Installation and Load packages

```
!pip install datasets peft -qq
!pip install accelerate -qq
!pip install bitsandbytes -qq

!pip install torch==2.2.0 torchvision==0.17.0 torchaudio==2.2.0 --
index-url https://download.pytorch.org/whl/cu118
!pip install --upgrade --pre transformers accelerate --extra-index-url
https://download.pytorch.org/whl/cu118
!pip install bitsandbytes==0.43.2 --prefer-binary --extra-index-url
https://pypi.org/simple

Defaulting to user installation because normal site-packages is not
writeable
Looking in indexes: https://download.pytorch.org/whl/cu118
Collecting torch==2.2.0
  Downloading https://download.pytorch.org/whl/cu118/torch-
2.2.0%2Bcu118-cp310-cp310-linux_x86_64.whl (811.7 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 811.7/811.7 MB 1.7 MB/s eta
0:00:0000:0100:01
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.2/6.2 MB 87.6 MB/s eta
0:00:00:00:01
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 3.3/3.3 MB 50.8 MB/s eta
0:00:00a 0:00:01
ent already satisfied: jinja2 in
/home/student/.local/lib/python3.10/site-packages (from torch==2.2.0)
(3.1.3)
Collecting nvidia-cufft-cu11==10.9.0.58
  Downloading
https://download.pytorch.org/whl/cu118/nvidia_cufft_cu11-10.9.0.58-
py3-none-manylinux1_x86_64.whl (168.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 168.4/168.4 MB 8.1 MB/s eta
0:00:0000:0100:01
anylinux1_x86_64.whl (58.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 58.1/58.1 MB 23.8 MB/s eta
0:00:0000:0100:01
anylinux1_x86_64.whl (204.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 204.1/204.1 MB 6.9 MB/s eta
0:00:0000:0100:01
e-cu11==11.8.89
  Downloading
https://download.pytorch.org/whl/cu118/nvidia_cuda_runtime_cu11-
11.8.89-py3-none-manylinux1_x86_64.whl (875 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 875.6/875.6 kB 51.6 MB/s eta
0:00:00
anylinux1_x86_64.whl (23.2 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 23.2/23.2 MB 62.2 MB/s eta
```

```
0:00:0000:0100:01
anylinux1_x86_64.whl (417.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 417.9/417.9 MB 3.3 MB/s eta
0:00:0000:0100:01
ent already satisfied: filelock in
/home/student/.local/lib/python3.10/site-packages (from torch==2.2.0)
(3.13.1)
Requirement already satisfied: sympy in
/opt/conda/lib/python3.10/site-packages (from torch==2.2.0) (1.12)
Requirement already satisfied: networkx in
/opt/conda/lib/python3.10/site-packages (from torch==2.2.0) (3.1)
Requirement already satisfied: fsspec in
/home/student/.local/lib/python3.10/site-packages (from torch==2.2.0)
(2024.2.0)
Collecting nvidia-nvtx-cu11==11.8.86
  Downloading https://download.pytorch.org/whl/cu118/nvidia_nvtx_cu11-
11.8.86-py3-none-manylinux1_x86_64.whl (99 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 99.1/99.1 kB 17.2 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (167.9 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 167.9/167.9 MB 8.4 MB/s eta
0:00:0000:0100:01
anylinux1_x86_64.whl (728.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 728.5/728.5 MB 1.9 MB/s eta
0:00:0000:0100:01
ent already satisfied: typing-extensions>=4.8.0 in
/home/student/.local/lib/python3.10/site-packages (from torch==2.2.0)
(4.10.0)
Collecting nvidia-cusolver-cu11==11.4.1.48
  Downloading
https://download.pytorch.org/whl/cu118/nvidia_cusolver_cu11-11.4.1.48-
py3-none-manylinux1_x86_64.whl (128.2 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 128.2/128.2 MB 10.8 MB/s eta
0:00:0000:0100:01
anylinux1_x86_64.whl (135.3 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 135.3/135.3 MB 10.3 MB/s eta
0:00:0000:0100:01
anylinux1_x86_64.whl (13.1 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.1/13.1 MB 19.5 MB/s eta
0:00:0000:0100:01
ent already satisfied: numpy in
/home/student/.local/lib/python3.10/site-packages (from
torchvision==0.17.0) (1.26.4)
Requirement already satisfied: pillow!=8.3.*,>=5.3.0 in
/home/student/.local/lib/python3.10/site-packages (from
torchvision==0.17.0) (10.2.0)
Requirement already satisfied: requests in
/home/student/.local/lib/python3.10/site-packages (from
torchvision==0.17.0) (2.31.0)
```

```
Requirement already satisfied: MarkupSafe>=2.0 in
/home/student/.local/lib/python3.10/site-packages (from jinja2-
>torch==2.2.0) (2.1.5)
Requirement already satisfied: certifi>=2017.4.17 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>torchvision==0.17.0) (2024.2.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>torchvision==0.17.0) (2.2.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>torchvision==0.17.0) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>torchvision==0.17.0) (3.6)
Requirement already satisfied: mpmath>=0.19 in
/opt/conda/lib/python3.10/site-packages (from sympy->torch==2.2.0)
(1.3.0)
Installing collected packages: triton, nvidia-nvtx-cu11, nvidia-nccl-
cu11, nvidia-cusparse-cu11, nvidia-curand-cu11, nvidia-cufft-cu11,
nvidia-cuda-runtime-cu11, nvidia-cuda-nvrtc-cu11, nvidia-cuda-cupti-
cu11, nvidia-cublas-cu11, nvidia-cusolver-cu11, nvidia-cudnn-cu11,
torch, torchvision, torchaudio
  WARNING: The scripts convert-caffe2-to-onnx, convert-onnx-to-caffe2
and torchrun are installed in '/home/student/.local/bin' which is not
on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress
this warning, use --no-warn-script-location.
Successfully installed nvidia-cublas-cu11-11.11.3.6 nvidia-cuda-cupti-
cu11-11.8.87 nvidia-cuda-nvrtc-cu11-11.8.89 nvidia-cuda-runtime-cu11-
11.8.89 nvidia-cudnn-cu11-8.7.0.84 nvidia-cufft-cu11-10.9.0.58 nvidia-
curand-cu11-10.3.0.86 nvidia-cusolver-cu11-11.4.1.48 nvidia-cusparse-
cu11-11.7.5.86 nvidia-nccl-cu11-2.19.3 nvidia-nvtx-cu11-11.8.86 torch-
2.2.0+cu118 torchaudio-2.2.0+cu118 torchvision-0.17.0+cu118 triton-
2.2.0
Defaulting to user installation because normal site-packages is not
writeable
Looking in indexes: https://pypi.org/simple,
https://download.pytorch.org/whl/cu118
Requirement already satisfied: transformers in
/opt/conda/lib/python3.10/site-packages (4.36.0)
Collecting transformers
  Downloading transformers-4.51.1-py3-none-any.whl (10.4 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 10.4/10.4 MB 60.5 MB/s eta
0:00:0000:010:01
ent already satisfied: accelerate in /opt/conda/lib/python3.10/site-
packages (0.25.0)

Collecting accelerate
  Downloading accelerate-1.6.0-py3-none-any.whl (354 kB)
```

```
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 354.7/354.7 kB 41.0 MB/s eta
0:00:00
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.0 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 3.0/3.0 MB 86.5 MB/s eta
0:00:00:00:01
anylinux_2_17_x86_64.manylinux2014_x86_64.whl (471 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 471.6/471.6 kB 27.3 MB/s eta
0:00:00
ent already satisfied: requests in
/home/student/.local/lib/python3.10/site-packages (from transformers)
(2.31.0)
Collecting huggingface-hub<1.0,>=0.30.0
  Downloading huggingface_hub-0.30.2-py3-none-any.whl (481 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 481.4/481.4 kB 45.8 MB/s eta
0:00:00
ent already satisfied: pyyaml>=5.1 in
/home/student/.local/lib/python3.10/site-packages (from transformers)
(6.0.1)
Requirement already satisfied: packaging>=20.0 in
/home/student/.local/lib/python3.10/site-packages (from transformers)
(24.0)
Requirement already satisfied: numpy>=1.17 in
/home/student/.local/lib/python3.10/site-packages (from transformers)
(1.26.4)
Requirement already satisfied: regex!=2019.12.17 in
/opt/conda/lib/python3.10/site-packages (from transformers)
(2023.12.25)
Requirement already satisfied: tqdm>=4.27 in
/home/student/.local/lib/python3.10/site-packages (from transformers)
(4.66.2)
Requirement already satisfied: filelock in
/home/student/.local/lib/python3.10/site-packages (from transformers)
(3.13.1)
Requirement already satisfied: torch>=2.0.0 in
/home/student/.local/lib/python3.10/site-packages (from accelerate)
(2.2.0+cu118)
Requirement already satisfied: psutil in
/opt/conda/lib/python3.10/site-packages (from accelerate) (5.9.0)
Requirement already satisfied: typing-extensions>=3.7.4.3 in
/home/student/.local/lib/python3.10/site-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (4.10.0)
Requirement already satisfied: fsspec>=2023.5.0 in
/home/student/.local/lib/python3.10/site-packages (from huggingface-
hub<1.0,>=0.30.0->transformers) (2024.2.0)
Requirement already satisfied: nvidia-nccl-cu11==2.19.3 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (2.19.3)
Requirement already satisfied: jinja2 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
```

```
>accelerate) (3.1.3)
Requirement already satisfied: nvidia-cuda-runtime-cu11==11.8.89 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.8.89)
Requirement already satisfied: nvidia-cublas-cu11==11.11.3.6 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.11.3.6)
Requirement already satisfied: nvidia-cusolver-cu11==11.4.1.48 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.4.1.48)
Requirement already satisfied: triton==2.2.0 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (2.2.0)
Requirement already satisfied: nvidia-cuda-cupti-cu11==11.8.87 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.8.87)
Requirement already satisfied: nvidia-cufft-cu11==10.9.0.58 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (10.9.0.58)
Requirement already satisfied: nvidia-cudnn-cu11==8.7.0.84 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (8.7.0.84)
Requirement already satisfied: nvidia-curand-cu11==10.3.0.86 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (10.3.0.86)
Requirement already satisfied: networkx in
/opt/conda/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (3.1)
Requirement already satisfied: nvidia-cuda-nvrtc-cu11==11.8.89 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.8.89)
Requirement already satisfied: nvidia-cusparse-cu11==11.7.5.86 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.7.5.86)
Requirement already satisfied: nvidia-nvtx-cu11==11.8.86 in
/home/student/.local/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (11.8.86)
Requirement already satisfied: sympy in
/opt/conda/lib/python3.10/site-packages (from torch>=2.0.0-
>accelerate) (1.12)
Requirement already satisfied: idna<4,>=2.5 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>transformers) (3.6)
Requirement already satisfied: certifi>=2017.4.17 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>transformers) (2024.2.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>transformers) (2.2.1)
```

```
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/student/.local/lib/python3.10/site-packages (from requests-
>transformers) (3.3.2)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/student/.local/lib/python3.10/site-packages (from jinja2-
>torch>=2.0.0->accelerate) (2.1.5)
Requirement already satisfied: mpmath>=0.19 in
/opt/conda/lib/python3.10/site-packages (from sympy->torch>=2.0.0-
>accelerate) (1.3.0)
Installing collected packages: safetensors, huggingface-hub,
tokenizers, accelerate, transformers
  Attempting uninstall: huggingface-hub
    Found existing installation: huggingface-hub 0.21.4
    Uninstalling huggingface-hub-0.21.4:
      Successfully uninstalled huggingface-hub-0.21.4
  WARNING: The script huggingface-cli is installed in
'/home/student/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress
this warning, use --no-warn-script-location.
  WARNING: The scripts accelerate, accelerate-config, accelerate-
estimate-memory, accelerate-launch and accelerate-merge-weights are
installed in '/home/student/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress
this warning, use --no-warn-script-location.
  WARNING: The script transformers-cli is installed in
'/home/student/.local/bin' which is not on PATH.
  Consider adding this directory to PATH or, if you prefer to suppress
this warning, use --no-warn-script-location.
Successfully installed accelerate-1.6.0 huggingface-hub-0.30.2
safetensors-0.5.3 tokenizers-0.21.1 transformers-4.51.1
Defaulting to user installation because normal site-packages is not
writeable
Looking in indexes: https://pypi.org/simple, https://pypi.org/simple
Collecting bitsandbytes==0.43.2
  Downloading bitsandbytes-0.43.2-py3-none-manylinux_2_24_x86_64.whl
(137.5 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 137.5/137.5 MB 6.5 MB/s eta
0:00:0000:0100:01
ent already satisfied: torch in
/home/student/.local/lib/python3.10/site-packages (from
bitsandbytes==0.43.2) (2.2.0+cu118)
Requirement already satisfied: numpy in
/home/student/.local/lib/python3.10/site-packages (from
bitsandbytes==0.43.2) (1.26.4)
Requirement already satisfied: jinja2 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (3.1.3)
Requirement already satisfied: nvidia-cuda-cupti-cu11==11.8.87 in
/home/student/.local/lib/python3.10/site-packages (from torch-
```

```
>bitsandbytes==0.43.2) (11.8.87)
Requirement already satisfied: networkx in
/opt/conda/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (3.1)
Requirement already satisfied: nvidia-cudnn-cu11==8.7.0.84 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (8.7.0.84)
Requirement already satisfied: sympy in
/opt/conda/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (1.12)
Requirement already satisfied: fsspec in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (2024.2.0)
Requirement already satisfied: nvidia-nccl-cu11==2.19.3 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (2.19.3)
Requirement already satisfied: filelock in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (3.13.1)
Requirement already satisfied: nvidia-cusparse-cu11==11.7.5.86 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (11.7.5.86)
Requirement already satisfied: nvidia-cublas-cu11==11.11.3.6 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (11.11.3.6)
Requirement already satisfied: nvidia-curand-cu11==10.3.0.86 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (10.3.0.86)
Requirement already satisfied: nvidia-cufft-cu11==10.9.0.58 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (10.9.0.58)
Requirement already satisfied: nvidia-cusolver-cu11==11.4.1.48 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (11.4.1.48)
Requirement already satisfied: nvidia-nvtx-cu11==11.8.86 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (11.8.86)
Requirement already satisfied: nvidia-cuda-nvrtc-cu11==11.8.89 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (11.8.89)
Requirement already satisfied: triton==2.2.0 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (2.2.0)
Requirement already satisfied: typing-extensions>=4.8.0 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (4.10.0)
Requirement already satisfied: nvidia-cuda-runtime-cu11==11.8.89 in
/home/student/.local/lib/python3.10/site-packages (from torch-
>bitsandbytes==0.43.2) (11.8.89)
```

```
Requirement already satisfied: MarkupSafe>=2.0 in
/home/student/.local/lib/python3.10/site-packages (from jinja2->torch-
>bitsandbytes==0.43.2) (2.1.5)
Requirement already satisfied: mpmath>=0.19 in
/opt/conda/lib/python3.10/site-packages (from sympy->torch-
>bitsandbytes==0.43.2) (1.3.0)
Installing collected packages: bitsandbytes
Successfully installed bitsandbytes-0.43.2

!pip install wandb scikit-learn

Defaulting to user installation because normal site-packages is not
writeable
Requirement already satisfied: wandb in
/home/student/.local/lib/python3.10/site-packages (0.19.9)
Requirement already satisfied: scikit-learn in
/home/student/.local/lib/python3.10/site-packages (1.6.1)
Requirement already satisfied: click!=8.0.0,>=7.1 in
/home/student/.local/lib/python3.10/site-packages (from wandb) (8.1.7)
Requirement already satisfied: sentry-sdk>=2.0.0 in
/home/student/.local/lib/python3.10/site-packages (from wandb)
(2.25.1)
Requirement already satisfied: pyyaml in
/home/student/.local/lib/python3.10/site-packages (from wandb) (6.0.1)
Requirement already satisfied: platformdirs in
/opt/conda/lib/python3.10/site-packages (from wandb) (4.2.0)
Requirement already satisfied: psutil>=5.0.0 in
/opt/conda/lib/python3.10/site-packages (from wandb) (5.9.0)
Requirement already satisfied: setuptools in
/opt/conda/lib/python3.10/site-packages (from wandb) (65.6.3)
Requirement already satisfied: gitpython!=3.1.29,>=1.0.0 in
/home/student/.local/lib/python3.10/site-packages (from wandb)
(3.1.44)
Requirement already satisfied: setproctitle in
/home/student/.local/lib/python3.10/site-packages (from wandb) (1.3.5)
Requirement already satisfied: requests<3,>=2.0.0 in
/home/student/.local/lib/python3.10/site-packages (from wandb)
(2.31.0)
Requirement already satisfied: pydantic<3 in
/home/student/.local/lib/python3.10/site-packages (from wandb) (2.6.4)
Requirement already satisfied: typing-extensions<5,>=4.4 in
/home/student/.local/lib/python3.10/site-packages (from wandb)
(4.10.0)
Requirement already satisfied: docker-pycreds>=0.4.0 in
/home/student/.local/lib/python3.10/site-packages (from wandb) (0.4.0)
Requirement already satisfied: protobuf!=4.21.0,!=5.28.0,<6,>=3.19.0
in /opt/conda/lib/python3.10/site-packages (from wandb) (4.25.3)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/home/student/.local/lib/python3.10/site-packages (from scikit-learn)
(3.6.0)
```

```
Requirement already satisfied: joblib>=1.2.0 in
/home/student/.local/lib/python3.10/site-packages (from scikit-learn)
(1.4.2)
Requirement already satisfied: numpy>=1.19.5 in
/home/student/.local/lib/python3.10/site-packages (from scikit-learn)
(1.26.4)
Requirement already satisfied: scipy>=1.6.0 in
/opt/conda/lib/python3.10/site-packages (from scikit-learn) (1.11.2)
Requirement already satisfied: six>=1.4.0 in
/home/student/.local/lib/python3.10/site-packages (from docker-
pycreds>=0.4.0->wandb) (1.16.0)
Requirement already satisfied: gitdb<5,>=4.0.1 in
/home/student/.local/lib/python3.10/site-packages (from gitpython!
=3.1.29,>=1.0.0->wandb) (4.0.12)
Requirement already satisfied: pydantic-core==2.16.3 in
/home/student/.local/lib/python3.10/site-packages (from pydantic<3-
>wandb) (2.16.3)
Requirement already satisfied: annotated-types>=0.4.0 in
/home/student/.local/lib/python3.10/site-packages (from pydantic<3-
>wandb) (0.6.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/home/student/.local/lib/python3.10/site-packages (from
requests<3,>=2.0.0->wandb) (3.3.2)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/student/.local/lib/python3.10/site-packages (from
requests<3,>=2.0.0->wandb) (2.2.1)
Requirement already satisfied: certifi>=2017.4.17 in
/home/student/.local/lib/python3.10/site-packages (from
requests<3,>=2.0.0->wandb) (2024.2.2)
Requirement already satisfied: idna<4,>=2.5 in
/home/student/.local/lib/python3.10/site-packages (from
requests<3,>=2.0.0->wandb) (3.6)
Requirement already satisfied: smmap<6,>=3.0.1 in
/home/student/.local/lib/python3.10/site-packages (from
gitdb<5,>=4.0.1->gitpython!=3.1.29,>=1.0.0->wandb) (5.0.2)
```

GPU - details

```python
import torch

print("Torch version:", torch.__version__)
print("CUDA available:", torch.cuda.is_available())

if torch.cuda.is_available():
    print("Device name:", torch.cuda.get_device_name(0))
else:
    print("No GPU detected.")
```

```
Torch version: 2.2.0+cu118
CUDA available: True
Device name: Tesla T4
```

Load libraries, Login HuggingFace API & WandB API

- **HuggingFace API:** To get access of Model Llama-3 (8 Billion)
- **WandB (Weigths & Biases):** To supervise perform of model and hyperparameter Tuning

```
# from google.colab import userdata
from huggingface_hub import login

login(token="YOUR_HF_API_KEY")

# Access Key for llama Model (HuggingFace)

from datasets import load_dataset, Dataset
from sklearn.model_selection import train_test_split

from transformers import (
    AutoTokenizer,
    AutoModelForCausalLM,
    TrainingArguments,
    DataCollatorForLanguageModeling,
    Trainer,
    BitsAndBytesConfig)

from peft import prepare_model_for_kbit_training, LoraConfig,
get_peft_model
from bitsandbytes.optim import AdamW8bit

# for hyperparameter tuning report
import wandb


wandb.login()
# YOUR_WANDB_API_KEY
```

```
wandb: Using wandb-core as the SDK backend.  Please refer to
https://wandb.me/wandb-core for more information.
wandb: Logging into wandb.ai. (Learn how to deploy a W&B server
locally: https://wandb.me/wandb-server)
wandb: You can find your API key in your browser here:
https://wandb.ai/authorize
wandb: Paste an API key from your profile and hit enter:

 ........
```

```
wandb: WARNING If you're specifying your api key in code, ensure this
code is not shared publicly.
wandb: WARNING Consider setting the WANDB_API_KEY environment
variable, or running `wandb login` from the command line.
wandb: No netrc file found, creating one.
wandb: Appending key for api.wandb.ai to your netrc file:
/home/student/.netrc
wandb: Currently logged in as: yashnayi00 (yashnayi00-university-of-
new-haven) to https://api.wandb.ai. Use `wandb login --relogin` to
force relogin

True
```

Load Llama-3.1-8B model

```
# model_name = "meta-llama/Llama-3.1-8B-Instruct"
model_name = "meta-llama/Llama-3.1-8B"

bnb_config = BitsAndBytesConfig(
    load_in_4bit=True,
    bnb_4bit_quant_type="nf4",
    bnb_4bit_compute_dtype=torch.bfloat16,
    bnb_4bit_use_double_quant=False
    )


tokenizer = AutoTokenizer.from_pretrained(model_name)

base_model = AutoModelForCausalLM.from_pretrained(
    model_name,
    device_map="auto",
    quantization_config=bnb_config,
)

if tokenizer.pad_token is None:
    tokenizer.pad_token = tokenizer.eos_token

# model.config.pretraining_tp = 1
# model.config.use_cache = False

{"model_id":"9cc356b8c34b42c5ae133f73ae4dda6b","version_major":2,"vers
ion_minor":0}

{"model_id":"819dd82ef0a9447aa5083b4bf6358b1b","version_major":2,"vers
ion_minor":0}

{"model_id":"0f6d4bb5028147999d2f66849b015e88","version_major":2,"vers
ion_minor":0}
```

{"model_id":"c0edb317fd894ab38f5d5919bdefe4ad","version_major":2,"version_minor":0}

{"model_id":"aae1c69734dc408eade664089e204879","version_major":2,"version_minor":0}

{"model_id":"45c9dd449c0b498387af098102fcc772","version_major":2,"version_minor":0}

{"model_id":"5679a1b53dbb40b8839d1e50c7f5c167","version_major":2,"version_minor":0}

{"model_id":"4a50e81f83c84cb18fb977c353a2a7b4","version_major":2,"version_minor":0}

{"model_id":"addd29f2fd2948c0abce3a31832630ff","version_major":2,"version_minor":0}

{"model_id":"4caf9879f2c3450198a23b831c954770","version_major":2,"version_minor":0}

{"model_id":"2ed9d30afb03467c8c672cc43cc5d1f3","version_major":2,"version_minor":0}

{"model_id":"50c512fbb9cb441e89ae5dcca3cc4f66","version_major":2,"version_minor":0}

```
print(f"meta-llama/Llama-3-8B:\n{base_model}")
```

```
meta-llama/Llama-3-8B:
LlamaForCausalLM(
  (model): LlamaModel(
    (embed_tokens): Embedding(128256, 4096)
    (layers): ModuleList(
      (0-31): 32 x LlamaDecoderLayer(
        (self_attn): LlamaAttention(
          (q_proj): Linear4bit(in_features=4096, out_features=4096, bias=False)
          (k_proj): Linear4bit(in_features=4096, out_features=1024, bias=False)
          (v_proj): Linear4bit(in_features=4096, out_features=1024, bias=False)
          (o_proj): Linear4bit(in_features=4096, out_features=4096, bias=False)
        )
        (mlp): LlamaMLP(
          (gate_proj): Linear4bit(in_features=4096, out_features=14336, bias=False)
          (up_proj): Linear4bit(in_features=4096, out_features=14336, bias=False)
          (down_proj): Linear4bit(in_features=14336,
```

```
out_features=4096, bias=False)
          (act_fn): SiLU()
        )
        (input_layernorm): LlamaRMSNorm((4096,), eps=1e-05)
        (post_attention_layernorm): LlamaRMSNorm((4096,), eps=1e-05)
      )
    )
    (norm): LlamaRMSNorm((4096,), eps=1e-05)
    (rotary_emb): LlamaRotaryEmbedding()
  )
  (lm_head): Linear(in_features=4096, out_features=128256, bias=False)
)

print(f"{base_model.config}")

LlamaConfig {
  "_attn_implementation_autoset": true,
  "architectures": [
    "LlamaForCausalLM"
  ],
  "attention_bias": false,
  "attention_dropout": 0.0,
  "bos_token_id": 128000,
  "eos_token_id": 128001,
  "head_dim": 128,
  "hidden_act": "silu",
  "hidden_size": 4096,
  "initializer_range": 0.02,
  "intermediate_size": 14336,
  "max_position_embeddings": 131072,
  "mlp_bias": false,
  "model_type": "llama",
  "num_attention_heads": 32,
  "num_hidden_layers": 32,
  "num_key_value_heads": 8,
  "pretraining_tp": 1,
  "quantization_config": {
    "_load_in_4bit": true,
    "_load_in_8bit": false,
    "bnb_4bit_compute_dtype": "bfloat16",
    "bnb_4bit_quant_storage": "uint8",
    "bnb_4bit_quant_type": "nf4",
    "bnb_4bit_use_double_quant": false,
    "llm_int8_enable_fp32_cpu_offload": false,
    "llm_int8_has_fp16_weight": false,
    "llm_int8_skip_modules": null,
    "llm_int8_threshold": 6.0,
    "load_in_4bit": true,
    "load_in_8bit": false,
    "quant_method": "bitsandbytes"
```

```
  },
  "rms_norm_eps": 1e-05,
  "rope_scaling": {
    "factor": 8.0,
    "high_freq_factor": 4.0,
    "low_freq_factor": 1.0,
    "original_max_position_embeddings": 8192,
    "rope_type": "llama3"
  },
  "rope_theta": 500000.0,
  "tie_word_embeddings": false,
  "torch_dtype": "float16",
  "transformers_version": "4.51.1",
  "use_cache": true,
  "vocab_size": 128256
}
```


Trainable parameters - Model

```
def trainable_parameters(model):
    """
    Prints the number of trainable parameters in the model.
    """
    trainable_params = 0
    all_param = 0
    for _, param in model.named_parameters():
        all_param += param.numel()
        if param.requires_grad:
            trainable_params += param.numel()
    return f"- Trainable model parameters: {trainable_params}.\n- All model parameters: {all_param}.\n- Percentage of trainable model parameters: {100 * trainable_params / all_param:.2f}%"

print(trainable_parameters(base_model))
```

```
- Trainable model parameters: 1050939392.
- All model parameters: 4540600320.
- Percentage of trainable model parameters: 23.15%
```

Assign datasetPH.json

Data is split in to train and test.

- Train size: 80%
- Test size: 20%

```
import json
with open("./dataset/datasetPH.json", "r") as f:
    data = json.load(f)
```

```python
if isinstance(data, dict):
    print("Data is a dictionary. Converting values to a list for splitting.")
    data = list(data.values())

train_data, test_data = train_test_split(data, test_size=0.2, random_state=42)

with open("./dataset/train_datasetPH.json", "w") as f:
    json.dump(train_data, f, indent=2)

with open("./dataset/test_datasetPH.json", "w") as f:
    json.dump(test_data, f, indent=2)

print(f"Train size: {len(train_data)}")
print(f"Test size: {len(test_data)}")
```

```
Data is a dictionary. Converting values to a list for splitting.
Train size: 160
Test size: 41
```

```python
data[0]
```

```
{'paper_id': 'ED012836',
 'title': 'Adult Basic Education Work Book in Basic Arithmetic. Parts I and II.',
 'author': 'Graham, Minnie M.',
 'publication_year': 1966,
 'source': 'Danbury Public Schools, Connecticut',
 'doi_or_url': '',
 'topic_category': 'Adult Education / Arithmetic Instruction',
 'document_type': 'Workbook',
 'abstract': 'These workbooks provide teaching materials and drill exercises in multiplication for adult basic education learners in Danbury, Connecticut. Part I covers multiplication by numbers two through nine, while Part II expands to ten through twelve, including dollars and cents, and offers speed and accuracy drills.',
 'key_findings': 'Instructional workbooks tailored for adult learners can assist in foundational arithmetic, especially multiplication, through structured drills and exercises.',
 'problem_statement': 'Adult learners require appropriately designed arithmetic materials to support basic educational needs at elementary levels.',
 'objectives': 'To provide instructors with multiplication drill materials suitable for adults operating at elementary grade levels.',
 'conclusion': 'The workbook is a supportive instructional aid that must be supplemented with additional materials and practice exercises to effectively meet adult learners’ needs.',
 'methodology': {'data_sources': 'Experience and requests from
```

```
arithmetic teachers of adult students.',
  'methods_used': 'Instructional material design and exercise
formulation.',
  'sample_size': None,
  'duration': '1966-1967 academic year',
  'research_design': 'Development and application of a structured
workbook for classroom use.'},
 'metrics_and_indicators': [{'metric_name': 'Speed and accuracy in
multiplication',
   'metric_value': None}],
 'policy_practice_implications': {'recommendations': 'Use the workbook
as a teaching aid for adult learners needing arithmetic instruction at
elementary level.',
  'implementation_notes': 'Instructors should supplement with
additional materials to ensure comprehensive understanding.'},
 'thematic_dimensions': {'demographic_focus': 'Adults in basic
education programs',
  'geographic_scope': 'Danbury, Connecticut',
  'domain_keywords': ['multiplication',
   'adult education',
   'arithmetic',
   'instructional materials',
   'workbook']},
 'comparative_and_qualitative_insights': {'comparative_data': '',
  'thematic_analysis': 'Focuses on gradual progression from simpler to
more complex multiplication problems with contextual financial
applications.',
  'limitations': 'Workbook alone is insufficient for comprehensive
instruction.',
  'future_work': 'Create additional supporting materials and broader
coverage of arithmetic topics.'},
 'supporting_materials': {'tables': ['Multiplication tables from 2
through 12'],
  'charts': [],
  'appendices': [],
  'external_links': []},
 'references': []}
```

Prompt Engineering

```
def build_prompt(entry):
    persona = "You are a public policy analyst specializing in
educational reform.\n"
    instruction = ("Summarize the key findings from the report below.
Your output should include:\n"
                   "- Three bullet points summarizing the findings\n"
                   "- One paragraph about implications\n"
                   "- A JSON tag with `impact` set to positive,
negative, or neutral\n")
```

```
    context = f"This report evaluates a new adult education
intervention implemented in {entry.get('thematic_dimensions',
{}).get('geographic_scope', 'a specific region')}.\n"
    format_guide = "Use professional and concise tone. Output must be
structured: bullet points, paragraph, then JSON.\n"
    few_shot = ("Example Input: \"The policy resulted in 70%
improvement in adult math scores and lowered dropout rates.\"\n"
                "Example Output:\n- Improved math proficiency by 70%\
n"
                "- Reduced dropout rates significantly\n"
                "- High engagement among learners\n"
                "Implication: These results show the program is
effective and could be scaled to other regions.\n"
                "{\"impact\": \"positive\"}\n")

    # Build the body of the report
    full_text = (
        f"Title: {entry.get('title', '')}\n"
        f"Abstract: {entry.get('abstract', '')}\n"
        f"Key Findings: {entry.get('key_findings', '')}\n"
        f"Problem Statement: {entry.get('problem_statement', '')}\n"
        f"Objectives: {entry.get('objectives', '')}\n"
        f"Conclusion: {entry.get('conclusion', '')}\n"
        f"Methodology: {entry.get('methodology',
{}).get('methods_used', '')}, "
        f"based on data from {entry.get('methodology',
{}).get('data_sources', '')}, "
        f"conducted over {entry.get('methodology', {}).get('duration',
'')}\n"
        f"Implications: {entry.get('policy_practice_implications',
{}).get('recommendations', '')} "
        f"{entry.get('policy_practice_implications',
{}).get('implementation_notes', '')}\n"
        f"Thematic Focus: {entry.get('thematic_dimensions',
{}).get('demographic_focus', '')} | "
        f"{entry.get('topic_category', '')}\n"
        f"Limitations:
{entry.get('comparative_and_qualitative_insights',
{}).get('limitations', '')}\n"
        f"Future Work:
{entry.get('comparative_and_qualitative_insights',
{}).get('future_work', '')}\n"
    )

    return persona + instruction + context + format_guide + few_shot +
"Now analyze this:\n" + full_text
```

Tokenization of dataset and normalization

```
# def tokenize_function(examples):
#     texts = []
#     for i in range(len(examples["title"])):
#         entry_parts = []

#         for key in examples.keys():
#             value = examples[key][i]
#             if isinstance(value, dict):
#                 for subkey, subval in value.items():
#                     entry_parts.append(f"{key}.{subkey}: {subval}")
#             elif isinstance(value, list):
#                 entry_parts.append(f"{key}: {', '.join(map(str,
value))}")
#             else:
#                 entry_parts.append(f"{key}: {value}")

#         combined_text = "\n".join(entry_parts)
#         texts.append(combined_text)

#     return tokenizer(texts, truncation=True, padding="max_length",
max_length=256)


def tokenize_function(examples):
    prompts = []
    for i in range(len(examples["title"])):
        entry = {key: examples[key][i] for key in examples}
        full_prompt = build_prompt(entry)
        prompts.append(full_prompt)

    return tokenizer(prompts, truncation=True, padding="max_length",
max_length=512)

def normalize_entry(entry):
    normalized = {}
    for key, value in entry.items():
        if isinstance(value, dict):
            for subkey, subval in value.items():
                normalized[f"{key}.{subkey}"] = str(subval) if subval
is not None else ""
        elif isinstance(value, list):
            normalized[key] = ", ".join(map(str, value))
        elif value is None:
            normalized[key] = ""
        else:
            normalized[key] = str(value)
    return normalized

# Normalize each entry
```

```
train_data_clean = [normalize_entry(entry) for entry in train_data]
test_data_clean = [normalize_entry(entry) for entry in test_data]

train_dataset_hf = Dataset.from_list(train_data_clean)
test_dataset_hf = Dataset.from_list(test_data_clean)
```

Train & Test - Tokenization

```
tokenized_train = train_dataset_hf.map(tokenize_function,
batched=True)
tokenized_train.set_format(type="torch")
print("Tokenization complete with all features.")
```


```
{"model_id":"e6b0ae19490345e79b4677909b632406","version_major":2,"vers
ion_minor":0}
```


```
Tokenization complete with all features.
```

```
tokenized_test = test_dataset_hf.map(tokenize_function, batched=True)
tokenized_test.set_format(type="torch")
print("Tokenization complete with all features.")
```


```
{"model_id":"237585d024634261b6256aff86e70cb9","version_major":2,"vers
ion_minor":0}
```


```
Tokenization complete with all features.
```

Configer - PEFT, LoRA & QLoRA

```
lora_config = LoraConfig(
    r=8,
    lora_alpha=32,
    target_modules=["q_proj", "v_proj"],
    lora_dropout=0.05,
    bias="none",
    task_type="CAUSAL_LM"
)

base_model.gradient_checkpointing_enable()

base_model = prepare_model_for_kbit_training(base_model)
peft_model = get_peft_model(base_model, lora_config)

peft_model.config.use_cache = False

print("After PEFT wrapping:")
print(trainable_parameters(peft_model))
```

```
After PEFT wrapping:
- Trainable model parameters: 3407872.
```

```
- All model parameters: 4544008192.
- Percentage of trainable model parameters: 0.07%

def compute_metrics(eval_pred):
    predictions, labels = eval_pred
    preds = predictions.argmax(-1)
    accuracy = (preds == labels).astype(float).mean().item()
    return {"accuracy": accuracy}
```

Train PH-Llama-3.0 Model & Evaluation

```
import torch
import os
data_collator = DataCollatorForLanguageModeling(tokenizer=tokenizer,
mlm=False)

os.environ["PYTORCH_CUDA_ALLOC_CONF"] = "expandable_segments:True"

training_args = TrainingArguments(
    output_dir="./PH-Llama-3.0",
    overwrite_output_dir=True,
    num_train_epochs=5,
    per_device_train_batch_size=1,
    per_device_eval_batch_size=1,
    gradient_accumulation_steps=1,
    learning_rate=2e-5,
    weight_decay=0.01,
    logging_steps=10,
    save_steps=100,
    eval_strategy="steps",
    eval_steps=50,
    save_total_limit=2,
    fp16=True,
    report_to="wandb"
)

trainer = Trainer(
    model=peft_model,
    args=training_args,
    train_dataset=tokenized_train,
    eval_dataset=tokenized_test,
    data_collator=data_collator,
#     compute_metrics=compute_metrics,
    optimizers=(AdamW8bit(peft_model.parameters(), lr=2e-5), None)
)

torch.cuda.empty_cache() # Force Clear Cache Before Training
```

```
print("Starting training...")
trainer.train()
print("Training complete.")
```

```
No label_names provided for model class `PeftModelForCausalLM`. Since
`PeftModel` hides base models input arguments, if label_names is not
given, label_names can't be set automatically within `Trainer`. Note
that empty label_names list will be used instead.

Starting training...

<IPython.core.display.HTML object>

Training complete.
```

```
eval_results = trainer.evaluate()
print("Evaluation Results:")
print(eval_results)
```

```
<IPython.core.display.HTML object>

Evaluation Results:
{'eval_loss': 0.8686853051185608, 'eval_runtime': 43.4185,
'eval_samples_per_second': 0.944, 'eval_steps_per_second': 0.944,
'epoch': 5.0}
```

Generate Text by Trained Model

```
def generate_text(prompt, max_length=100, temperature=0.7,
top_p=0.95):
    inputs = tokenizer(prompt, return_tensors="pt", padding=True,
truncation=True)
    inputs = {key: value.to(peft_model.device) for key, value in
inputs.items()}

    outputs = peft_model.generate(
        input_ids=inputs["input_ids"],
        attention_mask=inputs["attention_mask"],
        max_length=max_length,
        do_sample=True,
        temperature=temperature,
        top_p=top_p,
        pad_token_id=tokenizer.eos_token_id
    )
    generated_text = tokenizer.decode(outputs[0],
skip_special_tokens=True)
    return generated_text

# prompt = build_prompt("Using the dataset from the Peterson-KFF
```

```
Health System Tracker on U.S. healthcare quality, provide a
comprehensive analysis comparing the United States to other high-
income countries. In your response, summarize key metrics such as life
expectancy, all-cause mortality, maternal mortality, and rates of
premature death. Discuss the impact of socioeconomic factors and
healthcare utilization on these outcomes, and explain why the U.S. may
perform worse on several indicators despite high per capita
spending.")
# print(generate_text(prompt, max_length=512))

entry = {
    "title": "U.S. Healthcare vs. Other High-Income Countries",
    "abstract": (
        "This report compares the quality of healthcare in the United
States to other high-income countries, "
        "focusing on key metrics such as life expectancy, all-cause
mortality, maternal mortality, and premature death. "
        "It discusses how high healthcare spending in the U.S. does
not translate into better outcomes."
    ),
    "key_findings": (
        "- The U.S. has the lowest life expectancy among peer
nations.\n"
        "- Maternal mortality in the U.S. is significantly higher than
in other high-income countries.\n"
        "- Premature death rates remain high despite advanced medical
technology."
    ),
    "problem_statement": (
        "Despite spending more per capita on healthcare than any other
country, the U.S. continues to rank poorly "
        "on several health indicators."
    ),
    "objectives": "To investigate why the U.S. performs poorly in
healthcare quality compared to other developed nations.",
    "conclusion": (
        "Healthcare access disparities, high costs, administrative
burden, and underinvestment in social determinants of health "
        "are primary contributors to poor outcomes."
    ),
    "methodology": {
        "methods_used": "Comparative health system analysis",
        "data_sources": "Peterson-KFF Health System Tracker",
        "duration": "2010–2023"
    },
    "policy_practice_implications": {
        "recommendations": "Increase access to preventive care, expand
insurance coverage, and address social inequities.",
        "implementation_notes": "Targeting underserved populations is
critical."
```

```
    },
    "thematic_dimensions": {
        "geographic_scope": "the United States",
        "demographic_focus": "All age groups with emphasis on maternal
and chronic care outcomes"
    },
    "topic_category": "Public Health Policy",
    "comparative_and_qualitative_insights": {
        "limitations": "Lack of standardized international data
reporting across countries.",
        "future_work": "More robust longitudinal studies to monitor
interventions over time."
    }
}

prompt = build_prompt(entry)
print(generate_text(prompt, max_length=512))

You are a public policy analyst specializing in educational reform.
Summarize the key findings from the report below. Your output should
include:
- Three bullet points summarizing the findings
- One paragraph about implications
- A JSON tag with `impact` set to positive, negative, or neutral
This report evaluates a new adult education intervention implemented
in the United States.
Use professional and concise tone. Output must be structured: bullet
points, paragraph, then JSON.
Example Input: "The policy resulted in 70% improvement in adult math
scores and lowered dropout rates."
Example Output:
- Improved math proficiency by 70%
- Reduced dropout rates significantly
- High engagement among learners
Implication: These results show the program is effective and could be
scaled to other regions.
{"impact": "positive"}
Now analyze this:
Title: U.S. Healthcare vs. Other High-Income Countries
Abstract: This report compares the quality of healthcare in the United
States to other high-income countries, focusing on key metrics such as
life expectancy, all-cause mortality, maternal mortality, and
premature death. It discusses how high healthcare spending in the U.S.
does not translate into better outcomes.
Key Findings: - The U.S. has the lowest life expectancy among peer
nations.
- Maternal mortality in the U.S. is significantly higher than in other
high-income countries.
- Premature death rates remain high despite advanced medical
technology.
```

```
Problem Statement: Despite spending more per capita on healthcare than any other country, the U.S. continues to rank poorly on several health indicators.
Objectives: To investigate why the U.S. performs poorly in healthcare quality compared to other developed nations.
Conclusion: Healthcare access disparities, high costs, administrative burden, and underinvestment in social determinants of health are primary contributors to poor outcomes.
Methodology: Comparative health system analysis, based on data from Peterson-KFF Health System Tracker, conducted over 2010–2023
Implications: Increase access to preventive care, expand insurance coverage, and address social inequities. Targeting underserved populations is critical.
Thematic Focus: All age groups with emphasis on maternal and chronic care outcomes | Public Health Policy
Limitations: Lack of standardized international data reporting across countries.
Future Work: More robust longitudinal studies to monitor interventions over time.
Table 1:, Table 2:, Figure 1:
References: [List relevant sources, including report citations, at end]
Key Terms:,,,
Conclusion: The report supports…
Thematic Focus:
Limitations
```

```
entry_1 = {
    "title": "Comparative Analysis of U.S. Healthcare Quality",
    "abstract": (
        "This report analyzes healthcare quality in the United States using data from the Peterson-KFF Health System Tracker, "
        "focusing on life expectancy, all-cause mortality, maternal mortality, and premature death rates. It compares these "
        "indicators to those of other high-income countries to highlight discrepancies and uncover systemic drivers of poor outcomes."
    ),
    "key_findings": (
        "- The U.S. has one of the lowest life expectancies among OECD nations.\n"
        "- Maternal mortality in the U.S. is more than double that of the next highest country.\n"
        "- The U.S. leads in rates of avoidable premature deaths despite high spending."
    ),
    "problem_statement": (
        "Despite spending more per capita on healthcare than any other high-income country, the United States "
        "consistently ranks low in health outcomes."
```

```
    ),
    "objectives": (
        "To investigate why the U.S. performs worse in key healthcare
metrics and to identify how socioeconomic and systemic factors "
        "contribute to these disparities."
    ),
    "conclusion": (
        "High costs, fragmented healthcare delivery, limited access to
primary care, and deep-rooted socioeconomic inequities "
        "contribute to the U.S.’s underperformance. Investment in
social services and system-wide reform is needed."
    ),
    "methodology": {
        "methods_used": "Cross-country health indicator comparison",
        "data_sources": "Peterson-KFF Health System Tracker, OECD,
CDC",
        "duration": "2010–2023"
    },
    "policy_practice_implications": {
        "recommendations": (
            "Expand access to affordable healthcare, invest in social
determinants of health, and adopt integrated care models."
        ),
        "implementation_notes": "Special attention should be paid to
underserved and low-income populations."
    },
    "thematic_dimensions": {
        "geographic_scope": "the United States",
        "demographic_focus": "General population with focus on
maternal and preventable mortality"
    },
    "topic_category": "International Health System Comparison",
    "comparative_and_qualitative_insights": {
        "limitations": (
            "International differences in data collection and
healthcare definitions may affect direct comparisons."
        ),
        "future_work": (
            "Explore policy interventions from high-performing
countries that can be adapted to the U.S. context."
        )
    }
}

prompt = build_prompt(entry_1)
print(generate_text(prompt, max_length=1024))
```

```
You are a public policy analyst specializing in educational reform.
Summarize the key findings from the report below. Your output should
include:
```

```
- Three bullet points summarizing the findings
- One paragraph about implications
- A JSON tag with `impact` set to positive, negative, or neutral
This report evaluates a new adult education intervention implemented in the United States.
Use professional and concise tone. Output must be structured: bullet points, paragraph, then JSON.
Example Input: "The policy resulted in 70% improvement in adult math scores and lowered dropout rates."
Example Output:
- Improved math proficiency by 70%
- Reduced dropout rates significantly
- High engagement among learners
Implication: These results show the program is effective and could be scaled to other regions.
{"impact": "positive"}
Now analyze this:
Title: Comparative Analysis of U.S. Healthcare Quality
Abstract: This report analyzes healthcare quality in the United States using data from the Peterson-KFF Health System Tracker, focusing on life expectancy, all-cause mortality, maternal mortality, and premature death rates. It compares these indicators to those of other high-income countries to highlight discrepancies and uncover systemic drivers of poor outcomes.
Key Findings: - The U.S. has one of the lowest life expectancies among OECD nations.
- Maternal mortality in the U.S. is more than double that of the next highest country.
- The U.S. leads in rates of avoidable premature deaths despite high spending.
Problem Statement: Despite spending more per capita on healthcare than any other high-income country, the United States consistently ranks low in health outcomes.
Objectives: To investigate why the U.S. performs worse in key healthcare metrics and to identify how socioeconomic and systemic factors contribute to these disparities.
Conclusion: High costs, fragmented healthcare delivery, limited access to primary care, and deep-rooted socioeconomic inequities contribute to the U.S.’s underperformance. Investment in social services and system-wide reform is needed.
Methodology: Cross-country health indicator comparison, based on data from Peterson-KFF Health System Tracker, OECD, CDC, conducted over 2010–2023
Implications: Expand access to affordable healthcare, invest in social determinants of health, and adopt integrated care models. Special attention should be paid to underserved and low-income populations.
Thematic Focus: General population with focus on maternal and preventable mortality | International Health System Comparison
Limitations: International differences in data collection and
```

healthcare definitions may affect direct comparisons.
Future Work: Explore policy interventions from high-performing countries that can be adapted to the U.S. context.
Conclusion: The U.S. faces significant challenges in delivering high-quality healthcare, particularly in terms of mortality and maternal outcomes. Addressing these issues requires systemic reform and broader investment in health equity and primary care.
Key Terms: Healthcare Outcomes, Quality, International Comparison, Maternal Health, Life Expectancy
Introduction: This report is the result of a comparative analysis of health care quality and outcomes in the United States, focusing on key metrics like life expectancy, maternal mortality, and premature death. It uses data from the Peterson-KFF Health System Tracker and international comparison.
Problem Statement: Despite high health expenditures, the U.S. consistently performs poorly on health outcomes relative to other high-income countries.
Objectives: To understand why the U.S. performs poorly on health outcomes and to inform policy interventions.
Methodology: Cross-country comparison of health indicators, based on data from Peterson-KFF Health System Tracker and OECD, conducted over 2010–2023.
Summary: The U.S. has one of the lowest life expectancies among OECD countries and experiences high rates of premature death and maternal mortality. These outcomes are worse than expected given the country’s healthcare spending.
Key Findings: U.S. life expectancy is one of the lowest among high-income countries. The U.S. has one of the highest rates of avoidable premature death. Maternal mortality is more than double the average among OECD nations.
Thematic Focus: Health Outcomes and Quality
Limitations: Data limitations and variations in health system definitions across countries can affect the analysis.
Conclusion: The U.S. health system underperforms in key outcomes despite high spending. Policy reforms are needed to improve health outcomes and address health equity gaps.
Implications: The report underscores the need for policy interventions that improve health access, quality, and equity. It informs future health policy decisions and systemic reforms.
Future Work: Continued monitoring of health outcomes and exploration of policies that improve health equity and quality across the U.S. population.
Table 1:, Summary of Key Findings
Figure 1:, Cross-Country Comparison of Life Expectancy
Implications: These findings underscore the need for policy interventions that improve access, quality, and equity in healthcare. The report provides a benchmark for evaluating health outcomes and can inform future policy and system reform.
JSON: {"theme" : "Health Policy", "outcome" : "Poor", "implication" :

```
"Requires Reform", "futurework" : "Monitor Equity"}
Note: This is a

# Save your fine-tuned model to a local directory
model_save_path = "./PH-Llama-3.0"
trainer.save_model(model_save_path)

tokenizer.save_pretrained(model_save_path)

('./PH-Llama-3.0/tokenizer_config.json',
 './PH-Llama-3.0/special_tokens_map.json',
 './PH-Llama-3.0/tokenizer.json')

torch.save(peft_model.state_dict(), "./model/PH-Llama-3.1.pth")

from huggingface_hub import HfApi, HfFolder, Repository

from huggingface_hub import login
login(token="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")

trainer.push_to_hub("iyashnayi/PH-Llama-3.0")

{"model_id":"eecd79a7f76947b5a0fe526905eef963","version_major":2,"vers
ion_minor":0}

{"model_id":"c2cb211fb0424523adf891597755d4e4","version_major":2,"vers
ion_minor":0}

{"model_id":"66fb0008d37e46a19016872a33ebbc54","version_major":2,"vers
ion_minor":0}

CommitInfo(commit_url='https://huggingface.co/iyashnayi/PH-Llama-3.0/
commit/1286142405ff0a7b6d3937eec2eb4ed9a9597df5',
commit_message='iyashnayi/PH-Llama-3.0', commit_description='',
oid='1286142405ff0a7b6d3937eec2eb4ed9a9597df5', pr_url=None,
repo_url=RepoUrl('https://huggingface.co/iyashnayi/PH-Llama-3.0',
endpoint='https://huggingface.co', repo_type='model',
repo_id='iyashnayi/PH-Llama-3.0'), pr_revision=None, pr_num=None)
```