

CPU only



- [Nvidia GPU](#)
 - [Install with Apt](#)
 - [Install with Yum or Dnf](#)
 - [Configure Docker to use Nvidia driver](#)
 - [Start the container](#)
 - [AMD GPU](#)
 - [Vulkan Support](#)
 - [Run model locally](#)
 - [Try different models](#)
-

```
docker run -d -v ollama:/root/.ollama -p 11434:11434 --name ollama ollama/ollama
```

Nvidia GPU

Install the [NVIDIA Container Toolkit](#).

Install with Apt

1. Configure the repository

```
curl -fsSL https://nvidia.github.io/libnvidia-container/gpgkey \  
    | sudo gpg --dearmor -o /usr/share/keyrings/nvidia-container-toolkit-keyring.gpg  
curl -fsSL https://nvidia.github.io/libnvidia-container/stable/deb/nvidia-container-  
toolkit.list \  
    | sed 's#deb https://#deb [signed-by=/usr/share/keyrings/nvidia-container-  
toolkit-keyring.gpg] https://#g' \  
    | sudo tee /etc/apt/sources.list.d/nvidia-container-toolkit.list  
sudo apt-get update
```

2. Install the NVIDIA Container Toolkit packages

```
sudo apt-get install -y nvidia-container-toolkit
```

Install with Yum or Dnf

1. Configure the repository

```
curl -fsSL https://nvidia.github.io/libnvidia-container/stable/rpm/nvidia-container-toolkit.repo \
| sudo tee /etc/yum.repos.d/nvidia-container-toolkit.repo
```

2. Install the NVIDIA Container Toolkit packages

```
sudo yum install -y nvidia-container-toolkit
```

Configure Docker to use Nvidia driver

```
sudo nvidia-ctl runtime configure --runtime=docker
sudo systemctl restart docker
```

Start the container

```
docker run -d --gpus=all -v ollama:/root/.ollama -p 11434:11434 --name ollama ollama/ollama
```

If you're running on an NVIDIA JetPack system, Ollama can't automatically discover the correct JetPack version. Pass the environment variable `JETSON\_JETPACK=5` or `JETSON\_JETPACK=6` to the container to select version 5 or 6.

AMD GPU

To run Ollama using Docker with AMD GPUs, use the `rocm` tag and the following command:

```
docker run -d --device /dev/kfd --device /dev/dri -v ollama:/root/.ollama -p 11434:11434 --name ollama ollama/ollama:rocm
```

Vulkan Support

Vulkan is bundled into the `ollama/ollama` image and is enabled by default when the container can access the GPU devices.

```
docker run -d --device /dev/kfd --device /dev/dri -v ollama:/root/.ollama -p 11434:11434 --name ollama ollama/ollama
```

Use `OLLAMA_VULKAN=0` to disable Vulkan, or `GGML_VK_VISIBLE_DEVICES=<ids>` to select specific Vulkan devices.

Run model locally

Now you can run a model:

```
docker exec -it ollama ollama run llama3.2
```

Try different models

More models can be found on the [Ollama library](#).

`ollama` #2

`ollama` 2026-10-31 13:17:09 UTC

`ollama` 2026-10-31 13:17:10 UTC