

# ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

## DAY – 15

Date: Jul 11, 2025

### Running llama3 model on ubuntu

Llama 3, Meta's latest open-source AI model, represents a major leap in scalable AI innovation. Built with flexibility and performance in mind, Llama 3 is designed to handle various AI tasks, from natural language processing to interactive chat models, making it a powerful tool for developers, researchers, and AI enthusiasts alike. In this guide, we'll walk you through the steps to set up and run Llama 3 on a Linux system. Whether you're an experienced AI developer or new to machine learning, this step-by-step guide will help you get up and running, even on modest hardware.

#### Prerequisites

Before we get started, let's ensure your system meets the necessary hardware and software requirements to run Llama 3 efficiently. Here's what you'll need:

- **GPU:** An Nvidia GPU with at least 8GB of VRAM (12GB or more is recommended for better performance, especially with larger models).
- **Operating System:** Ubuntu 20.04 or a similar Linux distribution.
- **Python:** Version 3.8 or newer.
- **CUDA drivers:** Ensure that Nvidia's CUDA toolkit is properly installed and configured on your machine.

#### Step 1: Update Your System

Before starting the installation process, it is recommended to update your system to ensure all packages are up to date. Run the following command in the terminal:

```
sudo apt update && sudo apt upgrade -y
```

This command updates the package list and installs the latest software updates on your system.

```
step@step-HP-ProDesk-400-G5-SFF: ~  
step@step-HP-ProDesk-400-G5-SFF:~$ sudo apt update  
[sudo] password for step:  
Hit:1 http://in.archive.ubuntu.com/ubuntu focal InRelease  
Get:2 http://in.archive.ubuntu.com/ubuntu focal-updates InRelease [128 kB]  
Get:3 http://in.archive.ubuntu.com/ubuntu focal-backports InRelease [128 kB]  
Get:4 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages [3,954 kB]  
Get:5 http://security.ubuntu.com/ubuntu focal-security InRelease [128 kB]  
Get:6 http://in.archive.ubuntu.com/ubuntu focal-updates/main i386 Packages [1,114 kB]  
Get:7 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 DEP-11 Metadata [276 kB]  
Get:8 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted amd64 DEP-11 Metadata [212 B]  
Get:9 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 Packages [1,262 kB]  
Get:10 http://in.archive.ubuntu.com/ubuntu focal-updates/universe i386 Packages [824 kB]  
Get:11 http://security.ubuntu.com/ubuntu focal-security/main i386 Packages [881 kB]  
Get:12 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 DEP-11 Metadata [445 kB]  
Get:13 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 DEP-11 Metadata [940 B]  
Get:14 http://in.archive.ubuntu.com/ubuntu focal-backports/main amd64 DEP-11 Metadata [7,952 B]  
Get:15 http://in.archive.ubuntu.com/ubuntu focal-backports/restricted amd64 DEP-11 Metadata [216 B]  
Get:16 http://in.archive.ubuntu.com/ubuntu focal-backports/universe amd64 DEP-11 Metadata [30.5 kB]  
Get:17 http://in.archive.ubuntu.com/ubuntu focal-backports/multiverse amd64 DEP-11 Metadata [212 B]  
Get:18 http://security.ubuntu.com/ubuntu focal-security/main amd64 Packages [3,564 kB]  
Get:19 http://security.ubuntu.com/ubuntu focal-security/main amd64 DEP-11 Metadata [74.7 kB]  
Get:20 http://security.ubuntu.com/ubuntu focal-security/restricted amd64 DEP-11 Metadata [212 B]  
Get:21 http://security.ubuntu.com/ubuntu focal-security/universe amd64 Packages [1,040 kB]  
Get:22 http://security.ubuntu.com/ubuntu focal-security/universe i386 Packages [698 kB]  
Get:23 http://security.ubuntu.com/ubuntu focal-security/universe amd64 DEP-11 Metadata [160 kB]  
Get:24 http://security.ubuntu.com/ubuntu focal-security/multiverse amd64 DEP-11 Metadata [940 B]  
Fetched 14.7 MB in 15s (1,004 kB/s)  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
11 packages can be upgraded. Run 'apt list --upgradable' to see them.
```

Then install curl by running the following command:

```
sudo apt install curl
```

```
step@step-HP-ProDesk-400-G5-SFF:~$ sudo apt install curl  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following packages were automatically installed and are no longer required:  
  gir1.2-goa-1.0 libfwupdplugin1 libxmb1  
Use 'sudo apt autoremove' to remove them.  
The following NEW packages will be installed:  
  curl  
0 upgraded, 1 newly installed, 0 to remove and 11 not upgraded.  
Need to get 162 kB of archives.  
After this operation, 414 kB of additional disk space will be used.  
Get:1 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 curl amd64 7.68.0-1ubuntu2.25 [162 kB]  
Fetched 162 kB in 1s (186 kB/s)  
Selecting previously unselected package curl.  
(Reading database ... 189316 files and directories currently installed.)  
Preparing to unpack .../curl_7.68.0-1ubuntu2.25_amd64.deb ...  
Unpacking curl (7.68.0-1ubuntu2.25) ...  
Setting up curl (7.68.0-1ubuntu2.25) ...  
Processing triggers for man-db (2.9.1-1) ...  
step@step-HP-ProDesk-400-G5-SFF:~$ curl -fsSL https://ollama.com/install.sh | sh
```

## Step 2: Download the Ollama installation package

Next, download the Ollama installation package for Linux from the official website by running the following command:

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

## Step 3: Verify the Installation

After installing Ollama, it's time to verify that everything is working correctly. Run the following command to check the version of Ollama:

```
ollama --version
```

```
Processing triggers for man-db (2.9.1-1) ...
step@step-HP-ProDesk-400-G5-SFF:~$ curl -fsSL https://ollama.com/install.sh | sh
>>> Installing ollama to /usr/local
>>> Downloading Linux amd64 bundle
##### 100.0%
>>> Creating ollama user...
[sudo] password for step:
>>> Adding ollama user to render group...
>>> Adding ollama user to video group...
>>> Adding current user to ollama group...
>>> Creating ollama systemd service...
>>> Enabling and starting ollama service...
Created symlink /etc/systemd/system/default.target.wants/ollama.service → /etc/systemd/system/ollama.service.
>>> The Ollama API is now available at 127.0.0.1:11434.
>>> Install complete. Run "ollama" from the command line.
WARNING: No NVIDIA/AMD GPU detected. Ollama will run in CPU-only mode.
step@step-HP-ProDesk-400-G5-SFF:~$ ollama --version
ollama version is 0.9.5
```

## Step 4: Running Ollama with Llama 3.3

To run Ollama with the Llama 3.3 model, you can execute a command like the following:

```
ollama run llama3
```

```
step@step-HP-ProDesk-400-G5-SFF:~$ ollama pull llama3
pulling manifest
Error: pull model manifest: file does not exist
step@step-HP-ProDesk-400-G5-SFF:~$ ollama --version
ollama version is 0.9.5
step@step-HP-ProDesk-400-G5-SFF:~$ ollama pull llama3
pulling manifest
Error: pull model manifest: file does not exist
step@step-HP-ProDesk-400-G5-SFF:~$ ollama pull llama3
pulling manifest
pulling 6a0746a1ec1a: 100% 4.7 GB
pulling 4fa551d4f938: 100% 12 KB
pulling 8ab4849b038c: 100% 254 B
pulling 577073ffcc6c: 100% 110 B
pulling 3f8eb4da87fa: 100% 485 B
verifying sha256 digest
writing manifest
success
step@step-HP-ProDesk-400-G5-SFF:~$ ollama run llama3
>>> what is ai
A great question!

Artificial Intelligence (AI) refers to the development of computer systems that can perform tasks that would typically require human intelligence, such as:

1. **Learning**: AI systems can learn from data and improve their performance over time.
2. **Reasoning**: AI systems can draw conclusions based on given information and make decisions.
3. **Problem-solving**: AI systems can find solutions to complex problems.
4. **Perception**: AI systems can interpret and understand sensory information, such as images, speech, or text.

AI systems use various techniques, including:

1. **Machine learning**: AI systems learn from data and improve their performance over time.
2. **Deep learning**: A subset of machine learning that uses neural networks to analyze complex data.
3. **Natural language processing (NLP)**: AI systems can understand, generate, and process human language.
4. **Computer vision**: AI systems can interpret and understand visual information from images or videos.

AI has many applications across various industries, including:
```