

How to train Open source model like llama3 and mistral using ollama3 in python code step by step guide

Training models like **LLaMA 3** or **Mistral** is slightly different from using them in Ollama.

Important clarification first:

- **Ollama does NOT train base models from scratch.**
- Ollama supports **fine-tuning via adapters (LoRA)** using a `ModelFile`.
- Full training of models like **LLaMA 3** or **Mistral** requires massive GPU infrastructure (hundreds of GB VRAM).

So below is a **step-by-step guide to fine-tune (train) LLaMA 3 or Mistral using Ollama + Python.**

□ Step-by-Step: Fine-tune LLaMA 3 or Mistral with Ollama

We'll fine-tune:

- Meta's **LLaMA 3**
 - or **Mistral 7B**
-

□ Step 1: Install Ollama

Mac / Linux:

```
curl -fsSL https://ollama.com/install.sh | sh
```

Windows:

Download from:

<https://ollama.com>

Check installation:

```
ollama --version
```

☐ Step 2: Pull Base Model

Example with LLaMA 3:

```
ollama pull llama3
```

Or Mistral:

```
ollama pull mistral
```

Test:

```
ollama run llama3
```

☐ Step 3: Prepare Training Data

Ollama fine-tuning expects instruction format.

Create file: `dataset.txt`

Example format:

```
### Instruction:  
Explain quantum computing simply.
```

```
### Response:  
Quantum computing uses quantum bits...
```

You can include multiple instruction-response pairs.

☐ Step 4: Create Modelfile (LoRA Fine-tuning)

Create file named `Modelfile`

For LLaMA 3:

```
FROM llama3
```

```
PARAMETER temperature 0.7
```

SYSTEM You are a helpful AI tutor.

ADAPTER dataset.txt

For Mistral:

FROM mistral

SYSTEM You are a coding assistant.

ADAPTER dataset.txt

□ Step 5: Train (Fine-Tune) Model

Run:

```
ollama create my-custom-model -f Modelfile
```

This creates a **LoRA adapter** on top of base model.

Training time depends on:

- Dataset size
 - RAM
 - CPU/GPU
-

□ Step 6: Test Your Fine-Tuned Model

```
ollama run my-custom-model
```

□ Step 7: Use Fine-Tuned Model in Python

Install Python library:

```
pip install ollama
```

Example Python code:

```
import ollama

response = ollama.chat(
    model='my-custom-model',
```

```
        messages=[
            {'role': 'user', 'content': 'Explain quantum computing'}
        ]
    )

print(response['message']['content'])
```

□ Advanced: Proper Fine-Tuning with PyTorch (Better Method)

If you want **real training control**, use:

- Hugging Face Transformers
- PEFT (LoRA)
- bitsandbytes (QLoRA)

Example architecture:

```
pip install transformers datasets peft accelerate bitsandbytes
```

Basic structure:

```
from transformers import AutoModelForCausalLM, AutoTokenizer
from peft import LoraConfig, get_peft_model
import torch

model_name = "meta-llama/Meta-Llama-3-8B"

tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForCausalLM.from_pretrained(
    model_name,
    load_in_4bit=True,
    device_map="auto"
)

lora_config = LoraConfig(
    r=8,
    lora_alpha=16,
    target_modules=["q_proj", "v_proj"],
    lora_dropout=0.05,
    bias="none",
    task_type="CAUSAL_LM"
)

model = get_peft_model(model, lora_config)

print("Model ready for training")
```

Then after training, you can:

- Export adapter
- Convert to GGUF
- Load into Ollama

❑ Hardware Requirements

Method	GPU Needed
Ollama LoRA small dataset	16GB RAM OK
QLoRA (7B)	24GB GPU ideal
8B+ Full Fine-tune	80GB+ GPU

Dataset.txt file example :-

Instruction:

What computer courses do you offer?

Response:

We offer a variety of computer courses including Basic Computer Skills, MS Office, Advanced Excel, Python Programming, Data Science, Web Development, Graphic Design, Cybersecurity, and Full Stack Development.

Instruction:

What is the duration of the Python programming course?

Response:

The Python programming course typically lasts 3 months. It includes fundamentals, object-oriented programming, working with libraries, and building real-world projects.

Instruction:

What are the fees for the Data Science course?

Response:

The Data Science course fee is \$1200. This includes study materials, project guidance, and certification upon completion.

Instruction:

Do you provide online classes?

Response:

Yes, we provide both online and offline classes. Online sessions are conducted live with recorded lectures available for later access.

Instruction:

Is there any certification after completing the course?

Response:

Yes, students receive a recognized course completion certificate after successfully finishing the course and final assessment.

Instruction:

What are the class timings?

Response:

We offer flexible batch timings including morning, afternoon, evening, and weekend batches to accommodate students and working professionals.

Instruction:

Do I need prior knowledge to join the Web Development course?

Response:

No prior experience is required for the beginner-level Web Development course. We start from the basics and gradually move to advanced topics.

Instruction:

Is there any placement assistance?

Response:

Yes, we provide placement assistance including resume building, interview preparation, and job referrals to partner companies.

Instruction:

Can I pay the course fees in installments?

Response:

Yes, we offer flexible installment payment options. You can pay the fees in two or three installments depending on the course.

Instruction:

Where is your institute located?

Response:

Our institute is located in the city center, easily accessible by public transportation. Detailed address and directions are provided upon registration.

Instruction:

Do you offer demo classes?

Response:

Yes, we offer a free demo class so students can understand the teaching style and course structure before enrolling.

Instruction:

What is taught in the Advanced Excel course?

Response:

The Advanced Excel course covers formulas, pivot tables, data analysis tools, dashboards, macros, and real-world business applications.

Instruction:

How do I enroll in a course?

Response:

You can enroll by filling out the registration form on our website or visiting our institute directly. Our counselors will guide you through the process.

Instruction:

Are there any discounts available?

Response:

Yes, we offer seasonal discounts and special offers for students, early registrations, and group enrollments.

Instruction:

What is included in the Full Stack Development course?

Response:

The Full Stack Development course includes HTML, CSS, JavaScript, React, backend development, databases, APIs, deployment, and project work.

Tips for Better Fine-Tuning

1. Add **50–200 Q&A pairs** for better results.
2. Keep answers consistent in tone.
3. Avoid mixing formats.
4. Add real pricing/location if this is for a real institute.