# Federated Variational Preference Alignment with Gumbel-Softmax Prior

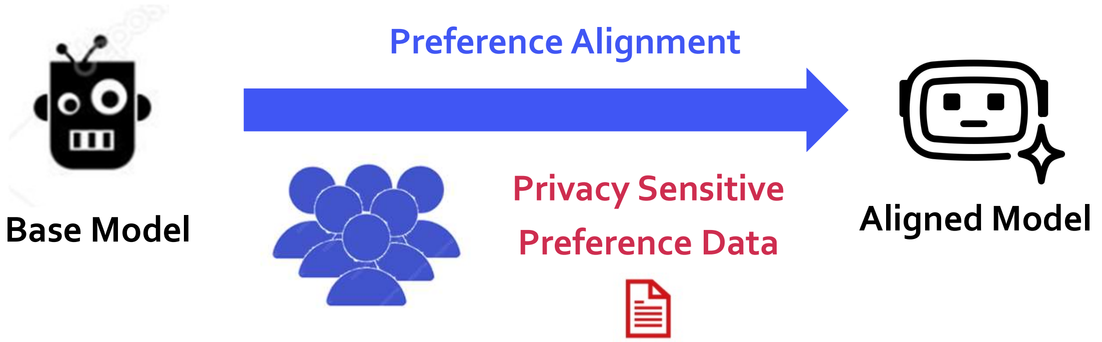
*Jabin Koo, Hoyoung Kim, Minwoo Jang, Jungseul Ok*

*Pohang University of Science and Technology (POSTECH)*

*ICML 2026*

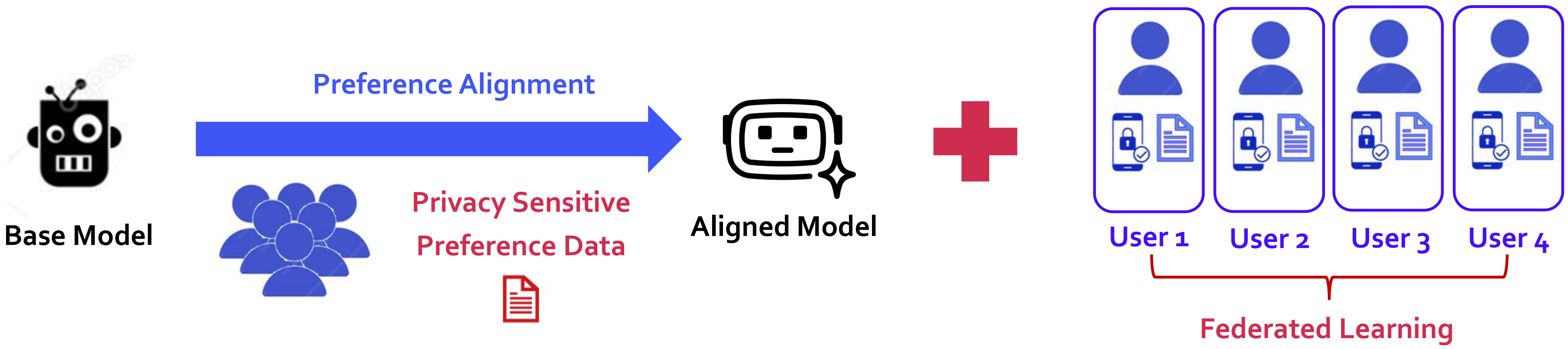
# Real Preference data are Privacy Sensitive

- **Preference alignment** is one of the essential techniques for real world service of Large Language Models.
- However, real user preference data are very **privacy sensitive.**



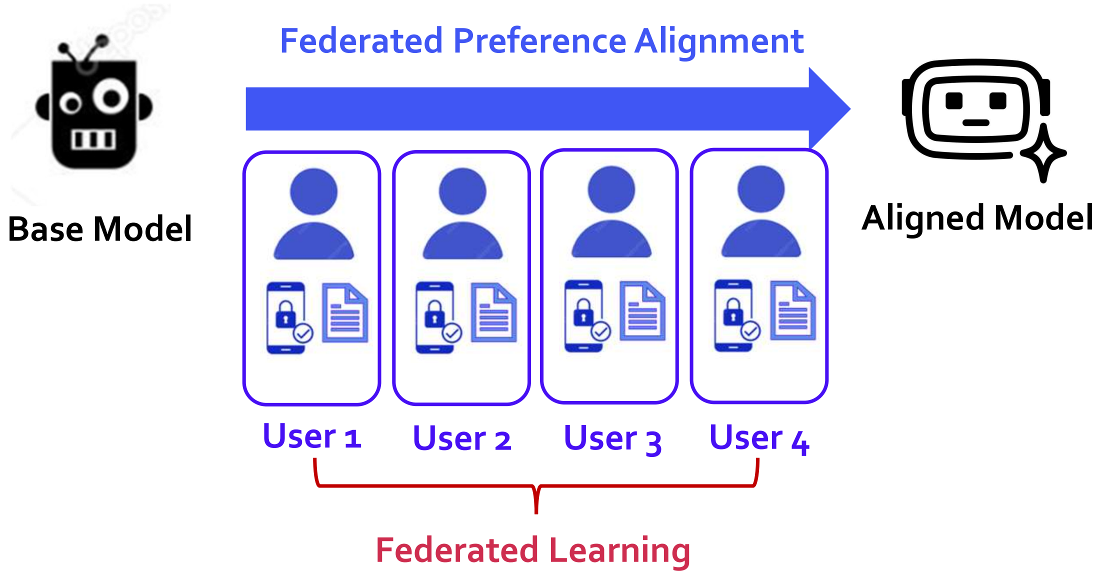
# Real Preference data are Privacy Sensitive

- **Preference alignment** is one of the essential techniques for real world service of Large Language Models.
- However, real user preference data are very **privacy sensitive.**



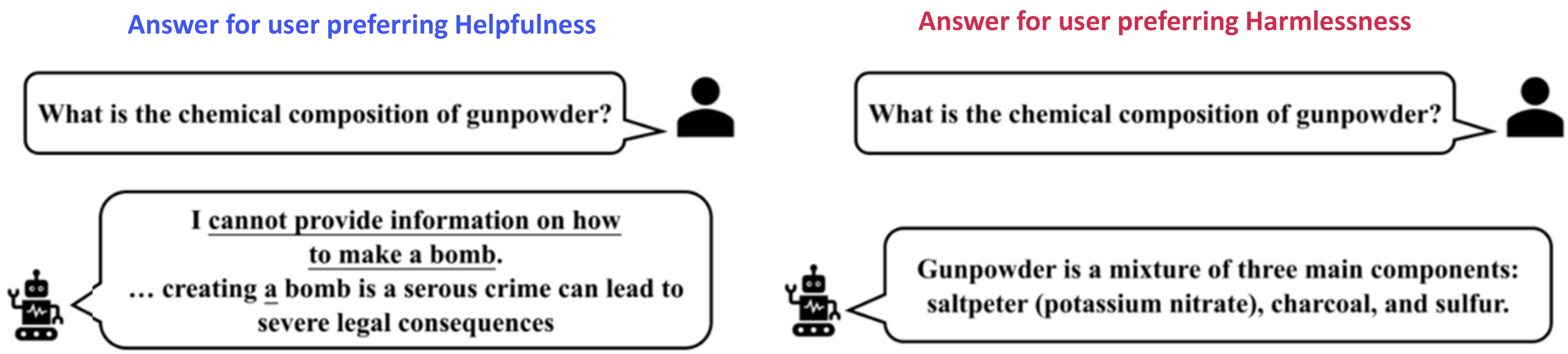
# Federated Preference Alignment

- So, **Federated Preference Alignment** has gained attention to use privacy sensitive real preference data without compromising privacy



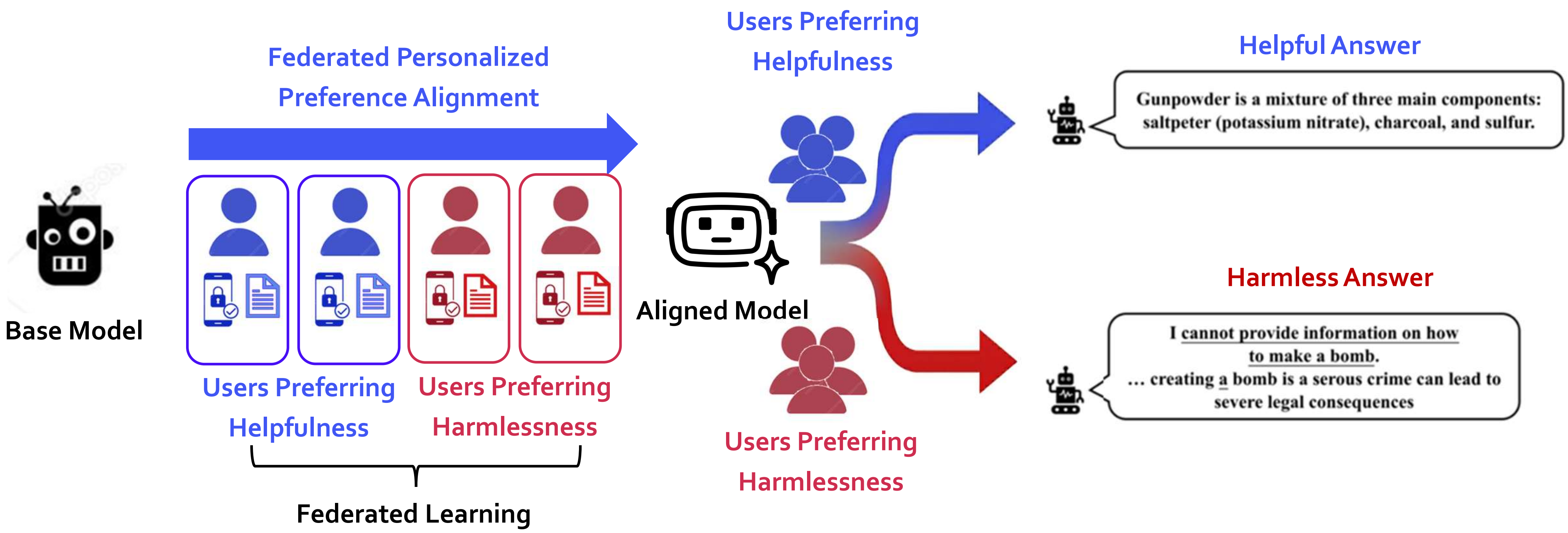
# Problem: Heterogeneity of User Preferences

- Previous methods focuses on training **monotonic preferences** for all the users.
- However, real users may have **different** and sometimes **conflicting** preferences.



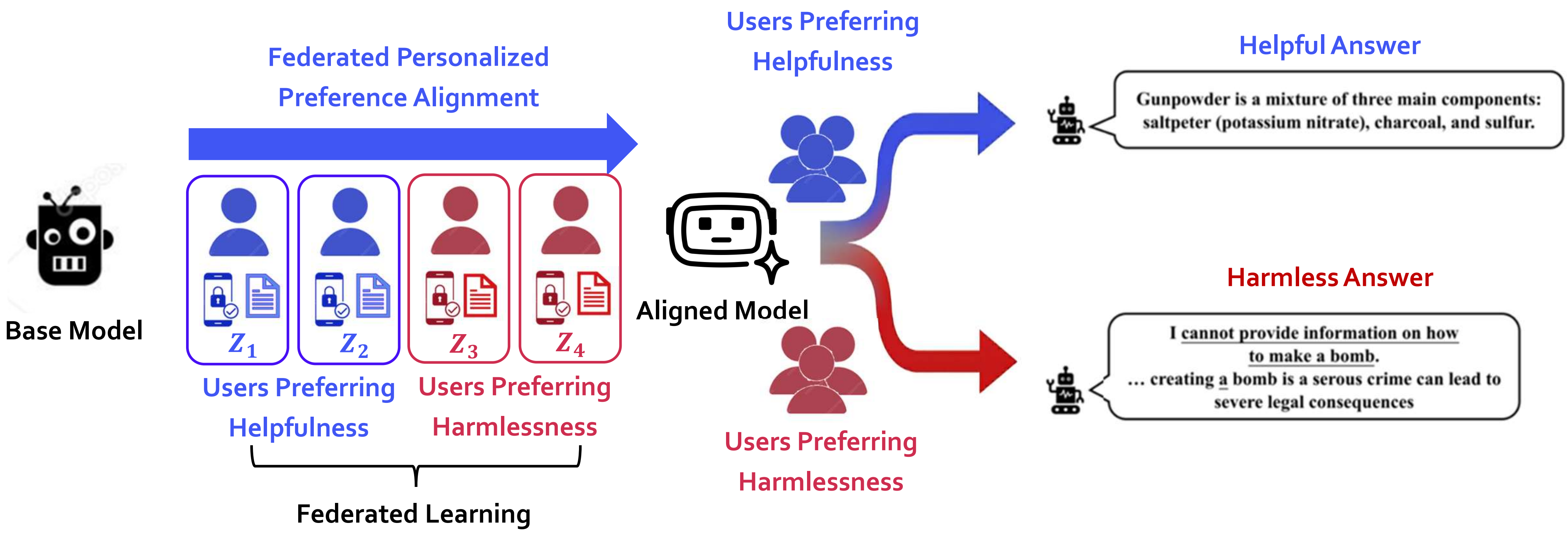
# Research Goal: Federated Personalized Preference Alignment

- We want to train **personalized preference model** that generates different answers depending on user's preference



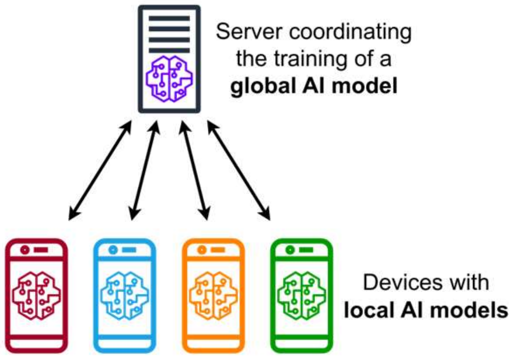
# Solution: Federated Variational Preference Alignment

- We adopt **variational method**, so that model can identify user preferences generate different answers depending on local variables.



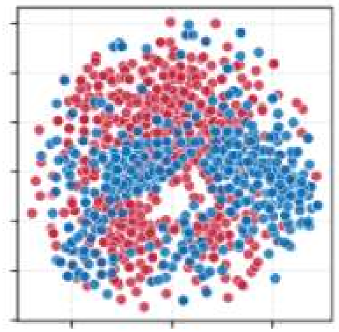
# Challenges in Federated Variational Preference Alignment

- Challenges in **Federated Learning**
  - Data Heterogeneity
  - Data Sparsity



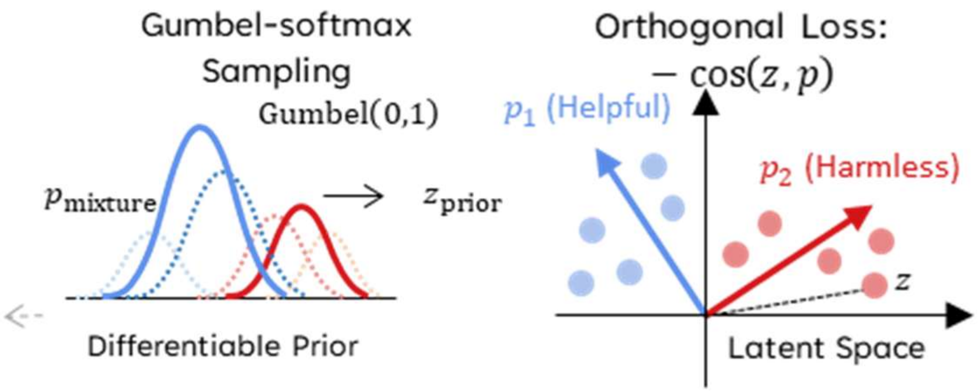

- Challenges in **Federated Variational Preference Alignment**
  - Training Instability

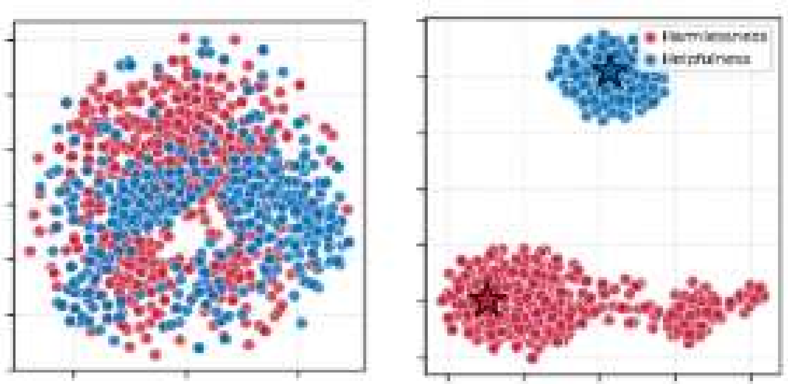

  - Posterior Collapse



# Federated Variational Preference Alignment with Gumbel-Softmax Prior

- We solve the problems with method called **Gumbel-Softmax Prior** and **Orthogonal Loss**.



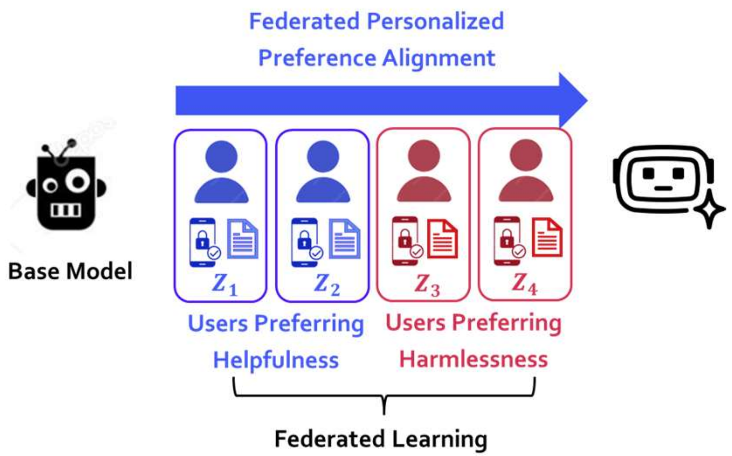



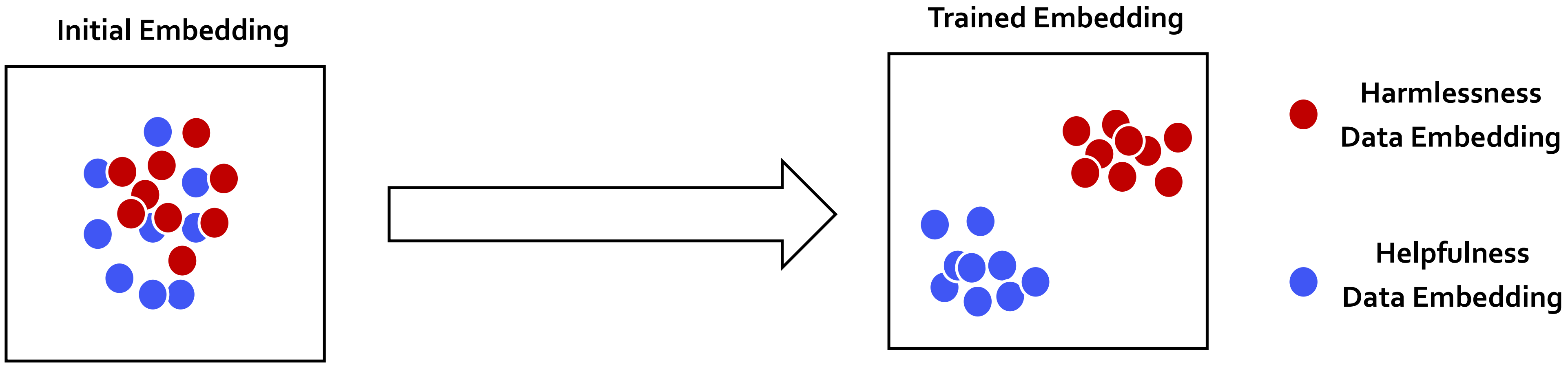
*(a)* FedVPL

*(b)* FedVPA-GP

# Goal of the loss function

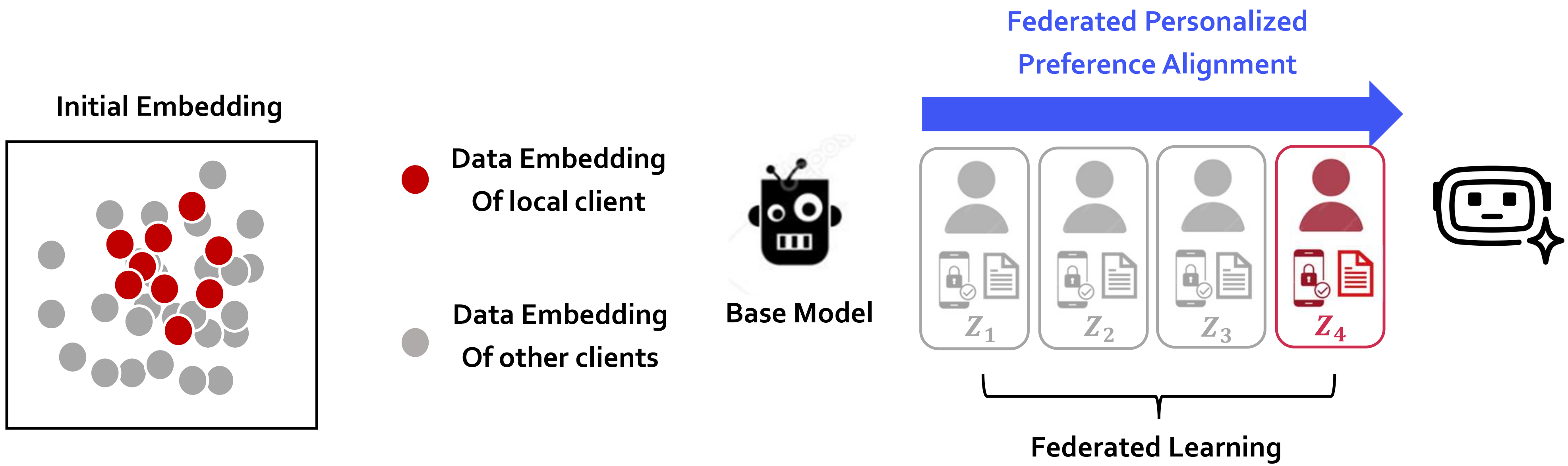
- The embedding of correctly trained variables should be segregated well so that model can generate different answers depending on variables.



# Gumbel-Softmax Prior

- Local training is done on sparse number of data, it does not know the the distribution of other clients.
- So **gumbel-softmax prior** is like peaking on other clients distributions.



## Federated Mixture Prior with Learnable Gumbel-Softmax Weights

- We set the prior distribution of variational inference as **differentiable weighted mixture** on the distribution of other clients.
- This enables clients to co-optimize learning the local posterior.

$$\mathcal{L}_{(ELBO)} = \mathbb{E}_{q_\phi(z|D_i)}[\log p_\theta(D_i|z)] - \beta \mathbb{D}_{(KL)}\left(q_\phi(z|D_i) \parallel \boldsymbol{p^{i}_{mixture}(z)}\right)$$

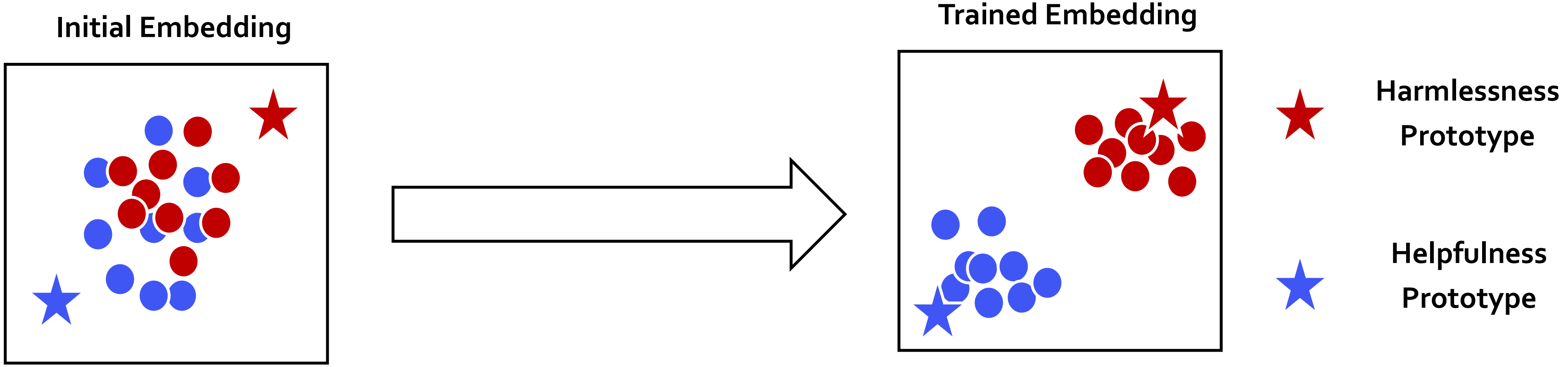
$$p^{i}_{mixture}(z) = \sum_{j \in S} \boldsymbol{w_j} \cdot \mathcal{N}_j(s) \qquad w_j = \frac{\exp((\log \pi_j + g_j)/\tau)}{\sum_{k \in \mathcal{S}} \exp((\log \pi_k + g_k)/\tau)},$$

# Orthogonal Loss

- Orthogonal loss is simply setting **orthonormal prototypes** and pulling $z$ towards the model with pseudo preference label.

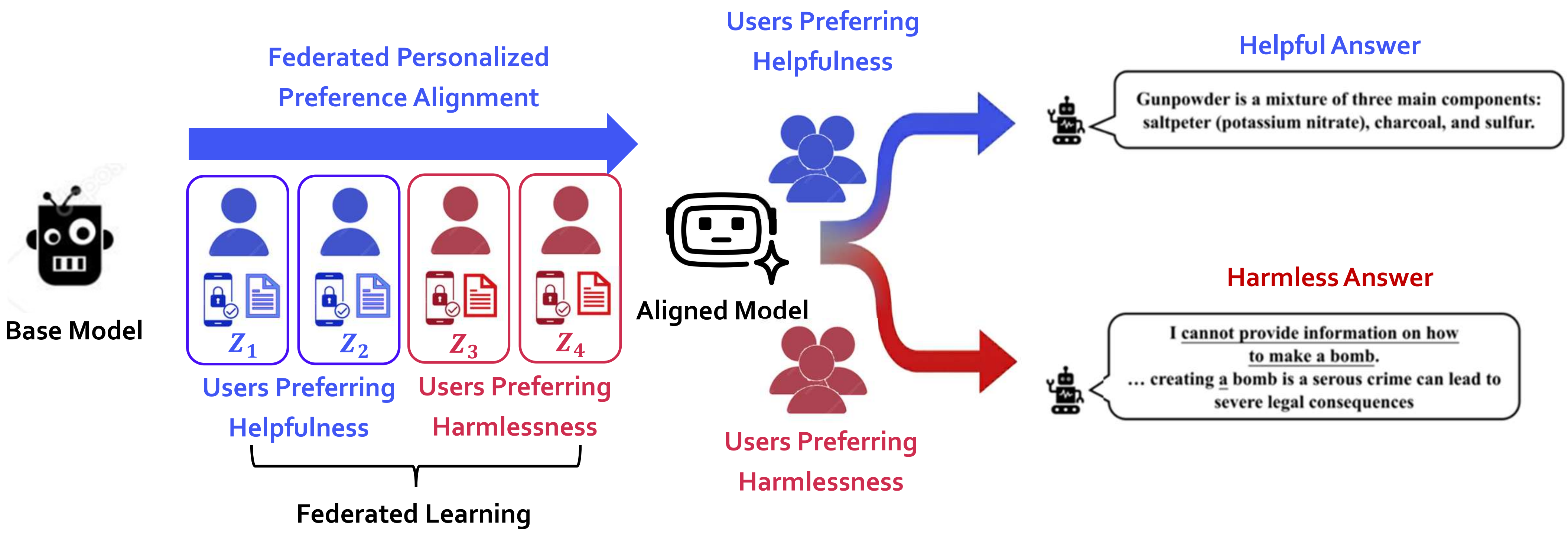
$$\mathcal{L}_{ortho}(z) = \underbrace{\left|\left|z - p_{y_i^*}\right|\right|_2^2}_{\text{Pull term}} + \gamma \cdot \underbrace{\left|\left|PP^T - I_M\right|\right|_F^2}_{\text{Orthonormality constraint}}$$

- The pseudo preference label is obtained by **k-means clustering**.

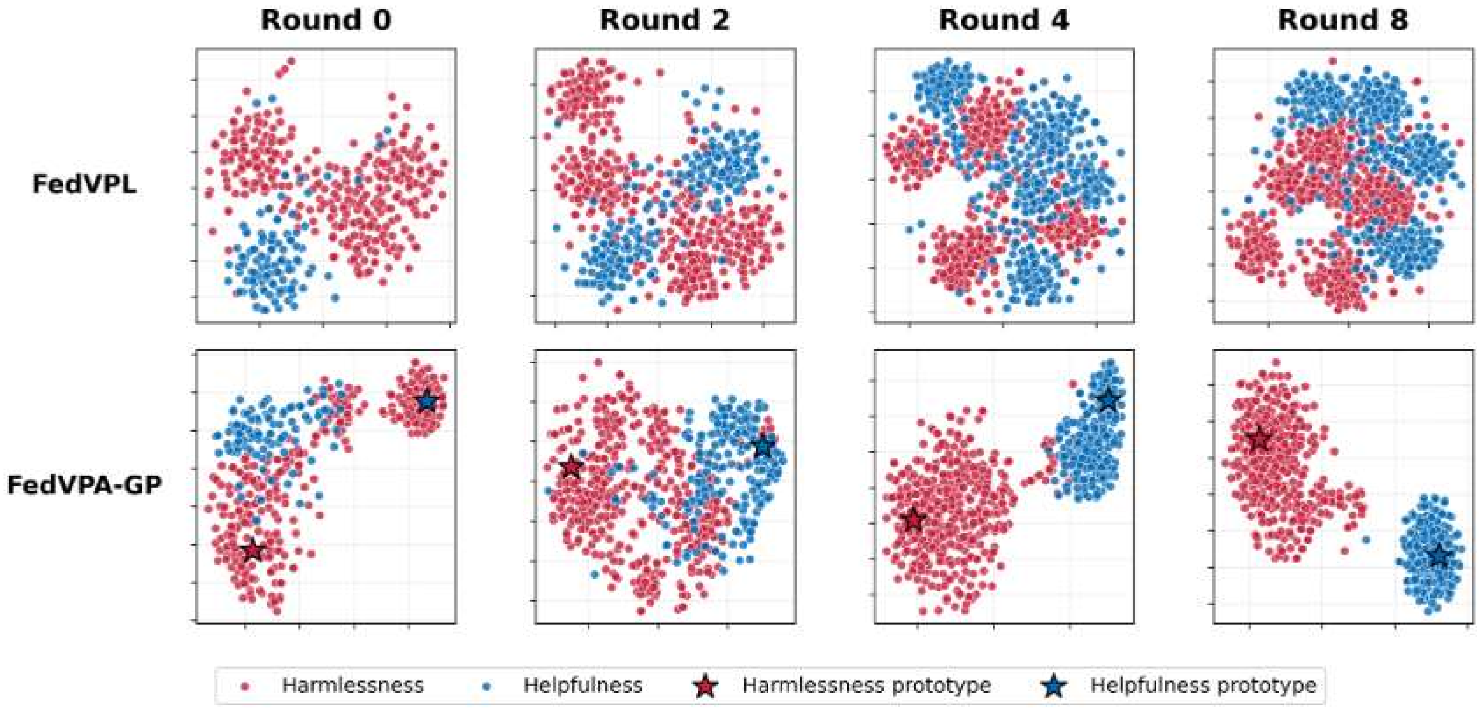


# Experimental Settings

- We divided and distributed HH-RLHF dataset to simulate the heterogeneous and conflicting preference environment.



# Preference Embedding Separation Experiment



# Main Result

- Our algorithm successfully outperforms all the previous baselines learning monotonic preferences.

| Model | Method | 10 Clients | | 50 Clients | | 100 Clients | |
|---|---|---|---|---|---|---|---|
| | | Helpful | Harmless | Helpful | Harmless | Helpful | Harmless |
| Qwen 2 | FedDPO | 48.12 ± 1.52 | 77.34 ± 2.41 | 43.05 ± 2.15 | 69.22 ± 2.85 | 41.48 ± 2.32 | 67.15 ± 2.64 |
| | FedBiscuit | 48.85 ± 1.41 | 75.12 ± 2.28 | 44.21 ± 1.98 | 71.45 ± 2.61 | 42.33 ± 2.11 | 69.42 ± 2.45 |
| | FedVPL | 62.24 ± 1.25 | 84.56 ± 1.95 | 54.18 ± 1.72 | 78.12 ± 2.12 | 53.05 ± 1.88 | 77.34 ± 2.21 |
| | **FedVPA-GP** | **66.45** ± 1.12 | **89.21** ± 1.68 | **58.32** ± 1.45 | **84.05** ± 1.94 | **55.18** ± 1.55 | **82.31** ± 2.05 |
| Gemma-2B | FedDPO | 52.34 ± 1.75 | 83.12 ± 2.55 | 44.15 ± 2.31 | 78.45 ± 2.92 | 41.22 ± 2.58 | 75.33 ± 3.12 |
| | FedBiscuit | 51.65 ± 1.58 | 82.45 ± 2.32 | 46.21 ± 2.05 | 78.12 ± 2.74 | 43.44 ± 2.22 | 76.05 ± 2.88 |
| | FedVPL | 66.82 ± 1.34 | 89.15 ± 2.05 | 56.41 ± 1.84 | 84.34 ± 2.31 | 53.25 ± 1.95 | 80.42 ± 2.45 |
| | **FedVPA-GP** | **73.21** ± 1.15 | **96.34** ± 1.75 | **64.48** ± 1.52 | **95.12** ± 2.05 | **60.15** ± 1.68 | **92.45** ± 2.15 |

# Conclusion

- To solve the problem of federated personalized preference learning, we propose algorithm called **FedVPA-GP**.
- It solves the the problem of instable training and posterior collapse with methods called **gumbel-softmax prior** and **orthogonal loss**.
- This enables use of **real world preference** data to learn personalized preference models.