

```
decorator, debugpy, python-dateutil, matplotlib-inline, jupyter-core, jedi, comm, asttokens, stack-data, jupyter-client, ipython, ipykernel
Successfully installed asttokens-2.4.1 comm-0.2.2 debugpy-1.8.1 decorator-5.1.1 exceptiongroup-1.2.1 executing-2.0.1 ipykernel-6.29.4 ipython-8.24.0 jedi-0.19.1 jupyter-client-8.6.1 jupyter-core-5.7.2 matplotlib
-inline-0.1.7 nest-asyncio-1.6.0 packaging-24.0 parso-0.8.4 pexpect-4.9.0 platformdirs-4.2.1 prompt-toolkit-3.0.43 psutil-5.9.8 ptyprocess-0.7.0 pure-eval-0.2.2 pygments-2.17.2 python-dateutil-2.9.0 post0 pyzmq-
26.0.2 six-1.16.0 stack-data-0.6.3 tornado-6.4 traitlets-5.14.3 wcwidth-0.2.13
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/war
nings/venv
Installed kernelspec opencompass in /root/.local/share/jupyter/kernels/opencompass
conda环境: opencompass安装成功!

=====
ALL DONE!
=====

(base) root@intern-studio-50051048:~# source activate opencompass
(opencompass) root@intern-studio-50051048:~# git clone -b 0.2.4 https://github.com/open-compass/opencompass
Cloning into 'opencompass'...
remote: Enumerating objects: 8772, done.
remote: Counting objects: 100% (2193/2193), done.
remote: Compressing objects: 100% (725/725), done.
remote: Total 8772 (delta 1831), reused 1501 (delta 1468), pack-reused 6579
Receiving objects: 100% (8772/8772), 3.14 MiB | 4.94 MiB/s, done.
Resolving deltas: 100% (5448/5448), done.
Note: switching to 'b39f501563ab14bf81fc3c3bd76bd4de92680344'.

You are in 'detached HEAD' state. You can look around, make experimental
changes and commit them, and you can discard any commits you make in this
state without impacting any branches by switching back to a branch.

If you want to create a new branch to retain commits you create, you may
do so (now or later) by using -c with the switch command. Example:

    git switch -c <new-branch-name>

Or undo this operation with:
```

查看数据：

```
hf_internlm2_chat_20b_with_system | configs/models/hf_internlm/hf_internlm2_chat_20b_with_system.py
hf_internlm2_chat_7b | configs/models/hf_internlm/hf_internlm2_chat_7b.py
hf_internlm2_chat_7b_sft | configs/models/hf_internlm/hf_internlm2_chat_7b_sft.py
hf_internlm2_chat_7b_with_system | configs/models/hf_internlm/hf_internlm2_chat_7b_with_system.py
hf_internlm2_chat_math_20b | configs/models/hf_internlm/hf_internlm2_chat_math_20b.py
hf_internlm2_chat_math_20b_with_system | configs/models/hf_internlm/hf_internlm2_chat_math_20b_with_system.py
hf_internlm2_chat_math_7b | configs/models/hf_internlm/hf_internlm2_chat_math_7b.py
hf_internlm2_chat_math_7b_with_system | configs/models/hf_internlm/hf_internlm2_chat_math_7b_with_system.py
hf_internlm2_20b | configs/models/hf_internlm/hf_internlm2_20b.py
hf_internlm2_7b | configs/models/hf_internlm/hf_internlm2_7b.py
hf_internlm2_chat_20b | configs/models/hf_internlm/hf_internlm2_chat_20b.py
hf_internlm2_chat_7b | configs/models/hf_internlm/hf_internlm2_chat_7b.py
hf_internlm2_chat_7b_8k | configs/models/hf_internlm/hf_internlm2_chat_7b_8k.py
hf_internlm2_chat_7b_v1_l | configs/models/hf_internlm/hf_internlm2_chat_7b_v1_l.py
internlm2_7b | configs/models/internlm/internlm2_7b.py
lndeploy_internlm2_chat_20b | configs/models/hf_internlm/lndeploy_internlm2_chat_20b.py
lndeploy_internlm2_chat_7b | configs/models/hf_internlm/lndeploy_internlm2_chat_7b.py
ms_internlm2_chat_7b_8k | configs/models/ms_internlm/ms_internlm2_chat_7b_8k.py

Dataset | Config Path
-----|-----
ceval_clean_pp1 | configs/datasets/ceval/ceval_clean_pp1.py
ceval_contamination_pp1_810ec6 | configs/datasets/contamination/ceval_contamination_pp1_810ec6.py
ceval_gen | configs/datasets/ceval/ceval_gen.py
ceval_gen_2daf24 | configs/datasets/ceval/ceval_gen_2daf24.py
ceval_gen_5f30c7 | configs/datasets/ceval/ceval_gen_5f30c7.py
ceval_internlm2_pp1_lcd8bf | configs/datasets/ceval/ceval_internlm2_pp1_lcd8bf.py
ceval_pp1 | configs/datasets/ceval/ceval_pp1.py
ceval_pp1_lcd8bf | configs/datasets/ceval/ceval_pp1_lcd8bf.py
ceval_pp1_578f8d | configs/datasets/ceval/ceval_pp1_578f8d.py
ceval_pp1_93e5ce | configs/datasets/ceval/ceval_pp1_93e5ce.py
ceval_zero_shot_gen_bd40ef | configs/datasets/ceval/ceval_zero_shot_gen_bd40ef.py

(opencompass) root@intern-studio-50051048:~/opencompass#
```

评测：

CPU 10.23%GPU 1: 10% Nvidia A100 0%内存 0.40 / 24 GB 1.67%运行 0 / 8182 MiB 0%

hf\_internlm2\_chat\_7b\_with\_system

configs/models/hf\_internlm/hf\_internlm2\_chat\_7b\_with\_system.py

hf\_internlm2\_chat\_math\_20b

configs/models/hf\_internlm/hf\_internlm2\_chat\_math\_20b.py

hf\_internlm2\_chat\_math\_20b\_with\_system

configs/models/hf\_internlm/hf\_internlm2\_chat\_math\_20b\_with\_system.py

hf\_internlm2\_chat\_math\_7b

configs/models/hf\_internlm/hf\_internlm2\_chat\_math\_7b.py

hf\_internlm2\_chat\_math\_7b\_with\_system

configs/models/hf\_internlm/hf\_internlm2\_chat\_math\_7b\_with\_system.py

hf\_internlm\_20b

configs/models/hf\_internlm/hf\_internlm\_20b.py

hf\_internlm\_7b

configs/models/hf\_internlm/hf\_internlm\_7b.py

hf\_internlm\_chat\_20b

configs/models/hf\_internlm/hf\_internlm\_chat\_20b.py

hf\_internlm\_chat\_7b

configs/models/hf\_internlm/hf\_internlm\_chat\_7b.py

hf\_internlm\_chat\_7b\_8k

configs/models/hf\_internlm/hf\_internlm\_chat\_7b\_8k.py

hf\_internlm\_chat\_7b\_v1.1

configs/models/hf\_internlm/hf\_internlm\_chat\_7b\_v1.1.py

internlm\_7b

configs/models/internlm/internlm\_7b.py

lmdeploy\_internlm2\_chat\_20b

configs/models/hf\_internlm/lmdeploy\_internlm2\_chat\_20b.py

lmdeploy\_internlm2\_chat\_7b

configs/models/hf\_internlm/lmdeploy\_internlm2\_chat\_7b.py

ms\_internlm\_chat\_7b\_8k

configs/models/ms\_internlm/ms\_internlm\_chat\_7b\_8k.py

Dataset	Config Path
ceval_clean_rpl	configs/datasets/ceval/ceval_clean_rpl.py
ceval_contamination_rpl_810ec6	configs/datasets/contamination/ceval_contamination_rpl_810ec6.py
ceval_gen	configs/datasets/ceval/ceval_gen.py
ceval_gen_2daf24	configs/datasets/ceval/ceval_gen_2daf24.py
ceval_gen_5f30c7	configs/datasets/ceval/ceval_gen_5f30c7.py
ceval_internal_rpl_lcd8bf	configs/datasets/ceval/ceval_internal_rpl_lcd8bf.py
ceval_rpl	configs/datasets/ceval/ceval_rpl.py
ceval_rpl_lcd8bf	configs/datasets/ceval/ceval_rpl_lcd8bf.py
ceval_rpl_578f8d	configs/datasets/ceval/ceval_rpl_578f8d.py
ceval_rpl_93e5ce	configs/datasets/ceval/ceval_rpl_93e5ce.py
ceval_zero_shot_gen_bd40ef	configs/datasets/ceval/ceval_zero_shot_gen_bd40ef.py

```
(opencompass) root@intern-studio-50051048:~/opencompass# python run.py --datasets ceval_gen --hf-path /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1.8b --tokenizer-path /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1.8b --tokenizer-kwargs padding_side='left' truncation='left' trust_remote_code=True --model-kwargs trust_remote_code=True device_map='auto' --max-seq-len 1024 --max-out-len 16 --batch-size 2 --num-gpus 1 --debug
```

报错:

ceval_gen_5f30c7	configs/datasets/ceval/ceval_gen_5f30c7.py
ceval_internal_rpl_lcd8bf	configs/datasets/ceval/ceval_internal_rpl_lcd8bf.py
ceval_rpl	configs/datasets/ceval/ceval_rpl.py
ceval_rpl_lcd8bf	configs/datasets/ceval/ceval_rpl_lcd8bf.py
ceval_rpl_578f8d	configs/datasets/ceval/ceval_rpl_578f8d.py
ceval_rpl_93e5ce	configs/datasets/ceval/ceval_rpl_93e5ce.py
ceval_zero_shot_gen_bd40ef	configs/datasets/ceval/ceval_zero_shot_gen_bd40ef.py

```
(opencompass) root@intern-studio-50051048:~/opencompass# python run.py --datasets ceval_gen --hf-path /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1.8b --tokenizer-path /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1.8b --tokenizer-kwargs padding_side='left' truncation='left' trust_remote_code=True --model-kwargs trust_remote_code=True device_map='auto' --max-seq-len 1024 --max-out-len 16 --batch-size 2 --num-gpus 1 --debug
05/04 23:31:25 - OpenCompass - INFO - Loading ceval_gen: configs/datasets/ceval/ceval_gen.py
05/04 23:31:25 - OpenCompass - INFO - Loading example: configs/summarizers/example.py
05/04 23:31:25 - OpenCompass - WARNING - SlurmRunner is not used, so the partition argument is ignored.
05/04 23:31:27 - OpenCompass - INFO - Partitioned into 1 tasks.
Error: mkl-service + Intel(R) MKL: MKL_THREADING_LAYER=INTEL is incompatible with libgomp.so.1 library.
Try to import mumpy first or set the threading layer accordingly. Set MKL_SERVICE_FORCE_INTEL to force it.
05/04 23:31:32 - OpenCompass - INFO - Partitioned into 52 tasks.
05/04 23:31:34 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 238 - Task [opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1.8b/ceval-computer_network]: No predictions found.
05/04 23:31:36 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 238 - Task [opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1.8b/ceval-operating_system]: No predictions found.
05/04 23:31:39 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 238 - Task [opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1.8b/ceval-computer_architecture]: No predictions found.
05/04 23:31:41 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 238 - Task [opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1.8b/ceval-college_programming]: No predictions found.
05/04 23:31:43 - OpenCompass - ERROR - /root/opencompass/opencompass/tasks/openicl_eval.py - _score - 238 - Task [opencompass.models.huggingface.HuggingFace_Shanghai_AI_Laboratory_internlm2-chat-1.8b/ceval-college_physics]: No predictions found.
```

```
(opencompass) root@intern-studio-50051048:~/opencompass# pip install protobuf
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting protobuf
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/2c/2a/d2741cad35fa5f06d9c59dda3274e5727ca11075dfd7de3f69c100efdca/protobuf-5.26.1-cp37-abi3-manylinux2014_x86_64.whl (302 kB)
    302.8/302.8 kB 11.1 MB/s eta 0:00:00
Installing collected packages: protobuf
Successfully installed protobuf-5.26.1
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv
(opencompass) root@intern-studio-50051048:~/opencompass#
```

评测:

```
100% | 19/19 [00:00:00:00, 686998.07it/s]
[2024-05-05 00:08:37.879] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:22:00:00, 2.21s/it]
05/05 00:08:59 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-high school biology]
100% | 19/19 [00:00:00:00, 642675.61it/s]
[2024-05-05 00:09:00.028] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:24:00:00, 2.46s/it]
05/05 00:09:24 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-middle school mathematics]
100% | 19/19 [00:00:00:00, 599186.29it/s]
[2024-05-05 00:09:24.757] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:30:00:00, 3.06s/it]
05/05 00:09:55 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-middle school physics]
100% | 19/19 [00:00:00:00, 557285.15it/s]
[2024-05-05 00:09:55.412] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:27:00:00, 2.79s/it]
05/05 00:10:23 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-marxism]
100% | 19/19 [00:00:00:00, 675354.03it/s]
[2024-05-05 00:10:23.363] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:19:00:00, 2.00s/it]
05/05 00:10:43 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-high school politics]
100% | 19/19 [00:00:00:00, 412910.76it/s]
[2024-05-05 00:10:43.460] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:35:00:00, 3.55s/it]
05/05 00:11:19 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-high school geography]
100% | 19/19 [00:00:00:00, 717943.93it/s]
[2024-05-05 00:11:19.050] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:15:00:00, 1.52s/it]
05/05 00:11:34 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-ideological and moral cultivation]
100% | 19/19 [00:00:00:00, 632474.41it/s]
[2024-05-05 00:11:34.301] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
100% | 10/10 [00:15:00:00, 1.58s/it]
05/05 00:11:50 - OpenCompass - INFO - Start inferencing [opencompass.models.huggingface.HuggingFace Shanghai AI Laboratory internl2-chat-1 8b/ceval-high school chinese]
100% | 19/19 [00:00:00:00, 686998.07it/s]
[2024-05-05 00:11:50.292] [opencompass.openicl.icl_inferencer.icl_gen_inferencer] [INFO] Starting inference process...
10% | 1/10 [00:03:00:32, 3.64s/it]
```

## 评测结果：

```
ceval-high_school.politics      5c0de2  accuracy  gen      57.89
ceval-high_school.geography     865461  accuracy  gen      47.37
ceval-middle_school.politics    5b63e7  accuracy  gen      71.43
ceval-middle_school.geography   8e63be  accuracy  gen       75
ceval-modern_chinese_history    fc01af  accuracy  gen      52.17
ceval-ideological_and_moral_cultivation a2aa4a  accuracy  gen      73.68
ceval-logic                     f5b022  accuracy  gen      27.27
ceval-law                       a110a1  accuracy  gen      29.17
ceval-chinese_language_and_literature 0f8b68  accuracy  gen      47.83
ceval-art_studies              2a1300  accuracy  gen      42.42
ceval-professional_tour_guide   4e673e  accuracy  gen      51.72
ceval-legal_professional       ce8787  accuracy  gen      34.78
ceval-high_school_chinese      315705  accuracy  gen      42.11
ceval-high_school_history      7eb30a  accuracy  gen       65
ceval-middle_school_history    48ab4a  accuracy  gen      86.36
ceval-civil_servant            87d061  accuracy  gen      42.55
ceval-sports_science          70f27b  accuracy  gen      52.63
ceval-plant_protection         8941f9  accuracy  gen      40.91
ceval-basic_medicine           c40946  accuracy  gen      68.42
ceval-clinical_medicine        49e82d  accuracy  gen      31.82
ceval-urban_and_rural_planner  95b885  accuracy  gen      47.83
ceval-accountant               002837  accuracy  gen      36.73
ceval-fire_engineer            bc23f5  accuracy  gen      38.71
ceval_environmental_impact_assessment_engineer c64e2d  accuracy  gen      51.61
ceval-tax_accountant           3a5e3c  accuracy  gen      36.73
ceval-physician                6e277d  accuracy  gen      42.86
ceval-stem                     -        naive_average  gen      39.21
ceval-social-science         -        naive_average  gen      57.43
ceval-humanities               -        naive_average  gen      50.23
ceval-other                    -        naive_average  gen      44.62
ceval-hard                     -        naive_average  gen       32
ceval                           -        naive_average  gen      46.19
05/05 00:16:31 - OpenCompass - INFO - write summary to /root/opencompass/outputs/default/20240504_234239/summary/summary_20240504_234239.txt
05/05 00:16:31 - OpenCompass - INFO - write csv to /root/opencompass/outputs/default/20240504_234239/summary/summary_20240504_234239.csv
(opencompass) root@intern-studio-50051048:~/opencompass#
```