



BlossomTuneLLM-MLX

Federated Fine-Tuning of LLMs on Apple Silicon

RMAIIG AI/ML Engineering group - 10/23/2025



About Me





The Origin: Built on Giants

Based on the **Flower Framework**

An open-source framework for **Federated Learning (FL)**.

Enables training ML models across many devices (servers, phones, laptops).

Key Principle: Train models **without** centralizing sensitive data.



The Inspiration

From FlowerTuneLLM...

A benchmarking project for federated supervised fine-tuning (SFT) of LLMs.

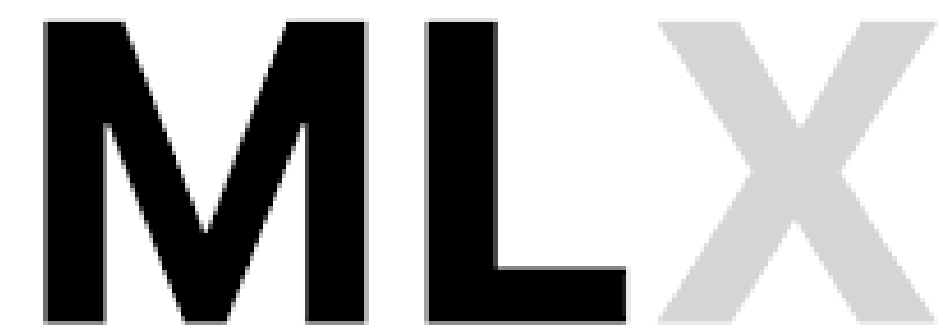
Built on Hugging Face
**transformers + bitsandbytes +
peft + trl (GPU-only)**

To BlossomTuneLLM 🌸

Adapted and refined the
original FlowerTuneLLM code.

Targets both local and cloud
environments **(Containers)**

Still requires a **GPU** setup.



The "Aha!" Moment: MLX

Apple Silicon Changed the Game

M-Series Chips: Incredible performance with Unified Memory.

MLX: Apple's native framework.

Familiar API (like PyTorch/JAX).
Designed from the ground up for Apple Silicon

The Idea

Flower's FL 
+ MLX's Native Performance 



BlossomTuneLLM-MLX

The PyTorch backend is completely replaced with **mlx-lm**.

Native Performance: Runs directly on your M1, M2, M3, or M4 Mac.

No Docker Required: A lightweight, simple, Python installation.

Accessible: Available to the entire Apple ecosystem (students, researchers, developers).

A LAN Party like it's 1999 - And a Real Federated Setup

Server (Superlink): One Mac starts the Flower server.

Clients (Supernodes): Other Macs join the network.

Clients fine-tune on their **own local data**.

Only **adapter weights (LoRA)** are sent back.



Thank You, RMAIIG AI/ML Group 🤖

Goal: Make decentralized, privacy-preserving AI/ML a practical reality for all developers.

Ready for the hands-on deep dive with Uche Ogbuji!

GitHub: <https://github.com/ethicalabs-ai/BlossomTuneLLM-MLX>