

Ollama Setup Instructions on Mac

1. Install Ollama

- Go to <https://ollama.com/download>
- Download the macOS installer (.dmg)
- Open the .dmg file and drag Ollama into your Applications folder
- Once installed, Ollama runs as a background service and provides a local API

2. Verify Installation

Open your Terminal and run:

```
ollama --version
```

If installed correctly, you'll see the version number.

3. Run Your First Model

To run Llama 3.1 locally:

```
ollama run llama3.1
```

This will pull the model if not already downloaded and start an interactive session.

4. Manage Models

- List models: `ollama list`
- Pull a new model: `ollama pull mistral`
- Remove a model: `ollama rm mistral`

5. Use Ollama in Development

Ollama exposes a local REST API at `http://localhost:11434`

Example (curl):

```
curl http://localhost:11434/api/generate -d '{"model": "llama3.1", "prompt": "Write me a haiku about autumn leaves"}'
```

Example (Python):

```
import requests
```

```
resp = requests.post("http://localhost:11434/api/generate", json={"model": "llama3.1", "prompt":  
"Explain quantum computing in simple terms"})  
for line in resp.iter_lines():  
    if line:  
        print(line.decode())
```

6. Optional: Autostart on Login

Ollama automatically starts as a background service when you log in. You can control it from the macOS menu bar icon (top-right).