EPFL

ml4ed

Scholé

Federative RLHF

Jérémy CHAVEROT

Supervised by
Prof. Dr. Tanja Käser

Advised by
Vinitra Swamy and
Paola Mejia

École
polytechnique
fédérale
de Lausanne

13th March 2025

Motivation

- We are witnessing an unprecedented surge in AI assistants and chatbots.
- These advancements are reshaping how users interact with technology and access knowledge.

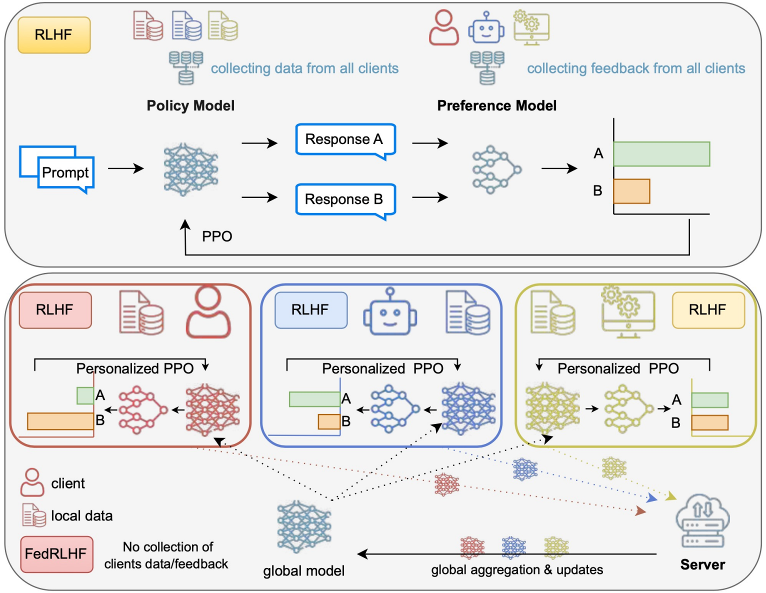
ScholéAI: A Highly Personalized Learning Experience

- **Mission:** Deliver a highly personalized data science learning experience.
- **Approach:** Finetune the Olé tutor by leveraging **Reinforcement Learning from Human Feedback (RLHF)** to refine responses based on user interactions and preferences.
- **Privacy:** Personalization happens locally, preventing real-world user preferences from being transmitted to a central server.

Overview

Scientific Related Work

- **FedRLHF: A Convergence-Guaranteed Federated Framework for Privacy-Preserving and Personalized RLHF [1]**

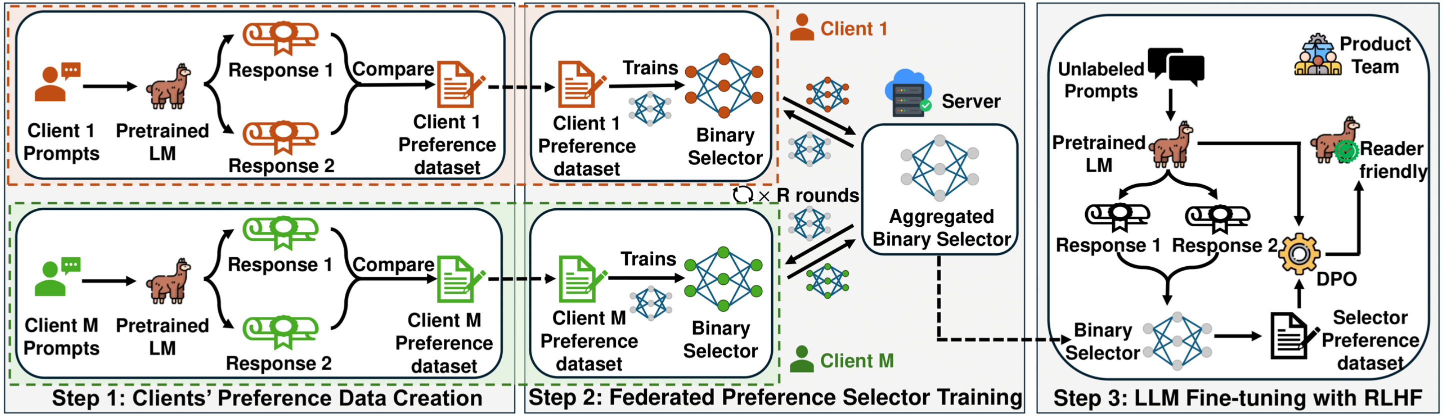


[1] Flint Xiaofeng Fan, Cheston Tan, Yew-Soon Ong, Roger Wattenhofer, & Wei-Tsang Ooi. (2025). FedRLHF: A Convergence-Guaranteed Federated Framework for Privacy-Preserving and Personalized RLHF.

Overview

Scientific Related Work

- **FedBiscuits: Towards Federated RLHF with Aggregated Client Preference for LLMs** [2]



[2] Feijie Wu, Xiaoze Liu, Haoyu Wang, Xingchen Wang, Lu Su, & Jing Gao. (2025). Towards Federated RLHF with Aggregated Client Preference for LLMs.

Research Questions

Objective: Implement a FedRLHF training pipeline (MLOps)

1) How can RLHF be effectively applied to personalize AI responses for data science learning without compromising user privacy?

2) How can we balance the trade-off between personalization and computational efficiency when deploying an AI model locally for each user?

3) What are the key factors and metrics for evaluating the quality and effectiveness of personalized AI-driven responses ?

Methodology

1) How can RLHF be effectively applied to personalize AI responses for data science learning without compromising user privacy?

- **Track 1 – FedRLHF Reimplementation:** Build from scratch using 🤗 TRL, first with a single pipeline, then scale to multiple parallel pipelines in Docker containers with secure model aggregation (FedAvg [3], [4]).

- **Track 2 – FedBiscuit Integration:** Evaluate the existing codebase for compatibility; if suitable, refine and adapt it for seamless integration into ScholéAI.

[3] Sun, Tao, Dongsheng Li, and Bao Wang. "Decentralized federated averaging." IEEE Transactions on Pattern Analysis and Machine Intelligence 45.4 (2022): 4289-4301.

[4] Deng, Yuyang, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. "Distributionally robust federated averaging." Advances in neural information processing systems 33 (2020): 15111-15122.

Methodology

2) How can we balance the trade-off between personalization and computational efficiency when deploying an AI model locally for each user?

- **Track 1 - FedRLHF Reimplementation:** Leverage SoTA **parameter-efficient fine-tuning** (LoRA, QLoRA with 🤗 PEFT); enable **distributed training** with 🤗 accelerate if multiple GPUs are available.

- **Track 2 - FedBiscuit:** A simpler approach with **built-in multi-GPU support** and adapter options for **LoRA parameter selection**.

Methodology

3) What are the key factors and metrics for evaluating the quality and effectiveness of personalized AI-driven responses ?

- **Phase 1:** Fine-tune on synthetic data for initial testing.
- **Phase 2:** Fine-tune on partnered organizations data (Decathlon, ETC, …).
- **Accuracy Goal:** Achieve ±5% of LLaMA 3.1 numbers on Preference Proxy Evaluation (PPE) benchmark metrics that qualifies a reward model's aptitude for RLHF [4].
- **A/B testing:** Show humans responses from both pre-RLHF and post-RLHF models and ask which they prefer.

[4] Frick, Evan, et al. "How to Evaluate Reward Models for RLHF." arXiv preprint arXiv:2410.14872 (2024).

Initial Results

- **Documentation:** Reviewing past work and asking key questions to fully grasp the project.
- **Local RLHF Pipeline:** Running a single RLHF pipeline on a random dataset locally.
- **Dockerization:** Preparing a Docker container to deploy on the RunAI cluster.
- **Synthetic Dataset & Reward Modeling:** Design a dataset with implicit & explicit user traces and define a method to convert entries into rewards/penalties for training the reward model.
- **Exploring FedBiscuit:** Newly discovered this week, next step is to run and test it ourselves.

Planning

Week 1-2: Literature Review & Project Planning

Week 3: Initial Experiments & Baseline Setup

Week 4-5: Implementing a Single RLHF Training Pipeline on RunAI

Week 6-8: Scaling to Multi-Container RLHF Pipelines with Model Aggregation

Week 9: Train on Real Data

Week 10-13: RLHF Model Evaluation & Optimization

Week 14: Finalization & Project Report

EPFL

ml4ed

Thanks

Questions?