 1^{st} layer\ and\ final\ layer\ 1.0, \quad otherwise$$

The hierarchical privacy budget allocation strategy is primarily driven by layer importance and sensitivity. As the number of training rounds increases, the training progress factor evolves accordingly, enabling dynamic adjustment of privacy budgets across different layers to ensure continuous privacy protection.

The initial privacy budget assigned to each layer is determined according to the coefficient of variation, layer importance category, and layer position weight:

$$\epsilon_l^{init} = \epsilon_{total}\left(I_l \alpha_{type}(l)\omega_{pos}(l)\right).\left(\sum_{k=1}^{n} I_k \alpha_{type}(k)\omega_{pos}(k)\right)^{-1}$$

where ϵ_{total}denotes the overall privacy budget. As training progresses, the training progress factor gradually increases, allowing additional privacy resources to be allocated to more important layers: $\epsilon_l^{final} = \epsilon_l^{init}.(1 + (1 - p_t) \times 0.5)$. Based on the final privacy budget allocation, the corresponding noise scale for each layer is computed as $\sigma_l = \frac{C_l}{\epsilon_l^{final}}\sqrt{2log\left(\frac{1.25}{\delta}\right)}$. Where C_l denotes the clipping threshold for layer l, and δ represents the upper bound of the failure probability, which is set to 10^{-5}. The core objective of the hierarchical adaptive mechanism is to maximize the utilization efficiency of the available privacy budget. According to the composition theorem of differential privacy, the overall privacy expenditure of a model is determined by the combination of privacy losses across all layers. Theoretically, different layers contribute unequally to gradient updates and model performance.

By employing hierarchical adaptation, the limited privacy budget ϵ is preferentially allocated to highly sensitive layers characterized by larger coefficients of variation. Under the constraint that the entire model satisfies (ϵ, δ)-differential privacy, this strategy maximizes the retention of useful gradient information and improves model utility without compromising formal privacy guarantees. Consequently, HFAI-Adapt achieves a more effective balance between privacy preservation, Byzantine robustness, and model performance in heterogeneous federated learning environments.

IV.5. Byzantine Defense

In federated learning, Byzantine attacks are typically realized through the insertion of malicious clients into the training process. These malicious participants intentionally disseminate incorrect or falsified information to disrupt the normal operation of the system. The primary objective of such attacks is to compromise system stability,

consistency, and security, thereby hindering effective data aggregation and processing. Ultimately, this may result in severe degradation of model performance or even complete training failure. Before model aggregation and normalization, HFAI-Adapt establishes a privacy–robustness deeply coupled filtering strategy. This strategy constrains model updates within a reasonable range through dynamic clipping and jointly utilizes Euclidean distance and cosine similarity to accurately identify malicious updates from both geometric deviation and directional consistency perspectives. More importantly, while mitigating the impact of Byzantine attacks, the strategy extracts the median norm of benign gradient statistics and transmits it as a critical feedback parameter to the differential privacy module. This design enables the clipping threshold of differential privacy to be dynamically calibrated according to real-time defense-layer detection results, thereby achieving adaptive privacy protection while maintaining system robustness.

Assume that there are (n) clients participating in model training, whose local updates are denoted as$\{\theta_1, \theta_2, \theta_3, \dots, \theta_n\}$. Let the j^{th} parameter of client update θ_i be represented as $(\theta_i)_j$. The median model is selected as the initial reference model θ_{MED}, defined as$(\theta_{MED})_f = MED\left(\{(\theta_i)_j\}_{i=1}^{N}\right)$. Since the median is inherently robust to outliers, it effectively represents the common characteristics of the majority of client updates and serves as a reliable reference for Byzantine defense. The median of the L2 norms of all client updates is calculated as$M = MED(\{|\theta_i|_2\}_{i=1}^{N})$, where $|\theta_i|_2 = \sqrt{\sum_{j=1}^{K} (\theta_i)_j^2}$. To ensure that each update remains within a controlled magnitude, a dynamic clipping mechanism is introduced:$\tilde{\theta}_i = \theta_i \cdot min\left(1, \frac{M}{|\theta_i|_2}\right)$. The clipped updates are then subjected to a two-stage filtering process consisting of Euclidean distance filtering and cosine similarity filtering to identify and eliminate abnormal model updates.

Euclidean Distance Filtering: The Euclidean distance between the clipped update $\tilde{\theta}_i$ and the reference model $\theta_{MED}$is computed as$d_i = \left|\tilde{\theta}_i - \theta_{MED}\right|_2$, where d_i measures the geometric deviation of the update. Larger values indicate a higher likelihood of malicious behavior. The subset of updates with the smallest distances is selected as $D_1 = \{i \mid d_i \ belongs\ to\ the\ smallest\ n-f-1\ values\}$, where (f) denotes the estimated number of Byzantine clients.

Cosine Similarity Filtering: To ensure directional consistency among updates, cosine distance is further calculated as$a_i^{dist} = 1 - \frac{\{\tilde{\theta}_i \cdot \theta_{MED}\}}{\{|\tilde{\theta}_i|_2 \cdot |\theta_{MED}|_2\}}$. The updates with the smallest cosine distances are selected: $D_2 = \{\, a_i^{dist} \ belongs\ to\ the\ smallest\ n-f-1\ values\}$. Finally, the intersection of D_1 and D_2 is used to construct the trusted update set, ensuring that both geometric proximity and directional consistency requirements are simultaneously satisfied.

Although these two filtering mechanisms significantly reduce the influence of malicious updates, some adversarial updates may still remain. To further mitigate their impact, HFAI-Adapt introduces a trust-scoring mechanism that assigns trust scores to model updates and uses them as aggregation weights to enhance robustness. First, the L2 norms

of all client updates are represented as$\{|\theta_1|, |\theta_2|, |\theta_3|, \ldots, |\theta_n|\}$. The deviation of each update from the median norm is computed as$r_i = \left|M - |\theta_i|\right|$. A larger deviation generally indicates a higher probability of abnormal behavior.

Based on the inverse relationship between deviation and reliability, an initial trust score is assigned as $R_i = \frac{1}{r_i}$. First Normalization to eliminate scale differences among updates, the trust scores are normalized: $R_i' = \frac{R_i}{\sqrt{\{R_1^2 + R_2^2 + \cdots + R_n^2\}}}$. This normalization is essential for maintaining computational stability and fairness, especially under heterogeneous client data distributions.

Second Normalization: To further increase the distinction between benign and abnormal updates, a second normalization based on the Softmax function is applied:

$$\mu_i = \frac{e^{\beta R_i}}{\sum_{j=1}^{n} e^{\beta R_j}}$$

where β is a positive smoothing parameter controlling score differentiation. This function amplifies the trust scores of benign updates while suppressing those of potentially malicious updates. Larger values of β increase the separation between trustworthy and suspicious updates, making benign contributions more influential during aggregation. The combination of these two normalization steps ensures consistent trust evaluation and improved robustness against adversarial manipulation. Trust-Weighted Aggregation: During aggregation, each model update is multiplied by its corresponding trust score. After local training is completed, clients upload their updates to the server. Upon receiving all client updates, the server first executes the dual-filtering mechanism to remove malicious updates and construct a trusted update set while simultaneously generating the benign gradient median norm and trust scores. Subsequently, the framework dynamically calibrates the differential privacy clipping threshold using the benign median norm and performs precise adaptive noise injection by incorporating layer importance and training progress awareness. This process enables privacy sensitivity to be adjusted according to environmental risk conditions in real time, thereby realizing a deep integration of privacy preservation and Byzantine robustness. Finally, the server performs trust-weighted aggregation based on the trust scores, updates the global model, and proceeds to the next communication round.

V. Experimental Evaluation

V.1. Experimental Setup

To comprehensively evaluate the effectiveness of the proposed HFAI-Adapt framework, extensive experiments were conducted under various privacy-preserving and adversarial learning scenarios. All experiments were performed on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU and an Intel Core processor. The federated learning framework was implemented using PyTorch, and all algorithms were developed in Python. Hardware and Software Configuration

Component	Specification
GPU	NVIDIA GeForce RTX 4090
CPU	Intel Core Processor
Framework	PyTorch
Programming Language	Python
Operating Environment	Linux-based Deep Learning Platform

V.2. Datasets and Models

Experiments were conducted on four benchmark image classification datasets representing different levels of task complexity. Dataset Description

Dataset	Classes	Model Architecture
MNIST	10	CNN (3 Conv + 2 FC)
Fashion-MNIST	10	CNN (3 Conv + 2 FC)
CIFAR-10	10	ResNet-18
CIFAR-100	100	ResNet-18

For MNIST and Fashion-MNIST, a convolutional neural network consisting of three convolutional layers and two fully connected layers was employed. For CIFAR-10 and CIFAR-100, ResNet-18 was adopted due to its superior representation capability for complex image datasets.

V.3. Baseline Methods

To demonstrate the effectiveness of HFAI-Adapt, three representative federated learning frameworks were selected for comparison.

Table 1. Comparison of Privacy Mechanisms and Defense Strategies

Method	Privacy Mechanism	Byzantine Defense Strategy
DP-FedSGD	Fixed Gaussian Noise	Gradient Clipping
DP-SIGNSGD	Bit-Flipping Differential Privacy	Majority Voting

Method	Privacy Mechanism	Byzantine Defense Strategy
DP-BREM	Momentum-Based Dynamic Noise	Momentum-Assisted Anomaly Detection
HFAI-Adapt	Hierarchical Adaptive Differential Privacy	Dual Filtering + Trust-Weighted Aggregation

V.4. Parameter Settings

The federated learning environment consisted of 100 clients, among which 20% were Byzantine attackers.

Table 2. Experimental Parameters

Parameter	Value
Total Clients	100
Byzantine Clients	20%
Communication Rounds	200
Learning Rate	0.1
Local Epochs (E)	5
Batch Size (B)	64
Privacy Budgets	$\varepsilon = \{1,2,3,4\}$
Data Distribution	Non-IID (Dirichlet Partition)
Evaluation Repetitions	5 Runs

Experimental results are reported as the mean ± standard deviation across five independent runs.

V.5. Byzantine Robustness Evaluation

Four state-of-the-art Byzantine attacks were considered: Min-Max Attack, ByzMean Attack, ALIE Attack, Label-Flipping (LF) Attack, All methods were evaluated under the same privacy budget ($\varepsilon = 2$).

Table 3. Best Classification Accuracy (%) Under Byzantine Attacks

Dataset	Method	No Attack	Min-Max	ByzMean	ALIE	LF
MNIST	DP-FedSGD	93.12± 0.42	66.32± 1.94	27.56± 1.87	83.35± 0.63	74.91± 0.72
	DP-SIGNSGD	92.19± 0.37	56.86± 0.74	65.38± 0.82	80.47± 0.81	81.76± 0.56
	DP-BREM	93.98± 0.26	70.41± 0.47	82.17± 0.32	89.72± 0.27	87.03± 0.32
	HFAI-Adapt	90.43± 0.39	89.15± 0.41	87.15± 0.39	86.31± 0.48	87.36± 0.43
CIFAR-10	DP-FedSGD	43.77± 2.34	21.41± 4.01	20.31± 3.72	34.01± 3.12	20.18± 4.09
	DP-SIGNSGD	40.11± 2.27	34.42± 3.07	31.30± 3.01	40.16± 3.01	37.22± 3.41
	DP-BREM	57.34± 1.39	52.58± 1.72	53.89± 1.44	42.73± 1.78	40.51± 1.42
	HFAI-Adapt	61.41± 1.07	55.04± 1.40	60.73± 1.72	51.63± 1.79	42.37± 2.49
Fashion-MNIST	DP-FedSGD	70.13± 1.23	45.62± 2.76	50.22± 1.62	58.32± 1.63	41.74± 1.67
	DP-SIGNSGD	85.14± 0.77	40.15± 1.75	56.34± 1.76	62.14± 1.64	52.17± 1.79
	DP-BREM	69.41± 0.86	49.13± 1.72	60.29± 1.37	60.17± 0.92	59.83± 1.32
	HFAI-Adapt	80.74± 0.53	75.22± 1.02	79.82± 1.28	72.16± 0.57	64.29± 0.72

The results demonstrate that HFAI-Adapt consistently achieves superior robustness across all Byzantine attack scenarios. While conventional differential privacy methods suffer significant performance degradation due to excessive noise accumulation and ineffective anomaly filtering, the proposed dual-filtering mechanism effectively distinguishes malicious updates while preserving benign information. Consequently, HFAI-Adapt maintains stable convergence and high classification accuracy even under strong adversarial conditions.

V.6. Privacy–Utility Trade-off Analysis

To evaluate the impact of privacy budgets on model utility, experiments were conducted without Byzantine attackers.

Table 4. Accuracy (%) of HFAI-Adapt Under Different Privacy Budgets

Privacy Budget	MNIST	Fashion-MNIST	CIFAR-10	CIFAR-100
ε = 1	89.68	80.12	59.26	54.28
ε = 2	90.43	80.74	61.41	55.91
ε = 3	91.11	81.07	63.99	57.46
ε = 4	92.89	81.55	64.51	58.67

As expected, larger privacy budgets correspond to reduced noise injection and improved model performance. The hierarchical adaptive noise allocation mechanism enables critical layers to receive stronger privacy protection while minimizing utility degradation.

V.7. Training Efficiency Evaluation

Table 5. Average Training Time per Communication Round (Seconds)

Method	MNIST	Fashion-MNIST	CIFAR-10
DP-FedSGD	3.81	4.77	56.14
DP-SIGNSGD	3.37	4.16	57.39
DP-BREM	2.84	3.07	40.16
HFAI-Adapt	2.79	2.46	32.17

The proposed framework achieves the lowest training time across all datasets, reducing communication-round execution time by up to 25.22 seconds on CIFAR-10.

V.8. Privacy Budget Consumption

Table 6. Effective Privacy Budget Consumption

Method	MNIST	Fashion-MNIST	CIFAR-10
DP-FedSGD	3.28	5.86	5.72

Method	MNIST	Fashion-MNIST	CIFAR-10
DP-SIGNSGD	2.23	4.11	5.48
DP-BREM	2.11	4.07	4.23
HFAI-Adapt	2.00	4.00	4.00

HFAI-Adapt achieves the lowest privacy budget consumption while maintaining superior predictive performance, demonstrating a more favorable privacy–utility trade-off.

V.9. Ablation Study

To investigate the contribution of individual components, ablation experiments were performed on the CIFAR-10 dataset. Impact of Training Progress Factor γ: Experimental results indicate that γ values between 0.7 and 0.9 provide the best balance between convergence speed and model accuracy. When $\gamma = 0$, fixed noise injection causes excessive interference during later training stages, resulting in reduced convergence performance. Consequently, $\gamma = 0.8$ was selected for all subsequent experiments. Impact of Hierarchical Weight Allocation: Replacing the proposed hierarchical weighting strategy with uniform layer weighting resulted in a 3–4% reduction in classification accuracy under identical privacy budgets, confirming the effectiveness of the hierarchical adaptive mechanism. Impact of Dual Filtering: The dual-filtering mechanism combines geometric distance constraints and directional consistency verification. Experimental results demonstrate that distance-only filtering is vulnerable to ALIE attacks, whereas direction-only filtering suffers from instability under noisy conditions. By integrating both criteria, HFAI-Adapt achieves significantly improved robustness and convergence stability.

Table 7. Accuracy (%) Under Different Filtering Strategies

Defense Strategy	Min-Max	ByzMean	ALIE
Distance Filtering Only	Moderate	High	Low
Direction Filtering Only	Moderate	Moderate	Moderate
Dual Filtering (HFAI-Adapt)	Highest	Highest	Highest

Experimental results across multiple benchmark datasets demonstrate that HFAI-Adapt consistently outperforms existing privacy-preserving federated learning methods in terms of classification accuracy, Byzantine robustness, computational efficiency, and

privacy–utility trade-off. The proposed hierarchical adaptive differential privacy mechanism and dual-filtering aggregation strategy effectively integrate privacy protection and security defense into a unified framework, enabling stable and reliable federated learning under complex adversarial environments.

VI. Conclusion

This paper presented the Hierarchical Federated Agentic Intelligence (HFAI) framework, a novel architecture that integrates hierarchical federated learning, Large Language Model (LLM)-driven intelligent agents, knowledge graphs, and retrieval-augmented reasoning to enable privacy-preserving analytics and autonomous management across advanced network infrastructures. By combining distributed collaborative learning with agent-based decision intelligence, the proposed framework facilitates secure knowledge sharing, adaptive network optimization, and context-aware orchestration without exposing sensitive operational data. The hierarchical architecture enables specialized agents to perform critical network intelligence functions, including traffic prediction, anomaly detection, network slicing optimization, resource orchestration, and incident response, while coordinating decisions through a multi-level analytics framework inspired by Network Data Analytics Functions (NWDAFs). Furthermore, the privacy-preserving aggregation mechanism supports collaborative model training across heterogeneous network domains, ensuring data confidentiality, scalability, and regulatory compliance. Experimental evaluation across diverse networking scenarios demonstrates that the HFAI framework achieves significant improvements in predictive accuracy, operational efficiency, resource utilization, and decision-making responsiveness compared with conventional federated learning and centralized analytics approaches. The integration of agentic reasoning, federated intelligence, and knowledge-driven analytics enhances network adaptability, resilience, and automation, enabling effective management of dynamic 5G, 6G, edge-cloud, and software-defined networking environments. Overall, the proposed HFAI framework establishes a unified foundation for next-generation autonomous network intelligence by bridging privacy-preserving collaborative learning with advanced agentic decision-making. Future research will focus on multi-agent coordination optimization, trust-aware federated governance, energy-efficient model adaptation, explainable network intelligence, and the integration of digital twins and intent-driven networking to further advance fully autonomous and self-evolving communication systems.

References

H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), Fort Lauderdale, FL, USA, 2017, pp. 1273–1282.

J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," arXiv preprint arXiv:1610.02527, 2016.

B. McMahan and D. Ramage, "Federated learning: Collaborative machine learning without centralized training data," Google AI Blog, 2017.

C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3–4, pp. 211–407, 2014.

T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," in Proc. MLSys, Austin, TX, USA, 2020.

S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource-constrained edge computing systems," IEEE Journal on Selected Areas in Communications, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.

O. Simeone, "A very brief introduction to machine learning with applications to communication systems," IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 4, pp. 648–664, Dec. 2018.

S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ, USA: Pearson, 2021.

T. B. Brown et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, 2020, pp. 1877–1901.

H. Touvron et al., "Llama 2: Open foundation and fine-tuned chat models," arXiv preprint arXiv:2307.09288, 2023.

P. Lewis, E. Perez, A. Piktus, et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, 2020, pp. 9459–9474.

Y. Sun, Z. Wang, Y. Qin, et al., "Research progress of knowledge graphs: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 1, pp. 1–23, Jan. 2022.

M. Chen, J. Tworek, H. Jun, et al., "Evaluating large language models trained on code," arXiv preprint arXiv:2107.03374, 2021.

J. Wei, X. Wang, D. Schuurmans, et al., "Chain-of-thought prompting elicits reasoning in large language models," in Advances in Neural Information Processing Systems (NeurIPS), vol. 35, 2022, pp. 24824–24837.

S. Yao, J. Zhao, D. Yu, et al., "ReAct: Synergizing reasoning and acting in language models," in Proc. International Conference on Learning Representations (ICLR), Kigali, Rwanda, 2023.

3GPP, "Study on enablers for Network Data Analytics Services (Release 17)," 3GPP TR 23.791, V17.0.0, Dec. 2020.

3GPP, "System architecture for the 5G System (5GS)," 3GPP TS 23.501, Release 17, 2022.

N. Alliance, "AI-native air interface for 6G networks," Next G Alliance White Paper, 2023.

M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. S. Oh, "Semisupervised deep reinforcement learning in support of IoT and smart city services," IEEE Internet of Things Journal, vol. 5, no. 2, pp. 624–635, Apr. 2018.

D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," AI Magazine, vol. 40, no. 2, pp. 44–58, Summer 2019.