I got the following error on ubuntu when using pytorch, can you help : """ File "/home/gongai/.local/lib/python3.10/site-packages/torch/cuda/init.py", line 298, in *lazy*init torch._C._cuda_init() RuntimeError: CUDA unknown error - this may be due to an incorrectly set up environment, """

Claude
1) check if your NVIDIA drivers are properly installed:
```
nvidia-smi
```

2) Verify your CUDA version:

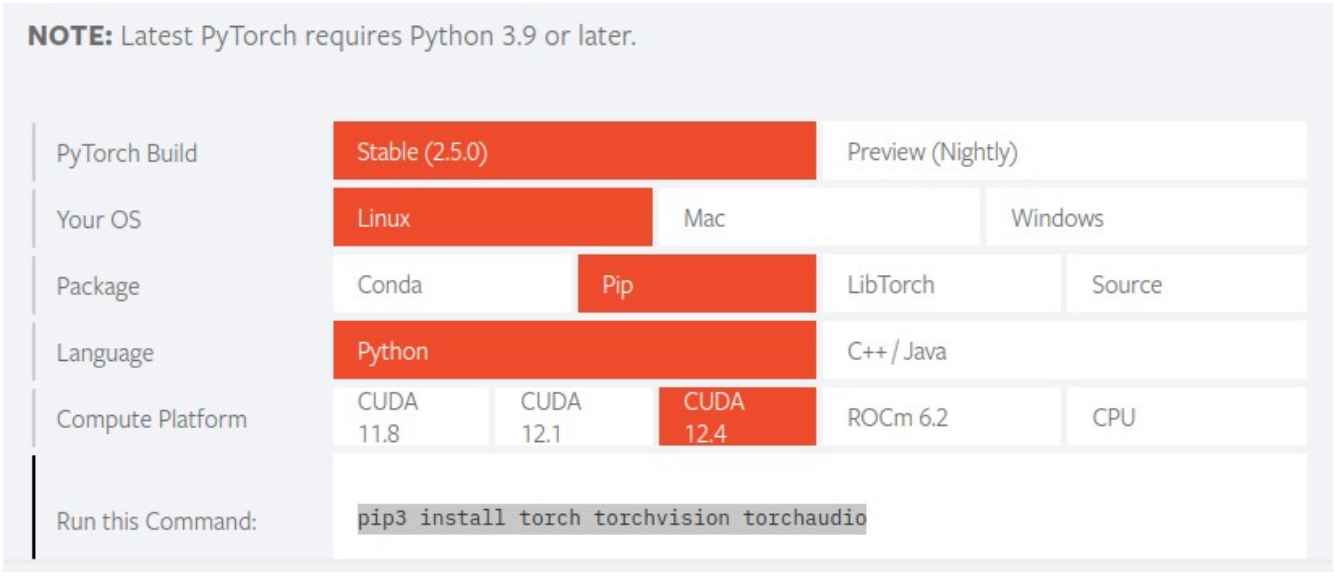
```
$ nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2023 NVIDIA Corporation
Built on Fri_Nov__3_17:16:49_PDT_2023
Cuda compilation tools, release 12.3, V12.3.103
Build cuda_12.3.r12.3/compiler.33492891_0
```

3) Check which version of PyTorch you have installed:
```
$ python3 -c "import torch; print(torch.__version__)"
2.1.2+cu121
```

```
(promptfix) gongai@ducklover1:~$ nvidia-smi
Sat Oct 26 10:52:15 2024
+-----------------------------------------------------------------------------------------+
| NVIDIA-SMI 550.107.02             Driver Version: 550.107.02     CUDA Version: 12.4     |
|-----------------------------------------+------------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id          Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |           Memory-Usage | GPU-Util  Compute M. |
|                                         |                        |               MIG M. |
|=========================================+========================+======================|
|   0  NVIDIA GeForce RTX 4060        Off |   00000000:01:00.0  On |                  N/A |
|  0%   36C    P8             N/A /  115W |     366MiB /   8188MiB |      0%      Default |
|                                         |                        |                  N/A |
+-----------------------------------------+------------------------+----------------------+

+-----------------------------------------------------------------------------------------+
| Processes:                                                                              |
|  GPU   GI   CI        PID   Type   Process name                              GPU Memory |
|        ID   ID                                                               Usage      |
|=========================================================================================|
|    0   N/A  N/A      9950      G   /usr/lib/xorg/Xorg                            116MiB |
|    0   N/A  N/A     10087      G   /usr/bin/gnome-shell                          140MiB |
|    0   N/A  N/A     12191      G   ...seed-version=20241025-130107.645000         98MiB |
+-----------------------------------------------------------------------------------------+
(promptfix) gongai@ducklover1:~$ nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2023 NVIDIA Corporation
Built on Fri_Nov__3_17:16:49_PDT_2023
Cuda compilation tools, release 12.3, V12.3.103
Build cuda_12.3.r12.3/compiler.33492891_0
(promptfix) gongai@ducklover1:~$ python3 -c "import torch; print(torch.__version__)"
2.1.2+cu121
```

https://pytorch.org/




```
pip3 install torch torchvision torchaudio
```

Instead of installing the version (which is old) you suggested, I went to pytorch.org, and ran this command to install the latest pytorch 2.5: """pip3 install torch torchvision torchaudio """, then verify its version by """ python3 -c "import torch; print(torch.version); print(torch.cuda.is_available())" """, it throws the following error: """ Traceback (most recent call last): File "<string>", line 1, in <module> File "/home/gongai/anaconda3/envs/promptfix/lib/python3.10/site-packages/torch/init.py", line 235, in <module> from torch._C import * # noqa: F403 ImportError: libcudnn.so.8: cannot open shared object file: No such file or directory """, libcudnn.so.8 share lib is missing, how to install that?

Need to install cuDNN (CUDA Deep Neural Network library)
```
sudo apt update
sudo apt install nvidia-cuda-toolkit

# Add NVIDIA repository key
wget https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64/cuda-ubuntu2204.pin
sudo mv cuda-ubuntu2204.pin /etc/apt/preferences.d/cuda-repository-pin-600
sudo apt-key adv --fetch-keys https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64/3bf863cc.pub
sudo add-apt-repository "deb https://developer.download.nvidia.com/compute/cuda/repos/ubuntu2204/x86_64/ /"

# Install cuDNN
sudo apt update
sudo apt install libcudnn8

python3 -c "import torch; print(torch.__version__); print(torch.cuda.is_available())"
```

New error:
Traceback (most recent call last):
  File "<string>", line 1, in <module>
  File "/home/gongai/anaconda3/envs/promptfix/lib/python3.10/site-packages/torch/__init__.py", line 235, in <module>
    from torch._C import * # noqa: F403
ImportError: libnccl.so.2: cannot open shared object file: No such file or directory

You need to install the NVIDIA Collective Communications Library (NCCL).

```
# Install NCCL library
sudo apt update
sudo apt install libnccl2 libnccl-dev

$ python3 -c "import torch; print(torch.__version__); print(torch.cuda.is_available())"
2.1.2+cu121
True
```

```

running `scripts/inference.py` throws this error:

```
  File "/home/gongai/anaconda3/envs/promptfix/lib/python3.10/site-packages/torch/cuda/amp/autocast_mode.py", line 119, i
n decorate_fwd
    return fwd(*_cast(args, cast_inputs), **_cast(kwargs, cast_inputs))
  File "/home/gongai/projects/wgong/PromptFix/./stable_diffusion/ldm/modules/diffusionmodules/util.py", line 145, in for
ward
    output_tensors = ctx.run_function(*ctx.input_tensors)
  File "/home/gongai/projects/wgong/PromptFix/./stable_diffusion/ldm/modules/attention.py", line 251, in _forward
    x = self.attn1(self.norm1(x)) + x
  File "/home/gongai/anaconda3/envs/promptfix/lib/python3.10/site-packages/torch/nn/modules/module.py", line 1518, in _w
rapped_call_impl
    return self._call_impl(*args, **kwargs)
  File "/home/gongai/anaconda3/envs/promptfix/lib/python3.10/site-packages/torch/nn/modules/module.py", line 1527, in _c
all_impl
    return forward_call(*args, **kwargs)
  File "/home/gongai/projects/wgong/PromptFix/./stable_diffusion/ldm/modules/attention.py", line 192, in forward
    sim = einsum('b i d, b j d -> b i j', q, k) * self.scale
  File "/home/gongai/anaconda3/envs/promptfix/lib/python3.10/site-packages/torch/functional.py", line 377, in einsum
    return _VF.einsum(equation, operands) # type: ignore[attr-defined]
torch.cuda.OutOfMemoryError: CUDA out of memory. Tried to allocate 6.19 GiB. GPU 0 has a total capacty of 7.74 GiB of wh
ich 1.64 GiB is free. Including non-PyTorch memory, this process has 5.58 GiB memory in use. Of the allocated memory 5.0
3 GiB is allocated by PyTorch, and 430.07 MiB is reserved by PyTorch but unallocated. If reserved but unallocated memory
 is large try setting max_split_size_mb to avoid fragmentation. See documentation for Memory Management and PYTORCH_CUD
A_ALLOC_CONF
```
```