Learning to Collaborate: An Orchestrated-Decentralized Framework for Peer-to-Peer LLM Federation

Inderjeet Singh[†], **Eleonore Vissaul-Gaudin, Andikan Otung, Motoyoshi Sekiya**

Fujitsu Research of Europe
Slough, United Kingdom
{inderjeet.singh, eleonore.gaudin, andikan.otung, motoyoshi.sekiya}@fujitsu.com

Abstract

Fine-tuning Large Language Models (LLMs) for specialized domains is constrained by a fundamental challenge: the need for diverse, cross-organizational data conflicts with the principles of data privacy and sovereignty. While Federated Learning (FL) provides a framework for collaboration without raw data exchange, its classic centralized form introduces a single point of failure and remains vulnerable to model inversion attacks. Decentralized FL (DFL) mitigates this risk by removing the central aggregator but typically relies on inefficient, random peer-to-peer (P2P) pairings, forming a collaboration graph that is blind to agent heterogeneity and risks negative transfer. This paper introduces KNEXA-FL, a novel framework for *orchestrated decentralization* that resolves this trade-off. KNEXA-FL employs a non-aggregating Central Profiler/Matchmaker (CPM) that formulates P2P collaboration as a contextual bandit problem, using a LinUCB algorithm on abstract agent profiles to learn an optimal matchmaking policy. It orchestrates direct knowledge exchange between heterogeneous, PEFT-based LLM agents via secure distillation, without ever accessing the models themselves. Our comprehensive experiments on a challenging code generation task show that KNEXA-FL yields substantial gains, improving Pass@1 by $\approx 50\%$ relative to random P2P collaboration. Critically, our orchestrated approach demonstrates stable convergence, in stark contrast to a powerful centralized distillation baseline which suffers from catastrophic performance collapse. Our work establishes adaptive, learning-based orchestration as a foundational principle for building robust and effective decentralized AI ecosystems.

Code — https://github.com/FujitsuResearch/knexa-fl

Introduction

Adapting Large Language Models (LLMs) to specialized domains presents a fundamental dilemma: the need for diverse, cross-organizational data clashes with inviolable principles of data sovereignty, privacy, and security. While Federated Learning (FL) (McMahan et al. 2017) offers a paradigm for collaborative training without raw data exchange, its canonical server-centric form introduces a trusted central aggregator. This entity becomes a single point of failure and a privileged attack surface for sophisticated model inversion attacks that can reconstruct sensitive training data (Geiping et al. 2020; Zhao et al. 2024).

Decentralized FL (DFL) addresses this by eschewing a central server in favor of direct peer-to-peer (P2P) communication (Roy et al. 2019; Itahara et al. 2023). However, in removing the central aggregator, DFL architectures typically regress to simplistic interaction strategies, such as random or static peer pairings. This random-walk approach to collaboration is statistically inefficient, blind to agent heterogeneity, and risks *negative transfer*, where poorly matched peers degrade each other's performance. It fails to strategically harness the network's latent collective intelligence.

This paper argues that the prevailing dichotomy between a vulnerable central aggregator and inefficient random P2P interaction is too limited. We introduce KNEXA-FL, a novel hybrid framework that enables *orchestrated decentralization*. It features a non-aggregating **Central Profiler/-Matchmaker (CPM)** whose sole purpose is to learn an optimal matchmaking policy for a federation of heterogeneous, autonomous LLM agents. The CPM operates on abstract, privacy-preserving agent profiles to solve a contextual bandit problem, intelligently routing knowledge exchange without ever accessing or storing the models themselves. The actual knowledge transfer, via prediction distillation compatible with parameter-efficient fine-tuning (PEFT), occurs directly and securely between the matched peers.

Our contributions are threefold:

- We designed and propose KNEXA-FL, a hybrid, centrally orchestrated - decentralized learning architecture for collaborative LLM fine-tuning. It resolves the matchmaking inefficiency of DFL without re-introducing the security vulnerabilities of a central model aggregator, thus offering a more robust and effective collaboration paradigm.
- We are the first to formulate and solve the P2P LLM collaboration problem as a contextual bandit. Our CPM employs a LinUCB algorithm to learn a pairing policy from interaction rewards, dynamically optimizing the P2P knowledge graph to maximize network utility.
- We provide comprehensive empirical validation on a challenging, heterogeneous code generation task. Our experiments show that KNEXA-FL substantially outper-

[†]Corresponding author.

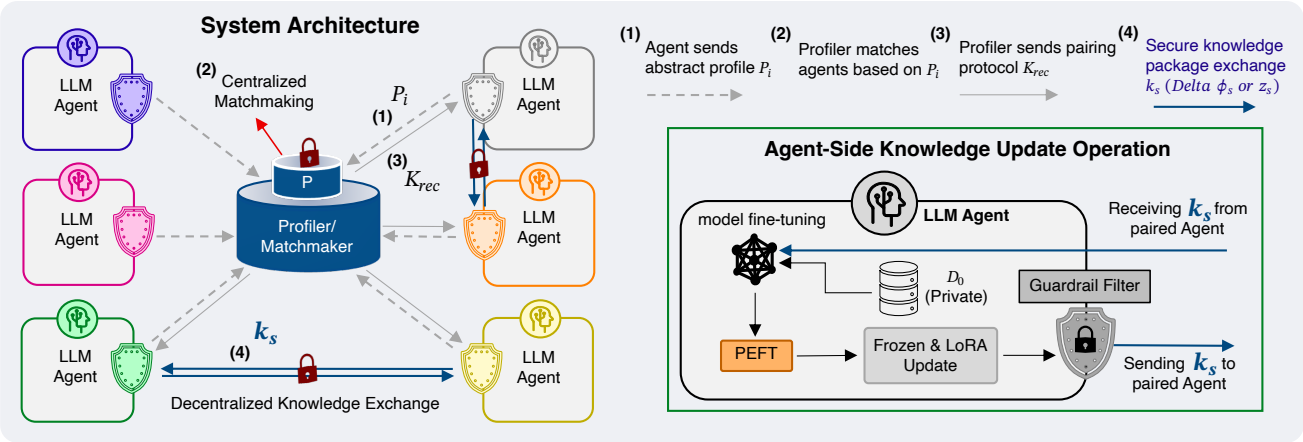

Figure 1: The KNEXA-FL architecture for orchestrated decentralization. Agents send abstract profiles ($\mathbf{p}_i$) to a non-aggregating Central Profiler/Matchmaker (CPM), which provides pairing directives (K_{rec}). Paired agents then conduct direct, secure P2P knowledge exchange (k_s). The agent-side view (right) details local PEFT fine-tuning and the privacy-enforcing Guardrail Filter.

forms isolated and random P2P baselines, and, critically, demonstrates stable convergence where a state-of-the-art centralized knowledge distillation approach suffers from catastrophic collapse.

Related Work

KNEXA-FL integrates ideas from four strands of literature: *(i)* decentralized / peer-to-peer federated learning, *(ii)* parameter-efficient LLM federation, *(iii)* security-aware governance in collaborative AI, and *(iv)* orchestration, data spaces, and multi-agent systems. We briefly position our contributions in each area.

Federated and Decentralized Learning

Federated Learning (FL) enables collaborative model training by exchanging parameter updates rather than raw data (McMahan et al. 2017). Canonical server-centric FL, however, struggles with statistical heterogeneity (non-IID data) and system heterogeneity (diverse model architectures). While methods like FedProx (Li et al. 2020) and SCAFFOLD (Karimireddy et al. 2020) address statistical drift, they maintain a central aggregator. Similarly, server-mediated knowledge distillation (e.g., FedMD (Li and Wang 2019)) handles architectural diversity but preserves the central server as a single point of failure and a potential privacy bottleneck.

Decentralized FL (DFL) obviates the server, typically relying on gossip-based protocols where agents average parameters with random neighbors (Hegedűs, Danner, and Jelasity 2021; Roy et al. 2019). Such methods generally assume model homogeneity and do not perform intelligent peer selection. More recent P2P frameworks have introduced greater flexibility through knowledge distillation (Itahara et al. 2023), sub-network exchange (Belal, Aksu, and Uluagac 2023), or topology-aware aggregation (Ryabinin et al. 2021). Despite these advances, the fundamental problem of *matchmaking* remains underdeveloped, often relying on random or handcrafted heuristics.

This is the critical gap our work addresses. In sharp contrast to prior art, our **CPM** formulates peer selection as a contextual bandit problem, learning to construct a high-utility interaction graph dynamically. Existing learning-based schedulers like Oort (Lai et al. 2021) and FedBalancer (Shin et al. 2022) optimize *server-client* selection for system metrics like throughput, a fundamentally different problem. The most relevant prior work, IPLS (Pappas et al. 2021), performs a *single, static peer grouping* at the outset based on model similarity. KNEXA-FL is the first framework to employ online learning to *continually* orchestrate the P2P collaboration graph, adapting to evolving agent knowledge while remaining fully non-aggregating.

Parameter-Efficient LLM Federation

The immense scale of LLMs has catalyzed research into federated fine-tuning using PEFT techniques. Centralized frameworks like FATE-LLM (Liu et al. 2023) and FedLoRA variants (Yang et al. 2024) federate PEFT modules (e.g., LoRA adapters) but rely on a central server for aggregation, which can lead to destructive parameter interference. Decentralized approaches such as FedSKD (Weng, Cai, and Zhou 2025) and KD-PDFL (Jeong and Kountouris 2023) use P2P knowledge distillation, but as noted, they lack an intelligent matchmaking mechanism. KNEXA-FL advances this line of work by uniquely combining support for heterogeneous LLM backbones and PEFT methods with CPM-guided P2P transfers, eliminating the risks of centralized aggregation while maximizing the efficacy of knowledge exchange.

Security and Governance in Collaborative AI

Security for FL and deployed ML systems covers AML risk analysis and robustness of perception/representation components (Bitton et al. 2023; Singh, Kakizaki, and Araki 2024), Byzantine-robust and cryptographic aggregation (Pillutla, Kakade, and Harchaoui 2022; Bonawitz et al. 2017), and decentralized trust via reputation or blockchain mechanisms (Kang et al. 2022; Katevas et al. 2020). However,

these approaches do not solve the problem of efficient partner selection. KNEXA-FL integrates governance directly into the learning process: the CPM's bandit model naturally learns to down-weight low-utility or malicious peers by observing their associated rewards. This is complemented by agent-side *Guardrail Filters* that enforce local data sovereignty policies. This learning-based approach to governance, where trust is an emergent property of observed utility, distinguishes our work from systems with static, rule-based policies.

Orchestration, Data Spaces, and Multi-Agent Systems

The vision of International Data Spaces (Otto and et al. 2022) emphasizes data sovereignty through policy-governed exchange. KNEXA-FL provides a concrete mechanism to realize a *knowledge data space*, where the exchanged assets are not raw data but abstract knowledge representations (distilled predictions: logits or decoded text), governed by policies enforced at both the agent level (Guardrails) and the system level (CPM orchestration).

From a mathematical and systems perspective, our framework can be viewed through the lens of Multi-Agent Systems (MAS). The problem of finding effective collaborators is a classic challenge in MAS, often addressed via coalition formation (Georgio et al. 2025) or task allocation. However, much of this work does not contend with the unique constraints of federated learning, namely extreme statistical heterogeneity and strict privacy requirements. Incentive-driven FL (Zhan et al. 2021) uses economic mechanisms to encourage participation but does not typically solve the problem of *who* should collaborate with whom for maximal learning efficacy. KNEXA-FL recasts the MAS coordination problem for federated LLMs as a contextual bandit problem. This is a significant departure from prior art: instead of relying on predefined rules or complex negotiation protocols, we leverage online learning to solve the coordination problem directly, optimizing for empirical performance in a dynamic, privacy-critical environment.

The KNEXA-FL Framework

The KNEXA-FL framework facilitates the collaborative enhancement of heterogeneous Large Language Models (LLMs) by enabling direct, orchestrated knowledge exchange among autonomous agents. It operates as a decentralized knowledge data space, circumventing data or model centralization. Interactions are guided by a CPM that analyzes dynamic, abstract agent profiles to recommend optimal P2P pairings and knowledge exchange protocols. This combination of adaptive P2P mechanisms with a learning-based central orchestrator addresses the dual challenges of LLM heterogeneity and effective, privacy-preserving knowledge sharing.

Problem Setting and System Architecture

We consider a dynamic set of N_t LLM agents $\mathcal{A}_t = \{a_1, \ldots, a_{N_t}\}$ at round t. Each agent a_i holds a frozen base model W_0 and a trainable PEFT module $\boldsymbol{\phi}_i$, primarily LoRA (Hu et al. 2021), on its private non-IID dataset D_i. The collective goal is a constrained multi-objective optimisation:

$$\min_{\{\boldsymbol{\phi}_i\}} \sum_{i=1}^{N_t} w_i \mathbb{E}_{(x,y)\sim D_i}\big[\ell(M_i(x), y)\big] \tag{1}$$

subject to a set of constraints including: privacy leakage, communication, and computation budgets.

The system architecture, illustrated in Figure 1, comprises three logical components. **(1) LLM Agents** ($\mathcal{A}$) are autonomous entities, each operating via a Local Gateway that enforces local policies and includes a Guardrail Filter to prevent sensitive data disclosure. **(2) The CPM** ($\mathcal{P}$) is a trusted, non-aggregating entity that receives abstract profiles $\mathbf{p}_i$ to orchestrate P2P interactions without accessing raw data or models. **(3) Secure P2P Communication Channels** are established directly between paired agents for ephemeral, encrypted knowledge exchange. This architecture establishes a managed data space where agents are sovereign knowledge providers, and $\mathcal{P}$ is the intelligent interaction broker.

P2P Collaboration Protocol

The KNEXA-FL protocol unfolds in iterative rounds of local training, profiling, matchmaking, and P2P knowledge exchange.

Agent-Side Operations. Each agent a_i first performs local PEFT updates by fine-tuning its module $\boldsymbol{\phi}_i$ on its private data D_i: $\boldsymbol{\phi}_i \leftarrow \boldsymbol{\phi}_i - \eta_L \nabla_{\boldsymbol{\phi}_i} \mathcal{L}_i(W_0, \boldsymbol{\phi}_i; D_i)$. Subsequently, it prepares a knowledge package κ_i for sharing. For Adaptive Knowledge Distillation (AKD), this package contains *teacher predictions* produced on a shared, privacy-vetted transfer set $\mathcal{X}_u$. These predictions may be represented as logits $\mathbf{z}_i(\mathbf{x})$ or as decoded text sequences $y_i(\mathbf{x})$; in our implementation we use the latter. All payloads are vetted by the agent's Guardrail Filter before packaging.

Adaptive Knowledge Distillation (AKD). AKD, our core exchange mechanism, uses **text-based distillation** for robustness to architectural and tokenizer heterogeneity. A teacher agent a_j generates text predictions $y_j(x)$ for each prompt $x \in \mathcal{X}_u$, which are sent to student a_i. The student a_i re-encodes the teacher's text $y_j(x)$ with its own tokenizer to create a "soft" target token sequence $\tilde{y}_j(x)$. It then minimizes a standard token-level cross-entropy loss, forcing its own output distribution $p_i(\cdot \mid x)$ to match the teacher's sequence:

$$\mathcal{L}^{\text{KD}}_{\text{total},i} = (1-\alpha_{\text{KD}})\,\mathcal{L}_i(D_i) \;+\; \alpha_{\text{KD}}\,\mathbb{E}_{x\in\mathcal{X}_u}\Big[\mathcal{L}_{\text{CE}}\big(\tilde{y}_j(x),\, p_i(\cdot \mid x)\big)\Big], \tag{2}$$

where $\mathcal{L}_{\text{CE}}$ is the token-level cross-entropy loss w.r.t. the re-encoded teacher tokens $\tilde{y}_j(x)$. This text-level approach makes the objective well-defined for any pair, sidestepping all tokenizer mismatch issues.

The Central Profiler/Matchmaker (CPM)

The CPM is the learning-based orchestrator. It intelligently pairs agents by solving a contextual combinatorial bandit problem, moving beyond random or heuristic matchmaking.

Table 1: Comparison with representative FL systems. KNEXA-FL is the first to integrate learned adaptive matchmaking into a decentralized, heterogeneous, LLM-native P2P framework. Legend: ✓= fully; (✓)= partial; —= not addressed.

Method	Core Architecture			Key Capabilities		Focus Area	
	Decentral. Arch.	P2P Exchange	Adaptive Matchmaking	Heterog. Support	Governance	Theory	LLM-Native
KNEXA-FL (ours)	✓	✓	✓(*Learned*)	✓	✓	✓	✓
Centralized Schedulers							
Oort (Lai et al. 2021)	—	—	✓(*Learned*)	—	(✓)	—	—
Decentralized Systems							
GossipLearn (Hegedűs, Danner, and Jelasity 2019)	✓	✓	—	—	(✓)	✓	—
IPLS (Pappas et al. 2021)	(✓)	✓	(✓)(*Static*)	(✓)	—	—	—
SparSFA (Wang et al. 2023)	✓	✓	—	—	✓	—	—
KD-PDFL (Jeong and Kountouris 2023)	✓	✓	—	✓	—	—	—
FedSKD (Weng, Cai, and Zhou 2025)	✓	✓	—	✓	—	—	(✓)
Centralized Baselines							
FedMD (Li and Wang 2019)	—	(✓)	—	✓	—	—	(✓)

Table 2: Comparison of Knowledge Exchange Mechanisms. AKD is the primary mechanism in KNEXA-FL due to its high heterogeneity tolerance and more favorable privacy-utility trade-off.

Mechanism	Comm. Cost	Comp. Overhead	Heterogeneity Tol.	Info Specificity / Privacy Risk
AKD (Teacher predictions)	Med–High (logits) or Low (text)	Medium–High (student train)	**High**	Medium
PEFT Module ($\Delta\phi$)	Low (e.g., MBs)	Low (Merge)	Low–Moderate (Needs $\mathbf{T}_{ij}$)	High (Specific param. changes, higher risk)

Agent Profiles and Contextual Bandit. Each agent a_i sends a privacy-preserving profile vector $\mathbf{p}_i \in \mathbb{R}^{d_p}$ to the CPM. This profile concatenates **static features** (e.g., LLM family, PEFT config), **dynamic features** (e.g., task performance, perplexity, privacy-preserving embeddings of local data distributions), and **historical/trust features** (e.g., success rates of past P2P interactions, CPM-maintained trust score). For a potential pair (a_i, a_j), the CPM forms a context vector $\mathbf{x}_{ij}^{(t)} = \varphi(\mathbf{p}_i^{(t)}, \mathbf{p}_j^{(t)}, S_{net}^{(t)})$, capturing their compatibility and the global network state.

LinUCB-based Matchmaking. We employ LinUCB (Li et al. 2010) to select a set of K_p disjoint pairs per round that maximize expected utility. The CPM models the expected reward of a pairing as $\hat{r}_{ij} = \hat{\boldsymbol{\theta}}^\top \mathbf{x}_{ij}$ and selects pairs based on an upper confidence bound (UCB) score to balance exploitation and exploration. The reward signal $r_{ij}^{(t)}$ provided by the receiving agent a_i after an exchange with a_j is a scalar value reflecting the utility of the interaction:

$$r_{ij}^{(t)} \;=\; \gamma\,(\mathcal{L}_i^{\text{pre}} - \mathcal{L}_i^{\text{post}}) \;-\; \delta\,\text{KB}_{ij}^{(t)}, \tag{3}$$

where the first term is the local loss reduction and the second penalizes communication cost ($\text{KB}_{ij}^{(t)}$). The CPM updates its bandit parameters $(\mathbf{A}, \mathbf{b})$ based on observed rewards, progressively learning the optimal matchmaking policy. The detailed matchmaking logic is presented in the Appendix.

Overall Protocol and Complexity

The complete KNEXA-FL protocol is specified in Algorithm 1. It formalizes the asynchronous, multi-phase loop involving parallel agent computation, centralized learning-based matchmaking, decentralized knowledge exchange, and the feedback mechanism that drives adaptation.

Complexity. The communication cost of a P2P exchange is dominated by the AKD payload. When sharing logits, this is $\mathcal{O}(|\mathcal{X}_u| \cdot |V|)$ bytes (e.g., FP16), which is about 62.5 MB for typical values. When sharing decoded text, the cost is $\mathcal{O}(|\mathcal{X}_u| \cdot L_{\text{avg}})$ tokens, which is substantially smaller in practice. The CPM's computational overhead is modest. Each LinUCB feedback update is $\mathcal{O}(d_p^2)$, where d_p is the profile dimensionality. The matchmaking step, if naively enumerating all $\binom{N_t}{2}$ pairs, would be $\mathcal{O}(N_t^2 d_p)$. However, by pre-filtering candidates using efficient approximate nearest neighbor search on profile embeddings, the practical complexity is reduced to a tractable $\mathcal{O}(N_t k d_p)$ for a small neighborhood size $k \ll N_t$. Our implementation with 20 agents completes a full round in under 16 minutes on a cluster of eight A100 GPUs.

Security, Privacy, and Theoretical Insight

KNEXA-FL's security is enhanced by several design principles. **Data Minimization** is achieved by exchanging only teacher predictions (logits or decoded text), never raw data. **Secure Communication** (e.g., mTLS with E2E payload encryption) ensures the CPM cannot decrypt knowledge packages. The **Non-Aggregating CPM** design mitigates central-point-of-failure risks. **Controlled Influence** is managed via the bandit, which learns to deprioritize malicious or low-quality peers, and through **Local Gateway Guardrail Fil-**

Algorithm 1: The KNEXA-FL Protocol

Initialize: Profiler $\mathcal{P}$ with LinUCB state ($\mathbf{A} \leftarrow \mathbf{I}_{d_p}, \mathbf{b} \leftarrow \mathbf{0}$).
procedure AGENTUPDATE($a_i, D_i, \boldsymbol{\phi}_i$)
    $\boldsymbol{\phi}_i \leftarrow \boldsymbol{\phi}_i - \eta_L \nabla_{\boldsymbol{\phi}_i} \mathcal{L}_i(W_0, \boldsymbol{\phi}_i; D_i)$ ▷ Local PEFT update
    $\mathbf{p}_i \leftarrow \text{GenProfile}(\boldsymbol{\phi}_i, D_i)$ ▷ Generate abstract profile
    return $\mathbf{p}_i$

for each communication round $t = 1, 2, \ldots, T$ **do**
// Phase 1: Asynchronous Profiling
    for all agent $a_i \in \mathcal{A}$ **in parallel do**
        $\mathbf{p}_i^{(t)} \leftarrow \text{AgentUpdate}(a_i, D_i, \boldsymbol{\phi}_i^{(t-1)})$
        Send profile $\mathbf{p}_i^{(t)}$ to $\mathcal{P}$.
// Phase 2: Centralized Matchmaking
    $\mathcal{P}$ computes $\hat{\boldsymbol{\theta}} \leftarrow \mathbf{A}^{-1}\mathbf{b}$ from its current state.
    $\mathcal{P}$ forms pairs $\mathcal{E}_t = \{(a_s, a_r, \mathbf{x}_{sr})\}$ by selecting K_p disjoint pairs that greedily maximize the LinUCB score: $\hat{\boldsymbol{\theta}}^\top \mathbf{x} + \beta\sqrt{\mathbf{x}^\top \mathbf{A}^{-1}\mathbf{x}}$.
    $\mathcal{P}$ dispatches matchmaking directives to agents in $\mathcal{E}_t$.
// Phase 3 & 4: P2P Exchange and Policy Update
    for all pair $(a_s, a_r, \mathbf{x}_{sr}) \in \mathcal{E}_t$ **in parallel do**
        a_r receives knowledge package κ_s from a_s and integrates it via AKD (Eq. 2).
        a_r computes reward $r_{sr}^{(t)}$ via Eq. 3.
        a_r reports feedback $(\mathbf{x}_{sr}, r_{sr}^{(t)})$ to $\mathcal{P}$.
        On feedback receipt, $\mathcal{P}$ updates its model:
        $\mathbf{A} \leftarrow \mathbf{A} + \mathbf{x}_{sr}(\mathbf{x}_{sr})^\top$; $\mathbf{b} \leftarrow \mathbf{b} + r_{sr}^{(t)}\mathbf{x}_{sr}$.

ters that scan outgoing knowledge packages for sensitive information. Future work can enhance **Verifiability** with DP-noise on logits or ZKPs for profile attestations.

Theoretically, the framework's convergence can be analyzed by viewing the CPM as inducing a dynamic collaboration graph $\mathcal{G}_t$. The LinUCB regret bounds (Li et al. 2010) ensure the CPM efficiently learns to form graphs with high-utility edges (i.e., positive expected rewards). Drawing from decentralized consensus literature (Boyd et al. 2006), the system's performance improvement is linked to the spectral properties (e.g., Fiedler value λ_2) of $\mathcal{G}_t$. As the CPM learns and adds more beneficial edges, $\mathbb{E}[\lambda_2(\mathcal{G}_t)]$ is non-decreasing, suggesting a trajectory towards monotonic performance improvement across the federation. We stress that this is an intuitive linkage to provide insight, not a formal convergence proof for the combined system.

Experiments

We conduct a comprehensive empirical study to validate KNEXA-FL, designed to answer three primary research questions:

RQ1: ***Overall Performance*** Does profiler-guided P2P collaboration in KNEXA-FL achieve superior task performance compared to isolated training, unguided P2P collaboration, and a conventional centralized knowledge distillation baseline?

RQ2: ***Matchmaking Efficacy*** To what extent are the performance gains attributable to the intelligent matchmaking of the Central Profiler/Matchmaker over heuristic or random pairing strategies?

RQ3: ***Robustness & Scalability*** How does KNEXA-FL's performance scale with the size of the federation and the degree of model and data heterogeneity among clients?

Experimental Setup

Datasets and Tasks. We focus on code generation, a challenging domain that demands complex reasoning and syntactic precision. We construct our primary dataset by merging the **HumanEval** (Chen et al. 2021) and **MBPP** (Austin et al. 2021) benchmarks, resulting in 464 unique programming problems. We create a 348/116 train/test split. To simulate realistic statistical heterogeneity, the 348 training problems are distributed among clients using a Dirichlet distribution with a concentration parameter $\alpha = 0.1$, ensuring a non-IID data landscape. An independent set of 128 problems is held out as the *knowledge-transfer set* ($\mathcal{X}_u$), used exclusively for knowledge distillation and unseen during local training.

Evaluation Metrics. Our primary metric is **Pass@k** ($k \in \{1, 5, 10\}$) (Chen et al. 2021), which measures the functional correctness of the generated code against unit tests. To assess syntactic and structural quality, we also report **CodeBLEU** (Ren et al. 2020).

Client Fleet and Models. To probe heterogeneity (RQ3), our main experiments feature a federation of **6** clients. Each client hosts a distinct open-source LLM backbone from a pool of models in the ≈500M parameter class, including `Qwen1.5-0.5B`, `Cerebras-GPT-590M`, `bloom-560m`, and `pythia-410m`. This 6-client setup is a deliberate stress test for high model/data heterogeneity, where naive collaboration fails and orchestration is most critical. All models are fine-tuned using LoRA (Hu et al. 2021), with ranks empirically optimized to keep trainable parameters between 2.2–3.0% of the total, significantly reducing communication payloads (details in Table 3).

Table 3: Heterogeneous 6-client configuration for our primary KNEXA-FL experiments. All backbones are tuned with empirically-optimized LoRA settings.

Client	Backbone Model	# Params	Train %	Data Train/Val
C0	`Qwen1.5-0.5B`	475M	2.39%	45 / 12
C1	`Cerebras-GPT-590M`	604M	2.34%	43 / 11
C2	`bloom-560m`	572M	2.20%	44 / 12
C3	`pythia-410m`	418M	3.01%	46 / 12
C4	`Qwen1.5-0.5B`	475M	2.39%	54 / 14
C5	`Cerebras-GPT-590M`	604M	2.34%	44 / 11

Baselines. We compare KNEXA-FL against a rigorous set of baselines:

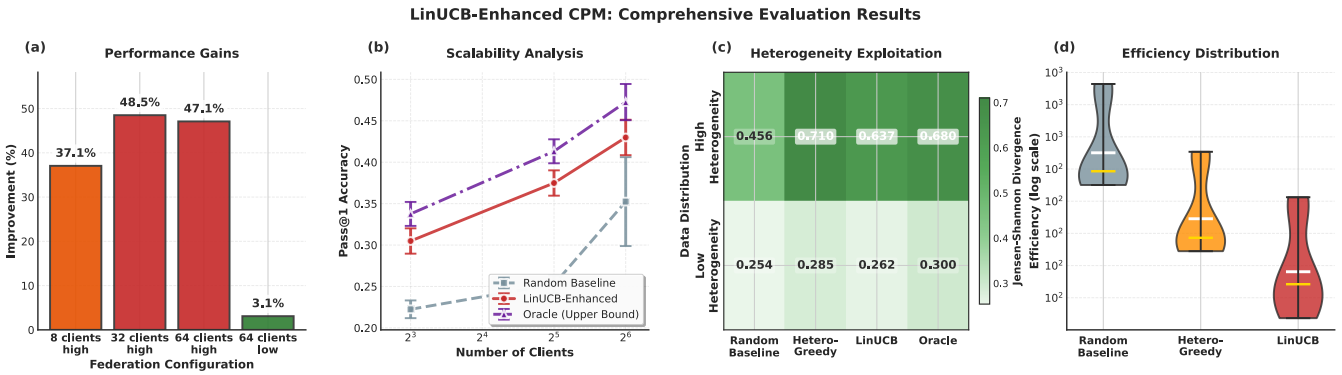

Figure 2: LinUCB-enhanced CPM comprehensive evaluation in a synthetic federation. (a) **Performance Gains:** The relative improvement over random pairing peaks at 48.5% for 32 clients in a high-heterogeneity setting and remains substantial (37.1%) even for 8 clients. Gains are modest (3.1%) in low-heterogeneity scenarios, confirming the CPM's primary value is in exploiting diversity. (b) **Scalability Analysis:** Pass@1 accuracy demonstrates our LinUCB approach consistently outperforms the random baseline across all federation sizes (8 to 64 clients) and robustly tracks towards the oracle upper bound. (c) **Heterogeneity Exploitation:** The Jensen-Shannon (JS) divergence of selected pairs reveals the CPM's learned strategy. While a naive `Hetero-Greedy` baseline maximizes JS divergence (0.710), our approach learns a superior trade-off, maintaining high divergence (≈0.64) while selecting for synergistic compatibility. (d) **Efficiency Distribution:** The computational efficiency, measured in rewards processed per second, remains highly practical, confirming that the CPM's intelligence does not introduce prohibitive overhead. Error bars and shaded regions represent 95% confidence intervals over five independent runs.

- **`LocalOnly`**: Establishes the performance frontier without collaboration. We report the average client performance after 12 rounds of isolated local fine-tuning.
- **`FedID-CentralKD`**: A modern centralized baseline, adapted from the Federated Interactive Distillation (`FedID`) framework (Ma et al. 2023). A server model (`Qwen-0.5B`) learns from client predictions on $\mathcal{X}_u$ via a confidence-weighted ensemble, then acts as a central teacher for all clients via text-level KD. This is privacy-preserving (server never sees private data).
- **`Central-KD`**: A strong centralized baseline inspired by FedMD (Li and Wang 2019). In each round, all clients submit their logits on the transfer set $\mathcal{X}_u$ to a central server. The server averages these logits to create an "ensemble teacher" distribution, which is then broadcast back to all clients for distillation. This represents a conventional approach to heterogeneous federated learning.
- **`Heuristic-P2P`**: A non-learning ablation using identical AKD. It replaces the CPM with a static heuristic: greedily pair clients to maximize data (JS) divergence, with the higher-performer as teacher.
- **`Random-P2P`**: A direct ablation of the CPM. This method uses the same P2P knowledge distillation as KNEXA-FL, but teacher-student pairings are selected uniformly at random in each of the 20 collaboration rounds.

Implementation Details. All experiments were conducted on a mix of NVIDIA H100s and A100s with clients simulated in parallel. Fixed random seed of 42 for reproducibility. Additional hyperparameters in the Appendix.

Results and Analysis

RQ1: Overall Performance. Table 4 validates our claims. The `LocalOnly` baseline (2.22% Pass@1) confirms collaboration is necessary. Our centralized baselines show extreme fragility under heterogeneity: the modern `FedID-CentralKD` failed to converge (1.11% Pass@1), reinforcing the `Central-KD` baseline's collapse (peak 18.33% → final 2.00%) and confirming aggregation is unreliable in this regime. The P2P baselines are most critical. While `Random-P2P` (8.89% Pass@1) confirms P2P's value, the `Heuristic-P2P` baseline, which maximizes data diversity (JS divergence), performed **worse** (6.67% Pass@1). This strongly suggests that naive, non-learning heuristics can be detrimental. In sharp contrast, **KNEXA-FL** achieves the highest performance (13.33% Pass@1), a **50%** relative gain over `Random-P2P` and **100%** over the failing heuristic. This demonstrates that our CPM's *learned* policy, which balances diversity and compatibility (Fig. 2c), successfully navigates the trade-offs for stable, superior performance.

The Instability of Centralized Distillation. A deeper look at `Central-KD` shows fundamental instability. Although it briefly peaked at 18.33% Pass@1 (6 clients), a controlled 4-client run with full logging confirmed a collapse to 2.00% final Pass@1. Forcing heterogeneous models to distill from a single averaged "ensemble teacher" overwrites specialized knowledge, triggering catastrophic forgetting, which KNEXA-FL's targeted, utility-driven P2P exchanges avoid.

RQ2: Matchmaking Efficacy. The performance gap between `Random-P2P` and KNEXA-FL directly points to the efficacy of the CPM. To isolate this effect, we mea-

Table 4: Average performance on the global test set. `LocalOnly` is evaluated after 12 rounds of isolated training. Collaborative methods are evaluated after 20 rounds, unless otherwise noted. Best final performance is in **bold**.

Method	Pass@1 (%)	Pass@5 (%)	Pass@10 (%)	CodeBLEU
LocalOnly	2.22	5.42	5.55	0.260
FedID-CentralKD	1.11	5.56	5.56	0.181
Central-KD	2.00 (18.33)[†]	7.80	10.00	0.268
Heuristic-P2P[‡]	6.67	16.67	27.78	0.392
Random-P2P	8.89	22.40	27.80	0.239
KNEXA-FL	**13.33**	**31.25**	**44.44**	**0.344**

[†] `Central-KD` was volatile; peaked at 18.33% (6-client) but collapsed to 2.00% (4-client instability analysis). [‡] `Heuristic-P2P` was evaluated for representative restricted rounds and data on the same 6-client setup.

sured the peak performance achieved by any student on the knowledge-transfer set during collaboration (Table 5). KNEXA-FL's CPM-guided pairings enabled a student to achieve **86.70%** Pass@1 on this set, a staggering **2.6**× improvement over the best pairing found by `Random-P2P`. This demonstrates that the CPM is not merely avoiding bad pairings but is actively discovering and exploiting highly synergistic knowledge transfers that random chance is unlikely to find.

Table 5: Peak student Pass@1 achieved on the 128-sample transfer set during collaboration. This metric isolates the quality of knowledge transfer.

Pairing Strategy	Peak Student Pass@1
`Random-P2P`	33.33%
KNEXA-FL (CPM-Guided)	**86.70%**

Profiler Ablation: LinUCB under a Controlled Synthetic Regime The gains in Table 5 imply that *who* collaborates with whom is decisive. To attribute this effect precisely to the LinUCB-driven Central Profiler/Matchmaker and to test its scalability beyond the six-client, real-model setup, we conducted a controlled ablation in a large, **synthetic** environment, an established protocol for bandit and FL research (Lattimore and Szepesvári 2020).

The results, presented in Figure 2, are definitive:

- **Substantial and Scalable Gains:** The LinUCB-enhanced CPM delivers major performance improvements over random pairing, peaking at a **48.5%** relative gain in the 32-client, high-heterogeneity case (Panel a). This advantage scales robustly, with performance consistently approaching the oracle upper bound as the federation grows (Panel b).
- **Learned Compatibility Trade-off:** The CPM learns a non-trivial strategy for exploiting diversity. While a naive 'Hetero-Greedy' baseline maximizes Jensen-Shannon divergence, our CPM intelligently trades a small amount of diversity for a large gain in synergistic compatibility, explaining its superior performance (Panel c). This is achieved with highly practical computational efficiency (Panel d).

Take-away. Under this isolated, reproducible setting, the LinUCB-driven CPM demonstrably *learns* a superior, scalable matchmaking policy that materially explains the end-to-end performance gains of KNEXA-FL, fully answering **RQ2**.

RQ3: Heterogeneous Federation. Per-client analysis of the 6-agent KNEXA-FL run shows that gains are distributed across the diverse federation. For instance, Client C2 (`bloom-560m`) evolved from a mid-tier performer to the strongest individual model, achieving a final local Pass@1 of 36.67%. Even the smallest model, C3 (`pythia-410m`), significantly surpassed its isolated performance, demonstrating that profiler-guided collaboration effectively lifts the entire ecosystem, not just the strongest members.

Discussion and Limitations

Our results show KNEXA-FL delivers substantial, stable gains with low overhead. The improvements stem from the CPM's learned P2P orchestration, contrasting sharply with the volatile, "one-size-fits-all" centralized baselines. By targeting knowledge sharing, KNEXA-FL mitigates the catastrophic forgetting seen in simpler schemes.

Limitations. Our study has limitations. Future work should: (i) validate on larger federations with realistic WAN latencies; (ii) explore more semantic (e.g., user-profile-based) data splits beyond Dirichlet partitioning; and (iii) benchmark against a wider array of advanced centralized FL optimizers.

Despite these constraints, the consistent and stable performance edge over strong baselines firmly positions profiler-guided P2P learning as a robust and promising paradigm for decentralized, collaborative LLM intelligence.

Conclusion

We introduced KNEXA-FL, a framework for orchestrated decentralization that resolves the trade-off between insecure centralized FL and inefficient random P2P collaboration. Its non-aggregating Central Profiler/Matchmaker (CPM) formulates P2P matchmaking as a contextual bandit problem, learning to optimize the collaborative graph. Empirically, on a heterogeneous code-generation task, our approach yields substantial gains (≈50% relative Pass@1 improvement over random P2P) and, critically, achieves stable convergence where a strong centralized distillation baseline catastrophically fails. Our work establishes learning-based orchestration as a core principle for robust decentralized AI. Future work includes: (i) scaling to larger federations, (ii) exploring more expressive (e.g., neural) bandit models, and (iii) integrating safeguards such as differential privacy, zero-knowledge proofs, and token-efficient disparity audits like TFDP (Singh et al. 2025).

References

Austin, J.; Odena, A.; Nye, M.; Bosma, M.; Michalewski, H.; Dohan, D.; Jiang, E.; Cai, C.; Terry, M.; Le, Q.; and Sutton, C. 2021. Program Synthesis with Large Language Models. arXiv:2108.07732.

Belal, A.; Aksu, E.; and Uluagac, A. S. 2023. P2P-FedMask: A Communication-Efficient and Byzantine-Robust Peer-to-Peer Federated Learning Approach. *IEEE TPDS*, 34(10): 2825–2837.

Bitton, R.; Maman, N.; Singh, I.; Momiyama, S.; Elovici, Y.; and Shabtai, A. 2023. Evaluating the cybersecurity risk of real-world, machine learning production systems. *ACM Computing Surveys*, 55(9): 1–36.

Bonawitz, K.; Ivanov, V.; Kreuter, B.; Marcedone, A.; McMahan, H. B.; Patel, S.; Ramage, D.; Segal, A.; and Seth, K. 2017. Practical Secure Aggregation for Privacy-Preserving Machine Learning. In *CCS*.

Boyd, S.; Ghosh, A.; Prabhakar, B.; and Shah, D. 2006. Randomized Gossip Algorithms. *IEEE Transactions on Information Theory*, 52(6): 2508–2530.

Chen, M.; Tworek, J.; Jun, H.; Yuan, Q.; Ponde de Oliveira Pinto, H.; Kaplan, J.; Edwards, H.; Burda, Y.; Joseph, N.; Brockman, G.; Ray, A.; Puri, R.; Krueger, G.; Petrov, M.; Khlaaf, H.; Sastry, G.; Mishkin, P.; Chan, B.; Gray, S.; Ryder, N.; Pavlov, M.; Power, A.; Kaiser, L.; Bavarian, M.; Winter, C.; Tillet, P.; Petroski Such, F.; Cummings, D.; Plappert, M.; Chantzis, F.; Barnes, E.; Herbert-Voss, A.; Guss, W. H.; Nichol, A.; Paino, A.; Tezak, N.; Tang, J.; Babuschkin, I.; Balaji, S.; Jain, S.; Saunders, W.; Hesse, C.; Carr, A. N.; Leike, J.; Achiam, J.; Misra, V.; Morikawa, E.; Radford, A.; Knight, M.; Brundage, M.; Murati, M.; Mayer, K.; Welinder, P.; McGrew, B.; Amodei, D.; McCandlish, S.; Sutskever, I.; and Zaremba, W. 2021. Evaluating Large Language Models Trained on Code. arXiv:2107.03374.

Geiping, J.; Bauermeister, H.; Dröge, H.; and Moeller, M. 2020. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems*, 33: 16937–16947.

Georgio, R. J.; Forder, C.; Deb, S.; Carroll, P.; and Gürcan, Ö. 2025. The Coral Protocol: Open Infrastructure Connecting The Internet of Agents. *arXiv preprint arXiv:2505.00749*.

Hegedűs, I.; Danner, G.; and Jelasity, M. 2019. Gossip learning as a decentralized alternative to federated learning. In *Distributed Applications and Interoperable Systems: 19th IFIP WG 6.1 International Conference, DAIS 2019, Held as Part of the 14th International Federated Conference on Distributed Computing Techniques, DisCoTec 2019, Kongens Lyngby, Denmark, June 17–21, 2019, Proceedings 19*, 74–90. Springer.

Hegedűs, I.; Danner, G.; and Jelasity, M. 2021. Decentralized learning works: An empirical comparison of gossip learning and federated learning. *Journal of Parallel and Distributed Computing*, 148: 109–124.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2021. LoRA: Low-Rank Adaptation of Large Language Models. *CoRR*, abs/2106.09685.

Itahara, S.; Nishio, T.; Koda, Y.; Yamamoto, K.; and Morikura, M. 2023. Peer-to-Peer Federated Knowledge Distillation for Heterogeneous Edge Devices. In *ICLR Workshop on Trustworthy Large-Scale ML*.

Jeong, H.; and Kountouris, I. 2023. Personalized Decentralized Federated Learning with Knowledge Distillation. *arXiv preprint arXiv:2302.12156*.

Kang, J.; Xiong, Z.; Niyato, D.; Yu, H.; Liang, Y.-C.; and Kim, D. I. 2022. Reputation-Based Federated Learning for Defending Against Byzantine Attacks. *IEEE Transactions on Information Forensics and Security*, 17: 2626–2641.

Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S. J.; Stich, S. U.; and Suresh, A. T. 2020. SCAFFOLD: Stochastic Controlled Averaging for Federated Learning. In *ICML*.

Katevas, K.; Bagdasaryan, E.; Waterman, J.; Safadieh, M. M.; Birrell, E.; Haddadi, H.; and Estrin, D. 2020. Policy-based federated learning. *arXiv preprint arXiv:2003.06612*.

Lai, F.; Tang, X.; Ezazi, M.; Wang, Q.; Zhang, H.; Yi, C.-C.; and Wang, C. 2021. Oort: Efficient Federated Learning via Guided Participant Selection. In *OSDI*.

Lattimore, T.; and Szepesvári, C. 2020. *Bandit Algorithms*. Cambridge University Press.

Li, D.; and Wang, J. 2019. Fedmd: Heterogenous federated learning via model distillation. *arXiv preprint arXiv:1910.03581*.

Li, L.; Chu, W.; Langford, J.; and Schapire, R. E. 2010. A Contextual-Bandit Approach to Personalized News Article Recommendation. In *Proceedings of the 19th International Conference on World Wide Web (WWW 2010)*, 661–670. Raleigh, NC, USA: ACM.

Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated Optimization in Heterogeneous Networks. In *Proceedings of MLSys*.

Liu, Y.; et al. 2023. FATE-LLM: An Industrial Grade Federated Learning Framework for Large Language Models. *arXiv preprint arXiv:2310.10049*.

Ma, X.; Liu, J.; Wang, J.; and Zhang, X. 2023. FedID: Federated Interactive Distillation for Large-Scale Pretraining Language Models. In Bouamor, H.; Pino, J.; and Bali, K., eds., *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 8566–8577. Singapore: Association for Computational Linguistics.

McMahan, H. B.; Moore, E.; Ramage, D.; Hampson, S.; and Agüera y Arcas, B. 2017. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *AISTATS*.

Otto, B.; and et al. 2022. Design Principles for Data Spaces. *Fraunhofer-Gesellschaft Report*.

Pappas, C.; Chatzopoulos, D.; Lalis, S.; and Vavalis, M. 2021. Ipls: A framework for decentralized federated learning. In *2021 IFIP Networking Conference (IFIP Networking)*, 1–6. IEEE.

Pillutla, K.; Kakade, S. M.; and Harchaoui, Z. 2022. Robust aggregation for federated learning. *IEEE Transactions on Signal Processing*, 70: 1142–1154.

Ren, S.; Guo, D.; Lu, S.; Zhou, L.; Liu, S.; Tang, D.; Sundaresan, N.; Zhou, M.; Blanco, A.; and Ma, S. 2020. CodeBLEU: a Method for Automatic Evaluation of Code Synthesis. arXiv:2009.10297.

Roy, A. G.; Siddiqui, S.; Pölsterl, S.; Navab, N.; and Wachinger, C. 2019. Braintorrent: A peer-to-peer environment for decentralized federated learning. *arXiv preprint arXiv:1905.06731*.

Ryabinin, M.; Gorbunov, E.; Plokhotnyuk, V.; and Pekhimenko, G. 2021. Moshpit sgd: Communication-efficient decentralized training on heterogeneous unreliable devices. *Advances in Neural Information Processing Systems*, 34: 18195–18211.

Shin, J.; Li, Y.; Liu, Y.; and Lee, S.-J. 2022. Fedbalancer: Data and pace control for efficient federated learning on heterogeneous clients. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, 436–449.

Singh, I.; Kakizaki, K.; and Araki, T. 2024. Advancing Deep Metric Learning With Adversarial Robustness. In *Asian Conference on Machine Learning*, 1231–1246. PMLR.

Singh, I.; Srinivasan, R.; Vainshtein, R.; and Kojima, H. 2025. TFDP: Token-Efficient Disparity Audits for Autoregressive LLMs via Single-Token Masked Evaluation. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, 24609–24626.

Wang, H.; Muñoz-González, L.; Hameed, M. Z.; Eklund, D.; and Raza, S. 2023. SparSFA: Towards robust and communication-efficient peer-to-peer federated learning. *Computers & security (Print)*, 129.

Weng, Z.; Cai, W.; and Zhou, B. 2025. FedSKD: Aggregation-free Model-heterogeneous Federated Learning using Multi-dimensional Similarity Knowledge Distillation. *arXiv preprint arXiv:2503.18981*.

Yang, J.; et al. 2024. Federated Learning of Large Language Models with Parameter-Efficient Prompt Tuning and Adaptive Optimization. *arXiv preprint arXiv:2310.15080*.

Zhan, Y.; Zhang, J.; Hong, Z.; Wu, L.; Li, P.; and Guo, S. 2021. A survey of incentive mechanism design for federated learning. *IEEE Transactions on Emerging Topics in Computing*, 10(2): 1035–1044.

Zhao, J. C.; Sharma, A.; Elkordy, A. R.; Ezzeldin, Y. H.; Avestimehr, S.; and Bagchi, S. 2024. Loki: Large-scale data reconstruction attack against federated learning through model manipulation. In *2024 IEEE Symposium on Security and Privacy (SP)*, 1287–1305. IEEE.