



ARTICLE

StabFedVAE: Stabilizing Global-Local Latent Representations for Federated Recommendation

Jincheng Cai¹ , Li Feng^{1,*} and Ni Zhao²

¹The School of Computer Science and Engineering, Macau University of Science and Technology, Macau, China

²The School of Finance and Economics, Shenzhen University of Information Technology, Shenzhen, China

*Corresponding Author: Li Feng. Email: lfeng@must.edu.mo

Received: 16 August 2026; Accepted: 11 September 2026; Published: 22 October 2026

ABSTRACT: Federated recommendation leverages distributed client data without centralizing interaction histories. However, sparse and non-identically distributed data can cause scale mismatches between global and local latent representations. We propose StabFedVAE, a global-local variational autoencoder (VAE) that disentangles shared and client-specific preference patterns. A sigmoid-gated multiplicative fusion regulates local latent contributions with provable contraction properties and bounded variance. While the global encoder-decoder are shared, local encoders remain private. We introduce a round-dependent local Kullback–Leibler (KL) penalty to stabilize the local posterior. Theoretical analysis proves that for fixed global gates, the fusion is contractive and bounds local latent variance. Empirical validation across four datasets confirms the observable variance inequality. Experiments show that moderate local KL regularization improves normalized discounted cumulative gain at cutoff 20 (NDCG@20) by 3.0%–10.3% over no regularization, while stronger penalties reduce posterior dispersion at the cost of ranking quality. Our results demonstrate that balancing latent stability and client personalization is essential rather than optimizing them independently.

KEYWORDS: Federated learning; personalized federated recommendation; variational autoencoder; latent representation

1 Introduction

Recommender systems learn preference structure from user–item interactions and now support information access in commerce, media, and online services [1–4]. Collaborative filtering is effective because it mines behavioral regularities directly from interaction records, but centralized collection creates privacy, governance, and deployment constraints. Federated learning (FL) instead keeps raw histories on clients while coordinating model updates [5,6]. Federated recommendation consequently has to transfer population-level preference knowledge without suppressing client-specific interests.

Recent surveys organize horizontal federated recommendation around its model, privacy, and federated-system dimensions [7]. Within this landscape, federated recommenders address the utility–personalization tension through neural collaborative filtering, private item representations, additive personalization, and multilevel personalized modeling [8–11], as well as model heterogeneity and communication-efficient variational learning [12–14]. These developments build on broader federated optimization analyses and classical latent-factor recommendation models [15,16]. They establish that parameter

ownership matters under non-IID interactions. However, a variational recommender introduces an additional issue: independently optimized clients can produce latent posteriors with very different scales and dispersion. The problem is especially visible in dual-encoder architectures such as FedDAE [17], where a shared encoder captures commonality and a private encoder captures individuality. Parameter separation alone does not prevent the local branch from becoming unbounded or, under excessive regularization, collapsing toward the prior.

Accordingly, this paper studies how to stabilize a private variational representation without removing its personalization capacity. The proposed STABFEDVAE has three components. Global and local encoders represent shared and client-specific evidence, respectively. A bounded global gate modulates the local latent coordinates through element-wise multiplication. A selective local Kullback–Leibler (KL) term, controlled by a communication-round schedule, constrains the private posterior without applying the same prior pressure to the shared branch. Unlike representation-consensus methods that explicitly attract clients to a common prototype or representation [18,19], STABFEDVAE controls local posterior scale while allowing nonzero client dispersion.

We investigate a non-monotonic utility–stability relationship. Without local regularization, posterior dispersion can become large; increasing regularization can control latent scale; excessive regularization can contract the local posterior toward the prior and impair ranking utility. The three-seed results support this trade-off, though the optimal coefficient is dataset- and seed-dependent.

The principal contributions are as follows:

- We develop a dual-encoder federated VAE with bounded multiplicative fusion, selective local KL regularization, and persistent client-private encoder states.
- We establish a local-channel contraction bound and define complementary dispersion and deterministic fused-variance diagnostics. The directly observable proxy consequence is verified on 7200 round-level evaluations.
- We evaluate the method on four public datasets through a fixed- $\beta_l^* = 1$ reference comparison and a six-value, three-seed KL sensitivity analysis, while separating validation-selected evidence from the common reference condition and from published baselines.

2 Related Work

2.1 Federated Recommendation

Early federated recommenders explored privacy-preserving matrix factorization and collaborative learning without centralizing ratings [7,20]. Recent work extends this foundation to nonlinear, heterogeneous, and personalized recommenders. FedNCF federates neural collaborative filtering [9]; PFedRec personalizes user and item representations [10]; FedRAP decomposes item representations into shared and additive personalized components [11]; HeteFedRec addresses model heterogeneity across clients [13]; and FPFRR combines user-local, cluster-level, and global models to study personalization and group fairness [8]. Privacy-oriented matrix-factorization variants further study lightweight protection mechanisms [14,21]. These methods improve adaptation to heterogeneous users, but their principal design object is the parameter or embedding partition. STABFEDVAE instead targets the stochastic latent posterior produced by a personalized variational branch.

2.2 Federated Variational Recommendation

Variational autoencoders (VAEs) represent an interaction vector by a posterior distribution and reconstruct item scores from a sampled latent variable [22]. Mult-VAE established the multinomial variational formulation for implicit feedback [23], and recent surveys summarize subsequent developments in VAE-based recommendation [4]. FedVAE transferred the variational recommender to FL through model-update exchange [24], while EFVAE reduced its federated cost through efficient importance-sampling-based optimization [12]. FedDAE introduced global and local encoders with adaptive gating for personalized federated collaborative filtering [17]. STABFEDVAE retains the useful global–local decomposition but changes the fusion and regularization roles: the global sample forms a bounded multiplicative gate, and only the local posterior receives the stabilizing KL term.

2.3 Personalization and Representation Stability

Personalized FL commonly uses private layers, client-specific heads, clustering, or regularization toward a global model, as synthesized in recent surveys [6,25]. FedBN handles feature-shift heterogeneity through local batch-normalization statistics, while pFedMe formulates personalization through a Moreau-envelope objective and parameter-space local models [26,27]. More recent representation-level methods align prototypes or common representations across heterogeneous clients. FedProto exchanges class prototypes [18], whereas FedCR learns a common representation with conditional mutual-information regularization [19]. These methods demonstrate the value of parameter- or representation-level control, but they do not study the scale–collapse trade-off of a private variational posterior in federated recommendation. STABFEDVAE addresses this gap with selective local prior regularization and evaluates it jointly through ranking utility and posterior dispersion. The contraction result is representational; it is not a convergence guarantee for FedBN, pFedMe, or the implemented federated optimizer.

3 Problem Formulation

Let $\mathcal{U} = \{1, \dots, N\}$ denote the client set and let $\mathcal{I} = \{1, \dots, M\}$ denote the item set. Client u stores a binary interaction vector $\mathbf{x}_u \in \{0, 1\}^M$. Raw vectors remain local. The server coordinates a sequence of communication rounds. In round t , a participating set $\mathcal{S}_t \subseteq \mathcal{U}$ performs local optimization and contributes updates to shared parameters.

For implicit-feedback recommendation, the model should rank the held-out positive item above unobserved candidates. For a local batch of B interaction vectors, we use the multinomial negative log-likelihood

$$\mathcal{L}_{\text{rec}}(\mathbf{x}, \widehat{\mathbf{x}}) = -\frac{1}{B} \sum_{b=1}^B \sum_{i=1}^M x_{bi} \log \text{softmax}(\widehat{\mathbf{x}}_b)_i, \quad (1)$$

where $\widehat{\mathbf{x}}_b \in \mathbb{R}^M$ contains the item logits and B is the local batch size; $B = 1$ in the reported client-level experiments. Evaluation masks the training interactions and computes Hit Rate at 20 (HR@20) and Normalized Discounted Cumulative Gain at 20 (NDCG@20) from held-out positive items.

The parameters are partitioned at round t into the shared state $\Theta_s^t = \{\phi_g^t, \theta^t\}$ and the client-private state $\Theta_u^t = \{\phi_{l,u}^t\}$. Here, ϕ_g^t , $\phi_{l,u}^t$, and θ^t parameterize the global encoder, client- u local encoder, and decoder, respectively. Partial aggregation collects only the gradients of ϕ_g^t and θ^t ; $\phi_{l,u}^t$ remains on client u and is restored when that client participates again.

4 The StabFedVAE Framework

Fig. 1 summarizes the shared and private model components, and Algorithm 1 details their federated optimization.

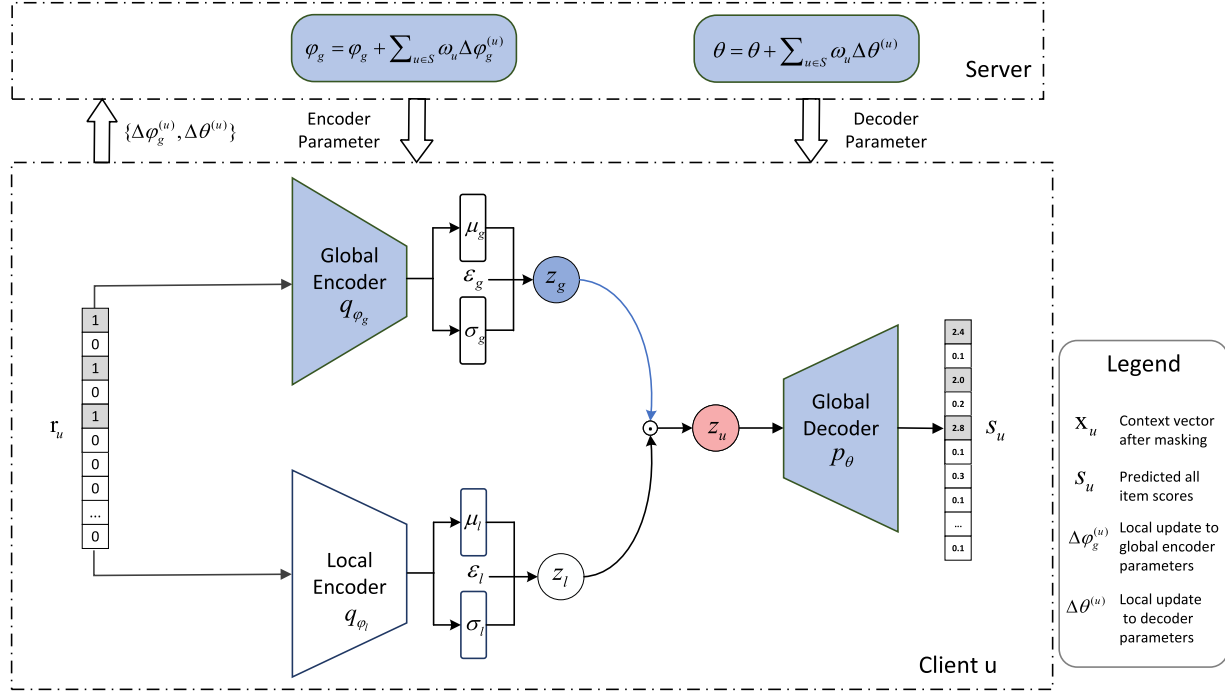


Figure 1: The StabFedVAE framework. The global encoder q_g and local encoder q_l produce latent representations fused by the parameter-free operation $\mathbf{z}_l \odot \text{sigmoid}(\mathbf{z}_g)$. Only the local posterior receives a scheduled KL penalty. The server broadcasts and updates the shared global encoder and decoder; each private local encoder persists on its client and is never uploaded.

4.1 Dual Global-Local Encoder

Given an interaction vector $\mathbf{x}_u$, the global encoder produces

$$q_{\phi_g}(\mathbf{z}_{g,u} | \mathbf{x}_u) = \mathcal{N}(\boldsymbol{\mu}_{g,u}, \text{diag}(\boldsymbol{\sigma}_{g,u}^2)), \quad (2)$$

and the local encoder produces

$$q_{\phi_l}(\mathbf{z}_{l,u} | \mathbf{x}_u) = \mathcal{N}(\boldsymbol{\mu}_{l,u}, \text{diag}(\boldsymbol{\sigma}_{l,u}^2)). \quad (3)$$

Both encoders have latent dimension d . A sample from branch $r \in \{g, l\}$ is obtained as $\mathbf{z}_{r,u} = \boldsymbol{\mu}_{r,u} + \boldsymbol{\sigma}_{r,u} \odot \boldsymbol{\epsilon}_{r,u}$, where $\boldsymbol{\epsilon}_{r,u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The global encoder is updated through the federated shared branch. The local encoder is restored from the client-specific state whenever the client participates again, preserving local adaptation across rounds. Both encoders have latent dimension d . In the reported configuration, each encoder maps $M \rightarrow 768 \rightarrow 512 \rightarrow 512 \rightarrow 2d$ with Tanh after the hidden linear layers; the final $2d$ coordinates are split into $d = 256$ -dimensional $\boldsymbol{\mu}$ and $\log \boldsymbol{\sigma}^2$. Dropout with probability 0.5 is applied before the multilayer perceptron, and no custom initializer is used, so the framework defaults for linear layers apply. A sample from branch $r \in \{g, l\}$ is obtained as $\mathbf{z}_{r,u} = \boldsymbol{\mu}_{r,u} + \boldsymbol{\sigma}_{r,u} \odot \boldsymbol{\epsilon}_{r,u}$, where $\boldsymbol{\epsilon}_{r,u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The global encoder is updated through the federated shared branch. The local encoder is restored from the client-specific state whenever the client participates again, preserving local adaptation across rounds.

Algorithm 1: Federated training of StabFedVAE

Require: client histories $\{\mathbf{x}_u\}_{u=1}^N$, rounds T , local epochs E , client ratio ρ , sampling rule, target β_l^* , schedule length T_w

```

1: Initialize shared encoder  $\phi_g^0$ , decoder  $\theta^0$ , and private states  $\{\phi_{l,u}^0\}$ 
2: for  $t = 0, \dots, T - 1$  do
3:   Select  $\lceil \rho N \rceil$  clients as  $\mathcal{S}_t$  and broadcast  $(\phi_g^t, \theta^t)$ 
4:   Set  $\beta_l(t)$  according to Eq. (8)
5:   for all  $u \in \mathcal{S}_t$  in parallel do
6:      $(\mathbf{G}_{\phi_g}^{(u)}, \mathbf{G}_{\theta}^{(u)}, \phi_{l,u}^{t+1}) \leftarrow \text{CLIENTUPDATE}(u, \phi_g^t, \theta^t, \phi_{l,u}^t, \beta_l(t), E)$ 
7:   end for
8:   Preserve  $\phi_{l,u}^{t+1} \leftarrow \phi_{l,u}^t$  for every  $u \notin \mathcal{S}_t$ 
9:   Sum local-epoch gradients, divide by  $|\mathcal{S}_t|$ , and apply one server Adam step to obtain  $(\phi_g^{t+1}, \theta^{t+1})$ 
10: end for
11: return  $\phi_g^T, \theta^T, \{\phi_{l,u}^T\}_{u=1}^N$ 
12: procedure CLIENTUPDATE( $u, \phi_g, \theta, \phi_{l,u}, \beta_l, E$ )
13:   Restore  $\phi_{l,u}$  and initialize local Adam optimizers
14:   for  $e = 1, \dots, E$  do
15:     Sample  $\mathbf{z}_{g,u} \sim q_{\phi_g}(\cdot | \mathbf{x}_u)$  and  $\mathbf{z}_{l,u} \sim q_{\phi_{l,u}}(\cdot | \mathbf{x}_u)$ 
16:      $\mathbf{z}_u \leftarrow \mathbf{z}_{l,u} \odot \text{sigmoid}(\mathbf{z}_{g,u})$ ; decode  $\widehat{\mathbf{x}}_u \leftarrow f_{\theta}(\mathbf{z}_u)$ 
17:     Evaluate Eq. (7) with coefficient  $\beta_l$ ; backpropagate
18:     Accumulate gradients for  $\phi_g$  and  $\theta$ ; apply one client Adam step to the local working model
19:   end for
20:   Persist  $\phi_{l,u}$  on client  $u$  and return  $(\mathbf{G}_{\phi_g}^{(u)}, \mathbf{G}_{\theta}^{(u)}, \phi_{l,u})$ 
21: end procedure

```

4.2 Gated Multiplicative Fusion

The fused representation is

$$\mathbf{z}_u = \mathbf{z}_{l,u} \odot \text{sigmoid}(\mathbf{z}_{g,u}), \quad (4)$$

where $\odot$ denotes element-wise multiplication. This fusion is parameter-free: the sigmoid term bounds the global modulation between zero and one without introducing a separate gate network. Thus the global branch acts as a dimension-wise gate on local evidence instead of being concatenated with or directly added to the local vector. The decoder maps the fused vector to item scores:

$$\widehat{\mathbf{x}}_u = f_{\theta}(\mathbf{z}_u). \quad (5)$$

The gate is an optimization mechanism that controls the contribution of local dimensions, not a calibrated probability.

4.3 Selective Local KL Stabilization

The standard diagonal-Gaussian KL divergence to a unit normal is

$$\mathcal{K}_l(\mathbf{x}_u) = -\frac{1}{2} \sum_{j=1}^d (1 + \log \sigma_{l,u,j}^2 - \mu_{l,u,j}^2 - \sigma_{l,u,j}^2). \quad (6)$$

STABFEDVAE minimizes

$$\mathcal{L}_u = \mathcal{L}_{\text{rec}}(\mathbf{x}_u, \widehat{\mathbf{x}}_u) + \beta_l(t) \mathcal{K}_l(\mathbf{x}_u), \quad (7)$$

where the KL penalty is applied only to the local posterior. The global posterior is not included in Eq. (7). This asymmetric treatment reflects the distinct roles of the two branches: the local posterior requires an explicit scale-control mechanism, whereas the global posterior must retain sufficient capacity for transferable preference structure.

The coefficient schedule is defined as:

$$\beta_l(t) = \begin{cases} \beta_l^*, & t = 0, \\ \beta_l^* \min(1, t/T_w), & t \geq 1, \end{cases} \quad (8)$$

where β_l^* is the target coefficient, $T_w = 20$, and t is zero-indexed. The first loss uses the target value because β_l is initialized before the first round-end update; the coefficient is then set to β_l^*/T_w and ramps through $t = 1, \dots, T_w$. Thus the schedule applies the full target coefficient at the first round and then ramps up from β_l^*/T_w , reaching the target on the 20th ramp update. This full-target initialization is retained to match the FedDAE baseline, ensuring that both methods are subject to the same early-prior pressure. Since all reported checkpoints are selected by validation NDCG@20 in late rounds, this early-stage behavior does not affect the reported values.

4.4 Federated Optimization

At each round, the server selects clients and broadcasts the shared state. Each selected client restores its local encoder, performs $E = 5$ local epochs, and accumulates the shared-branch gradients from those local updates. Under partial aggregation, only gradients of the global encoder and decoder are retained; the local encoder state is never uploaded. The server sums the returned shared gradients, divides by $|S_t|$, and applies one Adam update to the shared model. Clients not selected in a round retain their previous private state.

Client Participation Protocol

The experiments use dataset-specific participation: $\rho = 1.0$ for Amazon Video and ML-100K, and $\rho = 0.1$ for ML-1M and QB-Article. Subsets are sampled randomly with a no-immediate-repeat rule when $\rho < 1$. The results cover both full and partial participation; a controlled sweep over $\rho \in \{0.1, 0.3, 0.5, 1.0\}$ on ML-100K and Video is reported in the component study (Section 6.5). The completed FedDAE controls use the same dataset-specific ratios and nominal 100-round budget, which gives a matched participation/round-budget comparison; it does not establish parity of computation, memory, wall-clock time, or communication bytes.

4.5 Gating Stability Bound

The stochastic fusion in Eq. (4) admits a pointwise bound for every sampled gate. For the deterministic diagnostics, define $\mathbf{g}_u = \text{sigmoid}(\boldsymbol{\mu}_{g,u})$. The ideal deterministic-gate variance proxy is $\sigma_{f,u}^2 = \sigma_{l,u}^2 \odot \mathbf{g}_u^{\odot 2}$. The implementation adds $\varepsilon_0 = 10^{-8}$ inside the logarithm and therefore computes $\sigma_{f,u,\varepsilon}^2 = \sigma_{l,u}^2 \odot (\mathbf{g}_u + \varepsilon_0)^{\odot 2}$. This proxy is not the exact marginal variance of the product of two random variables.

Theorem 1 (Local-channel contraction under bounded gating): *Let $F(\mathbf{z}_l, \mathbf{z}_g) = \mathbf{z}_l \odot \text{sigmoid}(\mathbf{z}_g)$ for $\mathbf{z}_l, \mathbf{z}_g \in \mathbb{R}^d$. For any fixed $\mathbf{z}_g$ and every $p \in [1, \infty]$,*

$$\|F(\mathbf{z}_l, \mathbf{z}_g)\|_p \leq \|\mathbf{z}_l\|_p, \quad \left\| \frac{\partial F}{\partial \mathbf{z}_l} \right\|_2 = \|\text{sigmoid}(\mathbf{z}_g)\|_\infty < 1. \quad (9)$$

Consequently, $\mathbb{E}\|\mathbf{F}(\mathbf{Z}_l, \mathbf{Z}_g)\|_2^2 \leq \mathbb{E}\|\mathbf{Z}_l\|_2^2$ whenever the latter expectation is finite. For any nonempty client set $\mathcal{D}$, the deterministic variance diagnostic also satisfies

$$\bar{v}_f = \frac{1}{|\mathcal{D}|d} \sum_{u \in \mathcal{D}} \sum_{j=1}^d g_{u,j}^2 \sigma_{l,u,j}^2 \leq \bar{v}_l = \frac{1}{|\mathcal{D}|d} \sum_{u \in \mathcal{D}} \sum_{j=1}^d \sigma_{l,u,j}^2. \quad (10)$$

With the numerical safeguard, the corresponding implementation quantity obeys $\bar{v}_{f,\varepsilon} \leq (1 + \varepsilon_0)^2 \bar{v}_l$.

Proof: Every sigmoid coordinate lies strictly between zero and one, so $|F_j| \leq |z_{l,j}|$ coordinate-wise. Summation for finite p , or maximization for $p = \infty$, proves the norm inequality. The Jacobian with respect to $\mathbf{z}_l$ is $\text{diag}(\text{sigmoid}(\mathbf{z}_g))$, whose spectral norm is its largest diagonal entry. Squaring the pointwise ℓ_2 inequality and taking expectations gives the energy bound. Finally, $g_{u,j}^2 \sigma_{l,u,j}^2 \leq \sigma_{l,u,j}^2$ for every client and coordinate; averaging proves Eq. (10). Replacing $g_{u,j}$ by $g_{u,j} + \varepsilon_0 \leq 1 + \varepsilon_0$ proves the safeguarded bound. $\square$

The contraction is stated only with respect to the local input. For the global input, let $\mathbf{g} = \text{sigmoid}(\mathbf{z}_g)$; then

$$\left\| \frac{\partial \mathbf{F}}{\partial \mathbf{z}_g} \right\|_2 = \|\mathbf{z}_l \odot \mathbf{g} \odot (\mathbf{1} - \mathbf{g})\|_\infty \leq \frac{1}{4} \|\mathbf{z}_l\|_\infty, \quad (11)$$

so the joint map is not globally non-expansive unless the local code is bounded. This distinction prevents the representation-level result from being interpreted as a global optimization guarantee.

Theorem 1 is a representation-level property. It neither establishes convergence of federated optimization nor implies that increasing β_l^* must monotonically reduce posterior dispersion; the reconstruction term, client updates, and shared optimization jointly determine the learned posterior.

5 Latent Stability Diagnostics

Recommendation metrics alone cannot distinguish useful personalization from mere scale change. We therefore introduce the following diagnostics.

Posterior Dispersion Index

For a nonempty diagnostic client set $\mathcal{D}$, define the deterministic fused code $\tilde{\boldsymbol{\mu}}_u = \boldsymbol{\mu}_{l,u} \odot \text{sigmoid}(\boldsymbol{\mu}_{g,u})$ and its client average $\bar{\boldsymbol{\mu}} = |\mathcal{D}|^{-1} \sum_{u \in \mathcal{D}} \tilde{\boldsymbol{\mu}}_u$. This code is a diagnostic, not the analytic expectation of the nonlinear fused random variable. The posterior dispersion index (PDI) is

$$\text{PDI} = \frac{1}{|\mathcal{D}|} \sum_{u \in \mathcal{D}} \|\tilde{\boldsymbol{\mu}}_u - \bar{\boldsymbol{\mu}}\|_2^2. \quad (12)$$

For datasets with more than 200 users, $\mathcal{D}$ is a fixed random sample of up to 200 clients; otherwise all clients are used. The sensitivity analysis also reports the mean local-posterior norm $\bar{m}_l = |\mathcal{D}|^{-1} \sum_{u \in \mathcal{D}} \|\boldsymbol{\mu}_{l,u}\|_2$; the corresponding mean local variance $\bar{v}_l$ is defined in Eq. (10). PDI measures dispersion, not personalization quality by itself. A large value may reflect useful client differences or unstable scale. A value close to zero may indicate stable common coordinates or local posterior collapse. It must therefore be interpreted jointly with HR@20, NDCG@20, posterior norms, and posterior variances.

6 Experimental Protocol

6.1 Datasets and Evaluation

We use four implicit-feedback datasets: ML-100K, ML-1M, Amazon Video, and Tenrec QB-Article. Following the FedDAE protocol [17], users with fewer than ten interactions are removed and ratings are binarized. We apply leave-one-out evaluation: for each user, interactions are sorted by timestamp; all but the last two form training, the penultimate is validation, and the last is test. The reconstruction loss is the multinomial negative log-likelihood (NLL) over the full item catalog (Eq. (1)), without negative sampling. After preprocessing, the model item dimensions are 1678, 3703, 7417, and 4801, respectively. The QB-Article preprocessing differs slightly from the published baseline, so that comparison is descriptive.

We report HR@20 and NDCG@20. Training interactions are masked during evaluation, and each held-out item is ranked against the full catalog. With one held-out item per user, HR@20 equals Recall@20.

6.2 Main Comparison

Table 1 compares STABFEDVAE with the four most recent federated methods reported in the FedDAE benchmark. Published baseline values are reproduced as five-run mean $\pm$ standard deviation and converted from percentages to proportions. STABFEDVAE uses the common reference target $\beta_l^* = 1.0$ on every dataset; this is a prespecified reference condition, not a claim of a universal optimum. The current STABFEDVAE entries are means and sample standard deviations over three independent seeds $\{4, 123, 2024\}$. Each run selects the earliest checkpoint attaining the maximum validation NDCG@20; test HR@20, NDCG@20, and latent diagnostics are read from that checkpoint. The test set is not used for selection.

Table 1: Overall recommendation performance in proportion units. Published baselines are five-run mean $\pm$ standard deviation values from FedDAE Table 2 [17]; StabFedVAE reports current three-seed mean $\pm$ sample standard deviation at the common reference $\beta_l^* = 1.0$. This reference coefficient is not asserted to be optimal. Boldface marks the largest displayed mean, including ties.

Dataset	Metric	FedVAE [24] (2021)	PFedRec [10] (2023)	FedRAP [11] (2024)	FedDAE [17] (2025)	StabFedVAE (Ours)
ML-100K	HR@20	0.1200 $\pm$ 0.0040	0.0541 $\pm$ 0.0054	0.0626 $\pm$ 0.0032	0.1232 $\pm$ 0.0023	0.1711 $\pm$ 0.0046
	NDCG@20	0.0501 $\pm$ 0.0032	0.0115 $\pm$ 0.0026	0.0122 $\pm$ 0.0017	0.0503 $\pm$ 0.0014	0.0727 $\pm$ 0.0023
ML-1M	HR@20	0.0639 $\pm$ 0.0012	0.0600 $\pm$ 0.0002	0.0625 $\pm$ 0.0015	0.0650 $\pm$ 0.0039	0.0697 $\pm$ 0.0008
	NDCG@20	0.0250 $\pm$ 0.0006	0.0231 $\pm$ 0.0000	0.0240 $\pm$ 0.0003	0.0261 $\pm$ 0.0011	0.0281 $\pm$ 0.0002
Video	HR@20	0.0634 $\pm$ 0.0128	0.0138 $\pm$ 0.0064	0.0142 $\pm$ 0.0097	0.0654 $\pm$ 0.0080	0.0731 $\pm$ 0.0053
	NDCG@20	0.0227 $\pm$ 0.0036	0.0026 $\pm$ 0.0012	0.0028 $\pm$ 0.0019	0.0230 $\pm$ 0.0029	0.0269 $\pm$ 0.0030
QB-Article	HR@20	0.0116 $\pm$ 0.0031	0.0065 $\pm$ 0.0002	0.0116 $\pm$ 0.0026	0.0124 $\pm$ 0.0030	0.0098 $\pm$ 0.0004
	NDCG@20	0.0045 $\pm$ 0.0013	0.0033 $\pm$ 0.0000	0.0035 $\pm$ 0.0010	0.0052 $\pm$ 0.0014	0.0052 $\pm$ 0.0002

Table 2: Implementation check of Theorem 1 at the validation-selected checkpoints. Here, $\bar{v}_{f,\epsilon}$ is the safeguarded fused variance proxy and $R_{v,\epsilon} = \bar{v}_{f,\epsilon}/\bar{v}_l$.

Quantity	ML-100K	ML-1M	Video	QB-Article
β_l^*	0.0010	0.0010	0.1000	0.0100
Round	92	100	99	99
$\bar{v}_{f,\epsilon}$	0.4860	0.5449	0.8648	0.6387
$\bar{v}_l$	0.4862	0.5878	0.8649	0.6426
$R_{v,\epsilon}$	0.9996	0.9270	0.9999	0.9939

6.3 Compared Configurations

Table 1 compares STABFEDVAE with published baselines at the common reference $\beta_l^* = 1.0$. The local-KL sweep targets $\beta_l^* \in \{0, 0.001, 0.01, 0.1, 1, 10\}$ and is reported in Section 6.6. All runs use three seeds $\{4, 123, 2024\}$, latent dimension 256, Adam (10^{-3} , $L_2 = 10^{-8}$), five local epochs, three hidden layers, dropout 0.5, batch size 2048, and partial gradient aggregation. Experiments run on NVIDIA H800 GPUs under PyTorch 2.2.2 with CUDA 12.1. Test metrics do not enter coefficient or checkpoint selection.

6.4 Log-Based Check of the Variance Bound

The proof of Theorem 1 is analytic; experiments are used only to verify its observable implementation consequence. Define $R_{v,\varepsilon} = \bar{v}_{f,\varepsilon}/\bar{v}_l$. Table 2 reconstructs this ratio from four-decimal diagnostic values at the validation-selected targets and checkpoints used in the sensitivity sweep (Section 6.6). Unlike Table 1, this check does not fix $\beta_l^* = 1.0$ across datasets.

The check is state matched: $\bar{v}_{f,\varepsilon}$ and $\bar{v}_l$ are produced in the same diagnostic pass over the same clients. Across four datasets, five target values, and 100 rounds, all 2000 diagnostic evaluations satisfy $R_{v,\varepsilon} \leq (1 + \varepsilon_0)^2$ to the recorded precision; the largest rounded ratio is 1.0000. This result verifies the implementation; it does not constitute evidence of federated optimization convergence, which is beyond the scope of Theorem 1.

6.5 Component Studies: Fusion and Client Participation

The β_l^* sweep in Section 6 varies only the local-KL coefficient while keeping the gated multiplicative fusion fixed. To isolate the contribution of the fusion design itself, we ran a component study that fixes $\beta_l = 0.1$ and varies the fusion operation while holding the backbone, latent dimension 256, five local epochs, 100 rounds, dropout 0.5, three hidden layers, and partial aggregation fixed. Table 3 presents three-seed means over $\{4, 123, 2024\}$ at the best-test checkpoints for the two datasets (ML-100K and Video) for which all five fusion conditions completed at the time of writing; the learned Multilayer Perceptron (MLP) and linear gates are shared parameters updated through the federated path, so the comparison also checks whether a learned gate adds value beyond the parameter-free sigmoid that the architecture description states.

Table 3: Fusion component study at $\beta_l = 0.1$ (three-seed means $\pm$ sample standard deviation). The parameter-free sigmoid gate is the fusion used by StabFedVAE, and bold numbers denote the best (highest mean) result among fusion variants for each dataset–metric.

Dataset	Metric	Sigmoid Gate	Concat	MLP Gate	Linear Gate	Additive
ML-100K	HR@20	0.2037 $\pm$ 0.001	0.1527 $\pm$ 0.003	0.1941 $\pm$ 0.003	0.1934 $\pm$ 0.003	0.1778 $\pm$ 0.007
	NDCG@20	0.0863 $\pm$ 0.001	0.0651 $\pm$ 0.000	0.0818 $\pm$ 0.003	0.0833 $\pm$ 0.001	0.0762 $\pm$ 0.003
Video	HR@20	0.1152 $\pm$ 0.006	0.1123 $\pm$ 0.004	0.1078 $\pm$ 0.003	0.1069 $\pm$ 0.004	0.0617 $\pm$ 0.006
	NDCG@20	0.0479 $\pm$ 0.004	0.0435 $\pm$ 0.002	0.0463 $\pm$ 0.002	0.0459 $\pm$ 0.001	0.0235 $\pm$ 0.000

Note: Sigmoid gate: $\mathbf{z}_l \odot \sigma(\mathbf{z}_g)$; Concat: $\mathbf{W}[\mathbf{z}_l; \mathbf{z}_g]$ projected to d dimensions; MLP gate: $\mathbf{z}_l \odot \text{MLP}(\mathbf{z}_g)$; Linear gate: $\mathbf{z}_l \odot \sigma(\mathbf{W}\mathbf{z}_g)$; Additive: $0.5\mathbf{z}_l + 0.5\mathbf{z}_g$ (FedDAE-style). All variants produce a d -dimensional fused vector.

On both datasets, the parameter-free sigmoid gate attains the highest mean HR@20 and NDCG@20; the additive (FedDAE-style) fusion is the weakest, and the learned MLP and linear gates do not improve over the parameter-free gate despite adding shared parameters. The default gate is therefore competitive with, and never substantially worse than, the learned and concatenation alternatives on ML-100K while being the only option that adds neither shared nor private parameters. The additive fusion, which is the FedDAE-style combination, is among the weakest on both datasets.

Table 4 instead presents a controlled client-participation sweep over $\rho \in \{0.1, 0.3, 0.5, 1.0\}$ on ML-100K and Video at $\beta_l = 0.1$. Only ρ varies across its rows; the fusion mechanism, local-KL setting, schedule, and aggregation rule remain fixed. Reducing ρ from 1.0 to 0.1 lowers mean HR@20, NDCG@20, and PDI on both datasets, although Video PDI is not monotonic at the intermediate ratios. Table 4 therefore measures sensitivity to client participation, not fusion choice or KL placement.

Table 4: Partial client participation at $\beta_l = 0.1$ (three-seed means $\pm$ sample standard deviation). PDI is the posterior dispersion index at the validation-selected checkpoint. Bold numbers denote the best (highest mean) HR@20 and NDCG@20 among the tested participation rates.

ρ	ML-100K			Video		
	HR@20	NDCG@20	PDI	HR@20	NDCG@20	PDI
1.0	0.2037 $\pm$ 0.001	0.0863 $\pm$ 0.001	59.47 $\pm$ 1.42	0.1152 $\pm$ 0.006	0.0479 $\pm$ 0.004	89.21 $\pm$ 2.16
0.5	0.1902 $\pm$ 0.001	0.0804 $\pm$ 0.001	48.11 $\pm$ 0.08	0.1142 $\pm$ 0.008	0.0464 $\pm$ 0.002	90.86 $\pm$ 3.86
0.3	0.1775 $\pm$ 0.005	0.0771 $\pm$ 0.002	35.51 $\pm$ 3.04	0.1072 $\pm$ 0.002	0.0443 $\pm$ 0.002	81.49 $\pm$ 1.95
0.1	0.1640 $\pm$ 0.003	0.0695 $\pm$ 0.002	17.35 $\pm$ 2.71	0.0840 $\pm$ 0.006	0.0325 $\pm$ 0.002	66.05 $\pm$ 3.90

6.6 Results and Sensitivity Analysis

Table 5 presents the target-coefficient sweep across six values of β_l^* . For each coefficient, the checkpoint is selected solely by the highest validation NDCG@20. The sweep subsumes the two-point local-KL contrast ($\beta_l^* = 0$ vs. 0.01) previously shown separately.

Table 5: Three-seed sensitivity to the target local Kullback–Leibler coefficient β_l^* (means $\pm$ sample standard deviation over seeds {4, 123, 2024}). PDI is the posterior dispersion index at the validation-selected checkpoint. Boldface marks the largest mean HR@20 or NDCG@20 within each dataset.

β_l^*	Metric	ML-100K	ML-1M	Video	QB-Article
0	HR@20	0.2094 $\pm$ 0.0084	0.0887 $\pm$ 0.0061	0.0999 $\pm$ 0.0036	0.0146 $\pm$ 0.0006
	NDCG@20	0.0865 $\pm$ 0.0016	0.0358 $\pm$ 0.0025	0.0422 $\pm$ 0.0036	0.0060 $\pm$ 0.0001
	PDI	58172.44 $\pm$ 7259.20	1291.03 $\pm$ 83.90	57378.10 $\pm$ 4539.41	1191.27 $\pm$ 129.94
10^{-3}	HR@20	0.2186 $\pm$ 0.0101	0.0904 $\pm$ 0.0035	0.1015 $\pm$ 0.0072	0.0141 $\pm$ 0.0007
	NDCG@20	0.0898 $\pm$ 0.0056	0.0362 $\pm$ 0.0019	0.0443 $\pm$ 0.0047	0.0060 $\pm$ 0.0002
	PDI	2170.35 $\pm$ 47.66	250.21 $\pm$ 3.44	1279.28 $\pm$ 146.62	160.20 $\pm$ 6.38
10^{-2}	HR@20	0.2235 $\pm$ 0.0142	0.0839 $\pm$ 0.0018	0.1053 $\pm$ 0.0045	0.0099 $\pm$ 0.0003
	NDCG@20	0.0921 $\pm$ 0.0043	0.0338 $\pm$ 0.0007	0.0448 $\pm$ 0.0018	0.0051 $\pm$ 0.0002
	PDI	447.76 $\pm$ 19.85	53.59 $\pm$ 1.62	325.92 $\pm$ 14.72	8.70 $\pm$ 1.95
0.1	HR@20	0.2037 $\pm$ 0.0006	0.0717 $\pm$ 0.0025	0.1152 $\pm$ 0.0062	0.0096 $\pm$ 0.0003
	NDCG@20	0.0863 $\pm$ 0.0011	0.0291 $\pm$ 0.0015	0.0479 $\pm$ 0.0040	0.0051 $\pm$ 0.0002
	PDI	59.47 $\pm$ 1.42	5.40 $\pm$ 1.85	89.21 $\pm$ 2.16	2.20 $\pm$ 0.23
1	HR@20	0.1711 $\pm$ 0.0046	0.0697 $\pm$ 0.0008	0.0731 $\pm$ 0.0053	0.0098 $\pm$ 0.0004
	NDCG@20	0.0727 $\pm$ 0.0023	0.0281 $\pm$ 0.0002	0.0269 $\pm$ 0.0030	0.0052 $\pm$ 0.0002
	PDI	5.65 $\pm$ 0.59	0.20 $\pm$ 0.09	4.26 $\pm$ 0.66	0.23 $\pm$ 0.12
10	HR@20	0.1378 $\pm$ 0.0027	0.0709 $\pm$ 0.0009	0.0712 $\pm$ 0.0031	0.0106 $\pm$ 0.0008
	NDCG@20	0.0576 $\pm$ 0.0006	0.0272 $\pm$ 0.0000	0.0252 $\pm$ 0.0014	0.0053 $\pm$ 0.0003
	PDI	0.06 $\pm$ 0.04	0.00 $\pm$ 0.00	0.16 $\pm$ 0.02	0.01 $\pm$ 0.00

Across all four datasets, increasing β_l^* consistently reduces PDI while moving the local posterior toward the unit-normal target. This stability gain is not monotonic in recommendation utility: targets $\beta_l^* \geq 1$ substantially reduce ranking quality. On ML-100K, $\beta_l^* = 0.01$ yields the highest mean HR@20 (0.2235) and NDCG@20 (0.0921); on Video, $\beta_l^* = 0.1$ is best (0.1152/0.0479). On ML-1M and QB-Article, the no-regularization condition ($\beta_l^* = 0$) performs best. These results confirm a dataset-dependent utility–stability trade-off rather than a universal optimum.

Fig. 2 reports the round-level NDCG@20 trajectories across all four datasets. On ML-100K, ML-1M, and Video, the two StabFedVAE variants generally improve beyond the FedDAE trajectory after the initial rounds; the regularized variant also shows the strongest late-round trajectory on ML-100K and Video. On QB-Article, the absolute differences are smaller and all curves remain noisier, although both StabFedVAE variants finish above FedDAE in the displayed run. Because each curve represents one run, these observations describe the plotted training dynamics rather than mean convergence or statistical superiority.

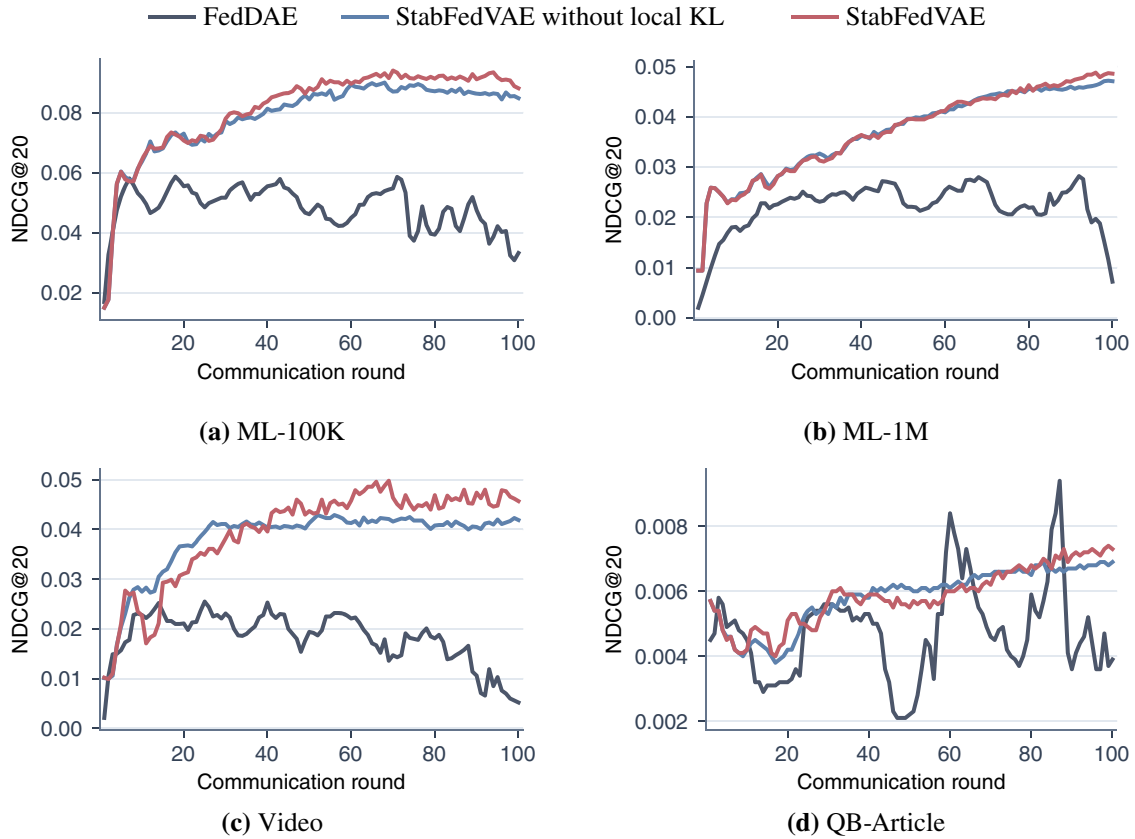


Figure 2: NDCG@20 convergence over 100 communication rounds on (a) MovieLens-100K, (b) MovieLens-1M, (c) Amazon Video, and (d) Tenrec QB-article. Both StabFedVAE and the FedDAE baseline use a matched participation/round-budget. The plotted trajectories come from one run per configuration and therefore do not represent confidence bands.

Fig. 3 complements the coefficient sweep on ML-100K. Its PDI trajectory shows rapid dispersion growth without local KL regularization, whereas the regularized StabFedVAE trajectory remains at a substantially lower scale. In the utility–stability map, the regularized point lies above and to the right of the no-local-KL point, indicating higher displayed NDCG@20 and lower displayed dispersion within the two StabFedVAE

configurations. This visual evidence is qualitative; the three-seed coefficient results in Table 5 provide the quantitative basis for the utility–stability conclusion.

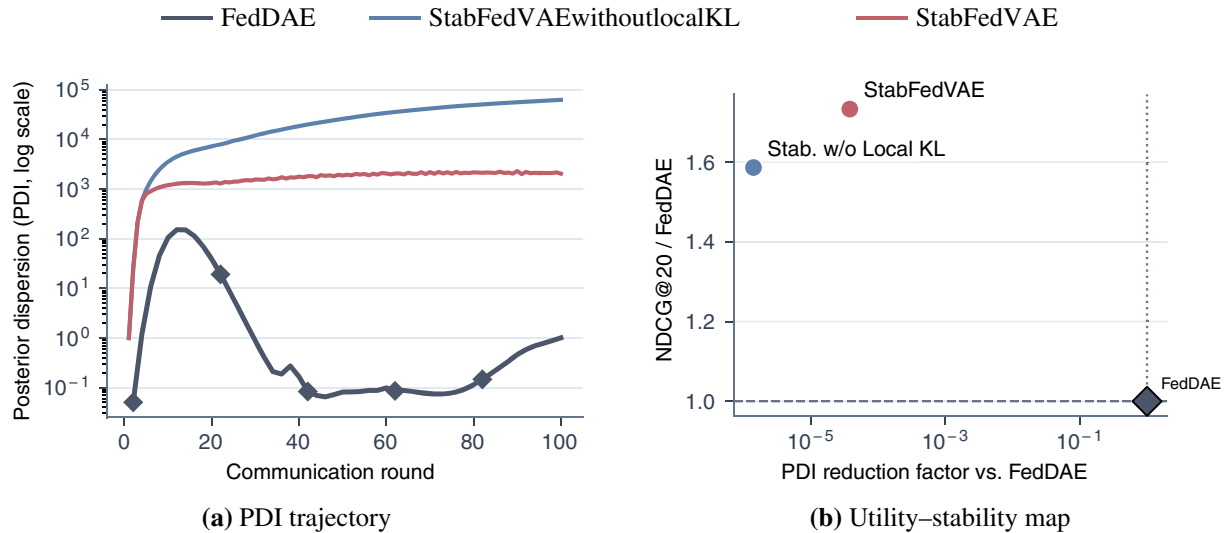


Figure 3: Utility–stability comparison on MovieLens-100K at $\beta_l = 0.1$. (a) Posterior dispersion over communication rounds on a logarithmic scale. (b) Validation-selected test NDCG@20 vs. the PDI reduction factor, both relative to the locally reproduced FedDAE run. FedDAE is explicitly anchored at (1,1); movement upward indicates higher utility and movement rightward indicates lower dispersion. Both FedDAE and StabFedVAE compute the PDI on the same diagnostic client set under the same seeds {2024, 4, 123}.

6.7 Reproducibility and Statistical Scope

The main comparison aggregates repeated STABFEDVAE runs, with each run selecting its checkpoint by validation NDCG@20 over 100 communication rounds; the reported HR@20/NDCG@20 values are read from the same selected checkpoint. The main comparisons, local-KL contrast, sensitivity sweep, and component study are reported as three-seed means over seeds {4, 123, 2024}. The convergence curves remain single-run for legibility and are labeled as such; they support descriptive mechanism and utility–stability observations but are not used for the quantitative comparisons, whose component differences with three independent seeds still have small sample sizes and do not establish formal statistical significance at conventional thresholds. Additional seeds would further strengthen the uncertainty assessment.

In each training record, the displayed $\beta_l(t)$ is the value after the round-end update, whereas the loss uses the value prescribed for that round by Eq. (8). Consequently, the displayed coefficient and the same line’s KLD cannot reconstruct the first 20 training losses.

7 Discussion

STABFEDVAE is based on a controlled separation of roles. The local encoder is allowed to retain client-specific information, but its posterior is prevented from becoming arbitrarily unconstrained by a moderate prior penalty. The global branch remains responsible for shared transfer and is updated through the federated path. The gate provides a low-cost interaction between the two branches.

The experiments further demonstrate that stability and collapse are distinct phenomena. A lower PDI is not automatically better, and a higher PDI is not automatically evidence of personalization. The useful regime is one in which recommendation quality remains high, local posterior statistics stay bounded, and client

dispersion remains nonzero. This motivates reporting utility metrics together with PDI, posterior norms, and posterior variances.

This study has the following limitations. First, the convergence curves are single-run for legibility, and the ablation and sensitivity analyses use three seeds but therefore still have small sample sizes; these component comparisons are descriptive and do not establish statistical significance at conventional thresholds. Second, the coefficient schedule applies the full target β_l^* at the first round, matching the FedDAE baseline's early-prior treatment; this is an intentional fairness choice and, because all reported checkpoints are late-round, it does not change the reported values. Third, the component study covers ML-100K and Video; the remaining datasets use dataset-specific participation and are not part of the controlled ρ sweep. Fourth, broader fusion and aggregation ablations are needed to isolate the contribution of each design choice.

8 Conclusion

This paper presented STABFEDVAE, a dual-encoder federated VAE with gated multiplicative fusion, selective local KL regularization, a round-dependent coefficient schedule, and persistent client-private encoders. We established a local-channel contraction bound and verified its same-state deterministic fused-variance proxy consequence experimentally; this proxy is not the exact marginal variance of the stochastic product, consistent with the main-text precision. In the main comparison, repeated runs at $\beta_l^* = 1.0$ are summarized by empirical means and standard deviations. The three-seed ablation associates moderate local regularization with higher validation-selected NDCG point estimates than the no-local-KL condition; excessive regularization contracts the local branch and reduces ranking accuracy. Thus, a useful regularization regime must control posterior scale while preserving client-specific variation. Repeated runs for the component analyses remain necessary for statistical comparison.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization, Jincheng Cai and Li Feng; methodology, Jincheng Cai; software, Jincheng Cai; validation, Jincheng Cai and Ni Zhao; formal analysis, Jincheng Cai; investigation, Jincheng Cai; resources, Li Feng and Ni Zhao; data curation, Jincheng Cai; writing—original draft preparation, Jincheng Cai; writing—review and editing, Li Feng and Ni Zhao; supervision, Li Feng; project administration, Li Feng. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets used in this study are publicly available: MovieLens-100K and MovieLens-1M are available from GroupLens (<https://grouplens.org/datasets/movielens/>); Amazon Instant Video is available from the UCSD Amazon review repository (https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/); and Tenrec QB-Article is available from the Tenrec benchmark repository (<https://github.com/yuangh-x/2022-NIPS-Tenrec>). The processed data and implementation details are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ko H, Lee S, Park Y, Choi A. A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics*. 2022;11(1):141. doi:10.3390/electronics11010141.
2. Zhang P, Huang W, Chen Y, Zhou MC, Al-Turki Y. A novel deep-learning-based QoS prediction model for service recommendation utilizing multi-stage multi-scale feature fusion with individual evaluations. *IEEE Trans Autom Sci Eng*. 2024;21(2):1740–53. doi:10.1109/tase.2023.3244184.
3. Zhao Q, Wu P, Liu G, An D, Lian J, Zhou MC. Sociological-theory-based multitopic self-supervised recommendation. *IEEE Trans Neural Netw Learn Syst*. 2025;36(6):11228–42. doi:10.1109/tnnls.2024.3477720.
4. Liang S, Pan Z, Liu, Yin J, Rijke M. A survey on variational autoencoders in recommender systems. *ACM Comput Surv*. 2024;56(10):1–40. doi:10.1145/3663364.
5. McMahan B, Moore E, Ramage D, Hampson S, Arcas BAY. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*; 2017 Apr 20–22; Fort Lauderdale, FL, USA. p. 1273–82.
6. Kairouz P, McMahan HB. Advances and open problems in federated learning. *Found Trends Mach Learn*. 2021;14(1–2):1–210. doi:10.1561/22000000083.
7. Wang L, Zhou H, Bao Y, Yan X, Shen G, Kong X. Horizontal federated recommender system: a survey. *ACM Comput Surv*. 2024;56(9):240. doi:10.1145/3656165.
8. Wang S, Tao H, Li J, Ji X, Gao Y, Gong M. Towards fair and personalized federated recommendation. *Pattern Recognit*. 2024;149(5):110234. doi:10.1016/j.patcog.2023.110234.
9. Perifanis V, Efrimidis PS. Federated neural collaborative filtering. *Knowledge-Based Syst*. 2022;242(3):108441. doi:10.1016/j.knsys.2022.108441.
10. Zhang C, Long G, Zhou T, Yan P, Zhang Z, Zhang C, et al. Dual personalization on federated recommendation. In: *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*; 2023 Aug 19–25; Macau, China. p. 4558–66.
11. Li Z, Long G, Zhou T. Federated recommendation with additive personalization. In: *Proceedings of the International Conference on Learning Representations (ICLR)*; 2024 May 7–11; Vienna, Austria. p. 11770–87.
12. Zhang L, Rong Q, Ding X, Li G, Yuan L. EFVAE: efficient federated Variational autoencoder for collaborative filtering. In: *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM)*; 2024 Oct 21–25; Boise ID USA. p. 3176–85.
13. Yuan W, Qu L, Cui L, Tong Y, Zhou X, Yin H. HeteFedRec: federated recommender systems with model heterogeneity. In: *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE)*; 2024 May 13–16; Utrecht, Netherlands. p. 1324–37.
14. Zhang H, Zhou X, Shen Z, Li Y. PrivFR: privacy-enhanced federated recommendation with shared hash embedding. *IEEE Trans Neural Netw Learn Syst*. 2025;36(1):32–46. doi:10.1109/tnnls.2024.3387757.
15. Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: challenges, methods, and future directions. *IEEE Signal Process Mag*. 2020;37(3):50–60. doi:10.1109/msp.2020.2975749.
16. Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. *Computer*. 2009;42(8):30–7. doi:10.1109/mc.2009.263.
17. Li Z, Long G, Zhou T, Jiang J, Zhang C. Personalized federated collaborative filtering: a Variational AutoEncoder approach. *Proc AAAI Conf Artif Intell*. 2025;39(17):18602–10.
18. Tan Y, Long G, Liu LU, Zhou T, Lu Q, Jiang J, et al. FedProto: federated prototype learning across heterogeneous clients. *Proc AAAI Conf Artif Intell*. 2022;36(8):8432–40. doi:10.1609/aaai.v36i8.20819.
19. Zhang H, Li C, Dai W, Zou J, Xiong H. FedCR: personalized federated learning based on across-client common representation with conditional mutual information regularization. In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*; 2023 Jul 23–29; Honolulu, HI, USA. p. 41314–30.
20. Chai D, Wang L, Chen K, Yang Q. Secure federated matrix Factorization. *IEEE Intell Syst*. 2021;36(5):11–20. doi:10.1109/mis.2020.3014880.
21. Zhang H, Luo F, Wu J, He X, Li Y. LightFR: lightweight federated recommendation with privacy-preserving matrix Factorization. *ACM Trans Inf Syst*. 2023;41(4):90. doi:10.1145/3578361.

22. Kingma DP, Welling M. Auto-encoding variational Bayes. In: Proceedings of the 2nd International Conference on Learning Representations; 2014 Apr 14–16; Banff, AB, Canada.
23. Liang D, Krishnan RG, Hoffman MD, Jebara T. Variational autoencoders for collaborative filtering. In: Proceedings of the 2018 World Wide Web Conference (WWW); 2018 Apr 23–27; Lyon, France. p. 689–98.
24. Polato M. Federated Variational autoencoder for collaborative filtering. In: Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN); 2021 Jul 18–22; Shenzhen, China. p. 1–8.
25. Tan AZ, Yu H, Cui L, Yang Q. Towards personalized federated learning. *IEEE Trans Neural Netw Learn Syst*. 2023;34(12):9587–603. doi:10.1109/tnnls.2022.3190359.
26. Li X, Jiang M, Zhang X, Kamp M, Dou Q. FedBN: federated learning on non-IID features via local batch normalization. In: Proceedings of the International Conference on Learning Representations (ICLR); 2021 May 3–7; Virtual.
27. Dinh CT, Tran NH, Nguyen TD. Personalized federated learning with Moreau envelopes. In: Proceedings of the 34th International Conference on Neural Information Processing Systems; 2020 Dec 6–12; Vancouver, BC, Canada. p. 21394–405.