Hands-On: Fine-Tuning Implementation

Practical Implementation with Hugging Face PEFT
Step-by-step guide to medical LLM fine-tuning

🔧 Setup & Installation

- **Libraries:** transformers, peft, datasets, bitsandbytes
- **Hardware:** Single GPU with 24GB+ memory
- **Environment:** Python 3.9+, PyTorch 2.0+, CUDA 11.8+

💻 Code Implementation

- **Step 1:** Load pre-trained model (LLaMA-7B, Mistral-7B)
- **Step 2:** Configure LoRA (rank=8, alpha=16)
- **Step 3:** Prepare medical dataset (tokenization)
- **Step 4:** Set training arguments (epochs=3, lr=3e-4)
- **Step 5:** Train with Trainer API
- **Step 6:** Merge LoRA weights and save

📈 Monitoring & Logging

- **TensorBoard:** Track loss curves and metrics
- **W&B:** Experiment tracking and comparison

- **GPU Monitoring:** nvidia-smi, watch -n1

30 min	2-8 hrs	100 lines
Setup Time	Training Time	Code