■ Full List of Fine-Tuning Frameworks & Tools (2025)

1. Hugging Face Transformers
Most popular for NLP/LLM fine-tuning. Integrates with Trainer, PEFT, Datasets, etc.
https://huggingface.co/docs/transformers

2. PEFT
Techniques: LoRA, Prefix-Tuning, Prompt-Tuning. Lightweight tuning of large models.
https://github.com/huggingface/peft

3. Unsloth
Fast, memory-efficient LoRA fine-tuning. Fine-tunes 7B models on 12GB GPUs.
https://github.com/unslothai/unsloth

4. Axolotl
Complete fine-tuning framework for open LLMs. Supports LoRA, QLoRA, DPO, etc.
https://github.com/OpenAccess-AI-Collective/axolotl

5. DeepSpeed
Distributed fine-tuning & training. Advanced memory optimization (ZeRO, Offloading).
https://www.deepspeed.ai

6. QLoRA
Quantized Low-Rank Adaptation. Fine-tune 33B/65B+ models on limited GPUs.
https://github.com/artidoro/qlora

7. LoRA
Efficient fine-tuning method. Plugged into PEFT, DeepSpeed, Axolotl.
https://github.com/microsoft/LoRA

8. TRL / TRLLM
Reinforcement Learning fine-tuning (RLHF, PPO, DPO). https://github.com/huggingface/trl

9. Colossal-AI
Distributed training for large models. https://www.colossalai.org

10. Lit-GPT
Fast GPT fine-tuning using Lightning AI. https://github.com/Lightning-AI/lit-gpt

11. FastChat
Fine-tune & serve chat models. https://github.com/lm-sys/FastChat

12. OpenLLM
Production-grade LLM serving & tuning. https://github.com/bentoml/OpenLLM

13. Text Generation Inference (TGI)
Inference server for LLMs. https://github.com/huggingface/text-generation-inference

14. LLaMA Factory
Easy fine-tuning for LLaMA models. https://github.com/hiyouga/LLaMA-Factory

15. S-LoRA / SpQR / EXL2 / GPTQ
Advanced quantization and sparse tuning methods. https://github.com/casper-hansen/AutoGPTQ