Appendix

A. Dynamic Adjustment Rule for the Mixing Parameter

Lemma 1. *Based on the control conditions outlined in Definition1, the dynamic adjustment rule for the mixing parameter α_i is defined as:*

$$\alpha_i(t+1) = \alpha_i(t) - \gamma \cdot \frac{dJ_{joint}(\alpha_i)}{d\alpha_i} \tag{3}$$

Here, γ is the learning rate that controls the step size for adjusting α_i. A larger γ leads to faster adjustment but may cause instability, while a smaller γ results in slower but more stable adjustments. Using this rule, the system can iteratively adjust the mixing parameter α_i to eventually find the balance point between the global and personalized models, minimizing the joint objective function $J_{\text{joint}}(\alpha_i)$.

B. Stopping Criterion

Definition 3 (Stopping Condition Control under Normal Convergence)**.** *To ensure that the training process stops at the appropriate time, we introduce a stopping criterion based on the stability of the joint objective function $J_{joint}(\alpha_i)$ and the rate of change in the losses of the global and personalized models. The stopping criterion is designed to ensure that the training process halts when the system has likely reached an optimal balance between the global and personalized models. The criterion involves three conditions: If*

$$\left| \frac{dJ_{joint}(\alpha_i)}{d\alpha_i} \right| \leq \zeta \text{ and } |\Delta J_{global}| \leq \epsilon \text{ and } |\Delta J_{local}| \leq \epsilon \tag{4}$$

then stop training.

$\frac{dJ_{\text{joint}}(\alpha_i)}{d\alpha_i}$ represents the derivative of the joint objective function with respect to the mixing parameter α_i. It indicates how sensitive the joint objective function is to changes in α_i. When this derivative is small, it suggests that further adjustments to α_i will have little effect on improving the balance between the global and personalized models. That is, the balancing condition of both models is reached.

ζ is a threshold parameter that determines when the derivative $\frac{dJ_{\text{joint}}(\alpha_i)}{d\alpha_i}$ is small enough to consider stopping the training. A smaller value of ζ implies that we require a very stable joint objective function before stopping, while a larger ζ allows for stopping the training even if there is still some potential for improvement.

ΔJ_{global} represents the rate of change in the loss of the global model, $J_{\text{global}}(g^*)$. It measures how much the global model's performance is improving or worsening during the training process. A smaller value of $|\Delta J_{\text{global}}|$ indicates that the global model's loss has stabilized, suggesting that further training may not yield significant improvements.

ΔJ_{local} represents the rate of change in the loss of the personalized model, $J_{\text{local}}(l_i^*)$. Similar to ΔJ_{global}, a small $|\Delta J_{\text{local}}|$ value indicates that the personalized model's loss has stabilized.

ϵ is a threshold parameter that determines when the rates of change in the global and personalized model losses, $|\Delta J_{\text{global}}|$ and $|\Delta J_{\text{local}}|$, are small enough to consider stopping the training. A smaller value of ϵ implies that we require the losses to be very stable before stopping. This is expected to be a human-controlled parameter.

Remark 3 (Dynamic Adjustment)**.** *We adopt this strategy employing a moving average smoothing approach to update the threshold $\zeta(t)$ in a dynamic way, which adjusts over time based on the observed changes in the global model loss:*

$$\zeta(t) = \kappa\zeta(t-1) + (1-\kappa)|\Delta J_{global}(t)| \tag{5}$$

$\zeta(t)$ is the dynamic threshold at iteration t, which is adjusted according to the moving average of recent changes in the global model loss. κ is the smoothing coefficient, that controls how much weight is given to past values of the threshold versus the current change in loss. This allows us to adjust the stopping criteria during the training process. By smoothing the threshold over time, the method can respond to fluctuating loss values, ensuring that temporary spikes or drops do not stop training prematurely. It is useful in cases where the loss function may exhibit volatility, allowing the training process to continue until a more stable convergence is detected.

Assumption 1 (Threshold-based Interrupt Condition)**.** *We analyze the threshold-based condition introduced in the assumption for terminating training before the final stopping criteria are met. The condition is designed to handle situations where continued training might lead to diminishing returns or potential overfitting.* ***Relative Change Monitoring****, to monitors the change rate of the global model loss $J_{global}(t)$ over consecutive iterations. The criterion for early stopping is based on the relative change of the current loss compared to the previous loss change rate:*

$$\left| \frac{\Delta J_{global}(t)}{\Delta J_{global}(t-1)} \right| \leq \theta_r \tag{6}$$

Here, $\Delta J_{\text{global}}(t)$ represents the change in the global model loss at iteration t. θ_r is the relative change threshold, typically set to a small value. When the relative change in the loss is below this threshold, it indicates that the training process is yielding increasingly smaller improvements, and thus may no longer be worth continuing. By comparing the rate of change across iterations, the method ensures that the system does not waste computational resources on minimal global model gains.