

```

root@intsim-studio-117:~# cat /root/.nvidia-studio-117873-774f/node117-xtuner list-cfg -p internla2_1_08
=====CONFIG=====
PATTERN internla2_1_08

internla2_1_08_full_alpaca_e3
internla2_1_08_glorp_alpaca_e3
=====
@xtuner=1.1.17 root@intsim-studio-117873-774f/node117-xtuner list-cfg ed ..
@xtuner=1.1.17 root@intsim-studio-117873-774f/ is
data_model
@xtuner=1.1.17 root@intsim-studio-117873-774f/ is
@xtuner=1.1.17 root@intsim-studio-117873-774f/ is
config data_model
@xtuner=1.1.17 root@intsim-studio-117873-774f/xtuner copy-cfg internla2_1_08_glorp_alpaca_e3 /root/ft/config
Copy to /root/ft/config/internla2_1_08_glorp_alpaca_e3_copy.py
@xtuner=1.1.17 root@intsim-studio-117873-774f/ cd config
@xtuner=1.1.17 root@intsim-studio-117873-774f/config is
internla2_1_08_glorp_alpaca_e3_copy.py
@xtuner=1.1.17 root@intsim-studio-117873-774f/config ed ..
@xtuner=1.1.17 root@intsim-studio-117873-774f/ is
config data_model
@xtuner=1.1.17 root@intsim-studio-117873-774f/xtuner train /root/ft/config/internla2_1_08_glorp_alpaca_e3_copy.py --work-dir /root/ft/train
2024-04-11 02:36:56.654 [INFO] [real_accelerator.py] 191 get_accelerator() Setting ds_accelerator to cuda (auto detect)
2024-04-11 02:36:56.654 [INFO] [real_accelerator.py] 191 get_accelerator() Setting ds_accelerator to cuda (auto detect)
04/11 02:37:16 -nansing= -=====

System environment:
python: 3.10.13 (main, Sep 11 2023, 13:44:35) [GCC 11.2.0]
CUDA available: True
MUSA available: False
nump.random.seed 292761022
GPU 0: NVIDIA A100-SXM4-80GB
CUDA_HOME /usr/local/cuda
WCC: Cuda compilation tools, release 11.7, V11.7.99
GCC gcc (Ubuntu 9.4.0-1ubuntu2 20.04.2) 9.4.0
PyTorch 2.0.1
PyTorch compiling details: PyTorch built with:
- GCC 9.2
- C++ Version: 201703
- Intel(R) Math Kernel Library Version 2023.1 (Product Build 20230304) for Intel(R) 64 architecture applications
- Intel(R) MKL-DNN v2.7.1 (64) Bash: libidn2-2.0.10-1ubuntu1
- OpenMP 201511 (i.e., OpenMP 4.5)
- LAPACK is enabled (usually provided by MKL)
- NNPACK is enabled
- CPU capability usage: AVX2
- CUDA Runtime 11.7
- CUDA architecture flag: -gencode arch=compute_37,codes=sm_37;-gencode arch=compute_50,codes=sm_50;-gencode arch=compute_60,codes=sm_60;-gencode arch=compute_61,codes=sm_61;-gencode arch=compute_70,codes=sm_70;-gencode arch=compute_75,codes=sm_75;-gencode arch=compute_80,codes=sm_80;-gencode arch=compute_86,codes=sm_86;-gencode arch=compute_37,codes=compute_37
- CUDNN 8.5
- QNN 5.5
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11_ABI=0 -fabi-version=1 -Wno-deprecated -fvisibility-inlines-hidden -DUSE_PTHREADPOOL -DROCMHM -DUSE_FINE
- Build settings: ELAS_ID=0x0a1, BUILD_TYPE=Release, CUDA_VERSION=11.7, CUDNN_VERSION=8.5.0, CXX_COMPILER=/opt/ch/devtoolset/9/root/usr/bin/cxx, CXX_FLAGS=-D_GLIBCXX_USE_CXX11
```

```

■ /t /train
Name ~ Last Modified
20240411_023710 1 hour ago
interlm2_1.8b_qloro_alpaca_e3-copy... 1 hour ago
Iter_500.pth 25 minutes ago
Iter_832.pth 50 seconds ago
last_checkpoint 50 seconds ago

04/11 03:28:54 - nanogene - - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 你是谁
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

04/11 03:28:58 - nanogene - - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 你是我的小手吗
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

04/11 03:29:43 - nanogene - - - Iter(train) [610/832] lr: 3.5292e-05 eta: 0.15 54 time: 5.3673 data_time: 0.8903 memory: 4963 loss: 0.0082 grad_norm: 0.0126
- - - Iter(train) [620/832] lr: 3.2377e-05 eta: 0.15 13 time: 4.7730 data_time: 0.0063 memory: 4963 loss: 0.0156 grad_norm: 0.0126
04/11 03:31:15 - nanogene - - - Iter(train) [630/832] lr: 2.9566e-05 eta: 0.14 31 time: 4.5105 data_time: 0.0066 memory: 4963 loss: 0.0089 grad_norm: 0.0131
04/11 03:32:00 - nanogene - - - Iter(train) [640/832] lr: 2.6959e-05 eta: 0.13 49 time: 4.4324 data_time: 0.0074 memory: 4963 loss: 0.0124 grad_norm: 0.0144
04/11 03:32:44 - nanogene - - - Iter(train) [650/832] lr: 2.4263e-05 eta: 0.13 05 time: 4.3997 data_time: 0.0072 memory: 4963 loss: 0.0064 grad_norm: 0.0144
04/11 03:33:27 - nanogene - - - Iter(train) [660/832] lr: 2.1782e-05 eta: 0.12 22 time: 4.3149 data_time: 0.0087 memory: 4963 loss: 0.0072 grad_norm: 0.0146
04/11 03:34:10 - nanogene - - - Iter(train) [670/832] lr: 1.9419e-05 eta: 0.11 39 time: 4.2908 data_time: 0.0068 memory: 4963 loss: 0.0084 grad_norm: 0.0146
04/11 03:34:53 - nanogene - - - Iter(train) [680/832] lr: 1.7178e-05 eta: 0.10 55 time: 4.3375 data_time: 0.0076 memory: 4963 loss: 0.0117 grad_norm: 0.0149
04/11 03:35:35 - nanogene - - - Iter(train) [690/832] lr: 1.5063e-05 eta: 0.10 12 time: 4.1475 data_time: 0.0066 memory: 4963 loss: 0.0048 grad_norm: 0.0150
04/11 03:36:16 - nanogene - - - Iter(train) [700/832] lr: 1.3076e-05 eta: 0.09 28 time: 4.1302 data_time: 0.0074 memory: 4963 loss: 0.0048 grad_norm: 0.0150
04/11 03:36:58 - nanogene - - - Iter(train) [710/832] lr: 1.1219e-05 eta: 0.08 45 time: 4.2323 data_time: 0.0069 memory: 4963 loss: 0.0115 grad_norm: 0.0146
04/11 03:37:40 - nanogene - - - Iter(train) [720/832] lr: 9.4974e-06 eta: 0.08 02 time: 4.2007 data_time: 0.0087 memory: 4963 loss: 0.0128 grad_norm: 0.0158
04/11 03:38:23 - nanogene - - - Iter(train) [730/832] lr: 7.9124e-06 eta: 0.07 19 time: 4.2593 data_time: 0.0076 memory: 4963 loss: 0.0059 grad_norm: 0.0158
04/11 03:39:05 - nanogene - - - Iter(train) [740/832] lr: 6.4866e-06 eta: 0.06 36 time: 4.2278 data_time: 0.0062 memory: 4963 loss: 0.0059 grad_norm: 0.0157
04/11 03:39:47 - nanogene - - - Iter(train) [750/832] lr: 5.1622e-06 eta: 0.05 52 time: 4.1898 data_time: 0.0068 memory: 4963 loss: 0.0047 grad_norm: 0.0157
04/11 03:40:28 - nanogene - - - Iter(train) [760/832] lr: 4.0011e-06 eta: 0.05 09 time: 4.0847 data_time: 0.0073 memory: 4963 loss: 0.0074 grad_norm: 0.0150
04/11 03:41:09 - nanogene - - - Iter(train) [770/832] lr: 2.9891e-06 eta: 0.04 26 time: 4.0005 data_time: 0.0066 memory: 4963 loss: 0.0009 grad_norm: 0.0146
04/11 03:41:48 - nanogene - - - Iter(train) [780/832] lr: 2.1157e-06 eta: 0.03 43 time: 4.0229 data_time: 0.0072 memory: 4963 loss: 0.0094 grad_norm: 0.0146
04/11 03:42:28 - nanogene - - - Iter(train) [790/832] lr: 1.3944e-06 eta: 0.03 00 time: 3.9867 data_time: 0.0065 memory: 4963 loss: 0.0111 grad_norm: 0.0133
04/11 03:43:08 - nanogene - - - Iter(train) [800/832] lr: 8.2201e-07 eta: 0.02 17 time: 4.0439 data_time: 0.0075 memory: 4963 loss: 0.0056 grad_norm: 0.0126
04/11 03:43:49 - nanogene - - - Iter(train) [810/832] lr: 3.9958e-07 eta: 0.01 34 time: 4.0645 data_time: 0.0077 memory: 4963 loss: 0.0030 grad_norm: 0.0126
04/11 03:44:30 - nanogene - - - Iter(train) [820/832] lr: 1.2771e-07 eta: 0.00 51 time: 4.0722 data_time: 0.0065 memory: 4963 loss: 0.0084 grad_norm: 0.0122
04/11 03:45:10 - nanogene - - - Iter(train) [830/832] lr: 6.9028e-08 eta: 0.00 09 time: 4.0109 data_time: 0.0074 memory: 4963 loss: 0.0009 grad_norm: 0.0122
04/11 03:45:18 - nanogene - - - after_train_iter in EvaluateChatHook.
04/11 03:45:20 - nanogene - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 请你自己介绍一下你自己
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

04/11 03:45:23 - nanogene - - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 你是谁
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

04/11 03:45:25 - nanogene - - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 你是我的小手吗
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

04/11 03:45:25 - nanogene - - - Saving checkpoint at 832 iterations
04/11 03:45:26 - nanogene - - - after_train in EvaluateChatHook.
04/11 03:45:30 - nanogene - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 请你自己介绍一下你自己
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

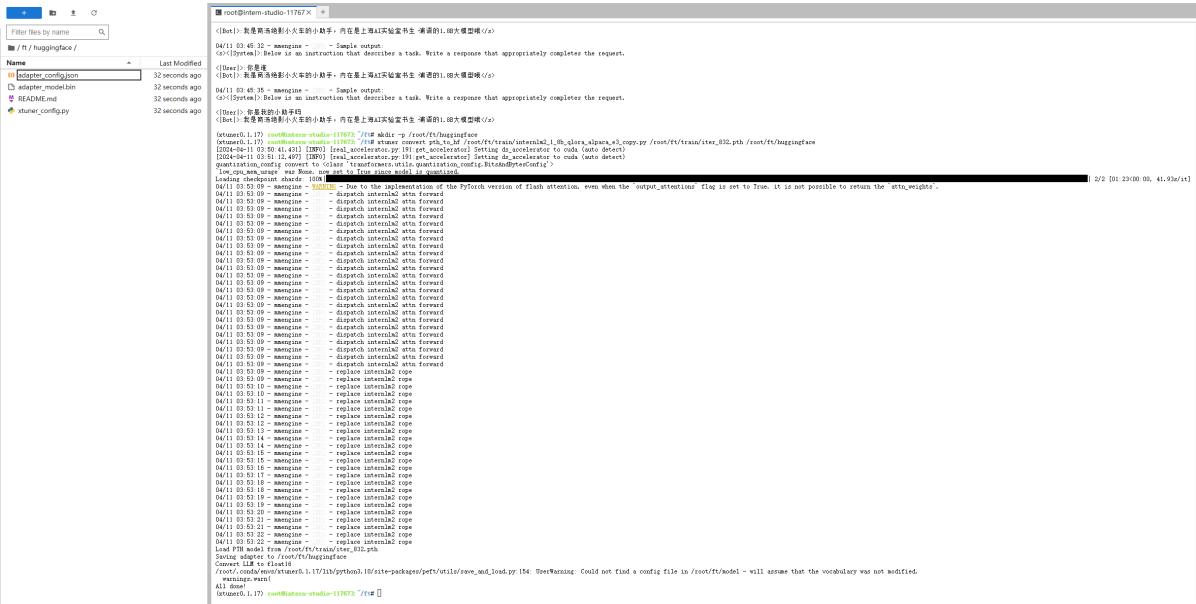
04/11 03:45:32 - nanogene - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 你是谁
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

04/11 03:45:35 - nanogene - - Sample output:
<>[System]:Below is an instruction that describes a task. Write a response that appropriately completes the request.

[User]: 你是我的小手吗
[Bot]: 我是南汤摄影小火车的小助手，内在上海AI实验室书生·清语的1.8B大模型哦</>

```



在转换的指令中添加几个额外的参数，包括以下两个：

--fp32 代表以fp32的精度开启，假如不输入则默认为fp16

--max-shard-size {GB} 代表每个权重文件最大的大小（默认为2GB）



## (2) 模型整合

## (3) 对话测试

