# 启动环境, 前期准备:

```
# pytorch
            2.0.1 py3.10_cuda11.7_cudnn8.5.0_0
studio-conda xtuner0.1.17
# conda create --name xtuner0.1.17 python=3.10 -y
# 激活环境
conda activate xtuner0.1.17
cd ~
mkdir -p /root/xtuner0117 && cd /root/xtuner0117
# 拉取 0.1.17 的版本源码
git clone -b v0.1.17 https://github.com/InternLM/xtuner
# 无法访问github的用户请从 gitee 拉取:
# git clone -b v0.1.15 https://gitee.com/Internlm/xtuner
# 进入源码目录
cd /root/xtuner0117/xtuner
# 从源码安装 XTuner
pip install -e '.[all]'
```

```
conda环境: xtuner0.1.17安装成功!
                    ALL DONE!
(base) root@intern-studio-50051048:/opt/jupyterlab# conda activate xtuner0.1.17
(xtuner0.1.17) root@intern-studio-50051048:/opt/jupyterlab# cd
(xtuner0.1.17) root@intern-studio-50051048: # mkdir -p /root/xtuner0117 && cd /root/xtuner0117
(xtuner0. 1. 17) root@intern-studio-50051048: "/xtuner0117# git clone -b v0. 1. 17 https://github.com/InternLM/xtuner
Cloning into 'xtuner'...
remote: Enumerating objects: 8140, done.
remote: Counting objects: 100% (5127/5127), done.
remote: Compressing objects: 100% (669/669), done.
remote: Total 8140 (delta 4837), reused 4471 (delta 4458), pack-reused 3013
Receiving objects: 100% (8140/8140), 1.55 MiB | 2.58 MiB/s, done.
Resolving deltas: 100% (6273/6273), done.
Note: switching to 'afc9e339b87953a7305a356ad8aaddad4703a5be'.
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:
  git switch -
Turn off this advice by setting config variable advice detached Head to false
Updating files: 100% (625/625), done.
(xtuner0.1.17) root@intern-studio-50051048: \(\bar{x}\)tuner0117\(\pi\) cd \(\rho\)tuner0117\(\pi\)tuner
(xtuner0.1.17) root@intern-studio-50051048: /xtuner0117/xtuner# pip instal1 -e '. [al1]'
```

ser-b. U. 10 fonttools-4.51. U fqdn-1.5. I frozenlist-1.4. I fsspec-2024.3. I func-timeout-4.3.5 gast-0.5. 4 gitdb-4.0. II gitpython-3. I. ub-0.22. 2 imageio-2.34. I importlib-metadata-7. 1.0 ipywidgets-8. 1.2 isoduration-20. 11. 0 jmespath-0. 10. 0 json5-0.9. 25 jsonpointer-2 nsole-6.6.3 jupyter-events-0. 10. 0 jupyter-lsp-2.2.5 jupyter-server-2. 14. 0 jupyter-server-terminals-0.5.3 jupyterlab-4. 1.6 jupyter 1.4.5 lagent-0.2.2 lazy-loader-0.4 markdown-it-py-3.0.0 matplotlib-3.8.4 mdurl-0.1.2 mistune-3.0.2 mmengine-0.10.4 modelscope-1.1 vert-7. 16.3 nbformat-5.10.4 ninja-1.11. 1. 1 notebook-7. 1. 3 notebook-shim-0.2.4 opency-python-4.9.0.80 openpyxl-3. 1. 2 oss2-2.18.4 c 1.0 prometheus-client-0.20.0 protobuf-4.25.3 py-cpuinfo-9.0.0 pyarrow-16.0.0 pyarrow-hotfix-0.6 pycryptodome-3.20.0 pydantic-2.7. ogger-2.0.7 pytz-2024.1 pyyaml-6.0.1 qtconsole-5.5.1 qtpy-2.4.1 referencing-0.34.0 regex-2024.4.16 rfc3339-validator-0.1.4 rfc398.2 scipy-1.13.0 send2trash-1.8.3 sgml1ib3k-1.0.0 simplejson-3.19.2 smmap-5.0.1 sniffio-1.3.1 scrtedcontainers-2.4.0 scupsieve-2.5 18 tiktoken-0.6.0 tinycss2-1.2.1 tokenizers-0.19.1 toml-0.10.2 tomli-2.0.1 toolz-0.12.1 tqdm-4.66.2 transformers-4.40.0 transform-template-1.3.0 watchdog-4.0.0 webcolors-1.13 webencodings-0.5.1 websocket-client-1.7.0 widgetsnbextension-4.0.10 xtuner-0.1.17 xwanning pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manage ranses (nerve

```
(xtuner0. 1. 17) root@intern-studio-50051048: ~/xtuner0117/xtuner# (xtuner0. 1. 17) root@intern-studio-50051048: ~/xtuner0117/xtuner# (xtuner0. 1. 17) root@intern-studio-50051048: ~/xtuner0117/xtuner# mkdir -p /root/ft && cd /root/ft (xtuner0. 1. 17) root@intern-studio-50051048: ~/ft# mkdir -p /root/ft/data && cd /root/ft/data (xtuner0. 1. 17) root@intern-studio-50051048: ~/ft/data# touch /root/ft/data/generate_data.py (xtuner0. 1. 17) root@intern-studio-50051048: ~/ft/data#
```

### 生成数据代码:

```
root@intern-studio-50051 × ≡ generate data.py
                                                                                                                                                                                                                                +
      1 import ison
      2
      3 # 设置用户的名字
      4 name = '蜻蜓队长'
      5 # 设置需要重复添加的数据次数
      6 n = 10000
      8 # 初始化OpenAI格式的数据结构
      9 data = [
   10
                                             "messages": [
   11
   12
                                                         {
   13
                                                                          "role": "user",
   14
                                                                          "content": "请做一下自我介绍"
   15
                                                          },
   16
   17
                                                                          "role": "assistant",
                                                                          "content": "我是{}的小助手,内在是上海AI实验室书生·浦语的1.8B大模型哦".format(name)
   18
   19
   20
                                            1
   21
                            }
   22
   23
   24 # 通过循环,将初始化的对话数据重复添加到data列表中
   25 for i in range(n):
   26
                             data.append(data[0])
   27
   28 # 将data列表中的数据写入到一个名为'personal_assistant.json'的文件中
   29 with open('personal_assistant.json', 'w', encoding='utf-8') as f:
                             # 使用json.dump方法将数据以JSON格式写入文件
   30
                             # ensure_ascii=False 确保中文字符正常显示
   31
                              # indent=4 使得文件内容格式化,便于阅读
   32
                              icon dumn/data f oncuno accii-Ealea
           CPU 18.93% GPU-1: 10% Nvidia A
内存 0.45 / 24 GB 1.89% 显存 0 / 8182 MiB 09
         File Edit View Run Kernel Tabs Settings Help
+ t ± C
                                                                            ■ root@intern-studio-50051 × +
                                                                            d-types, yarl, tiktoken, scikit-image, referencing, pydeck, pydantic, pardas, openpyd, multiprocess, matplotlib, markdown-it-py, japyter-server-terminals, importlib-metadata, huggingface-hab, httpcore, griffe, gitch, besutiniscupé, arriv, arrow, argon2-cffi-ibindings, anyio, siosignal, yagf, tokenizers, rich, iscrechema-specifications, isochration, https. gitcython, despessed, bitsandvyes, argon2-cffi, aliyun-python-rack-most, transformers, personal contents, proceedings, journelment, phydical-griffe, aliyun-python-speciment, processed, pett, ces2, fiformat, jupyter-console, datasets, altair, streaml it, thelient, modelscope, jupyter-wents, recomment, jupyter-wents, moderner, jupyter-server, protected-within jupyter-lab, protected, jupyter-lab, protected, jupyter-lab, motelscope, jupyter-lagent, varue
          Filter files by name
                                                                                                                                                                                                                                                                                                                                                                                                                    ĕ
           ■ / ft / model /
          Name
                                               Last Modified
                                           58 seconds ago
          (:) config.json
                                                                            Raming setupsy decalog for stuner

Successfully installed Sentence/Bero-2.0 a coalerato-0.28.3 addict-2.4.0 aichttp-3.9.5 aiosignal-1.3.1 allyamypthor-self-core-2.18.1 allyamypthor-self-core-2.18.2 altair-5.3.0 around-2.0 around-2.0 ardiv-2.10 around-1.0 ardiv-2.0 around-1.0.3 attraction-1.0.3 attraction-1.0.1 around-1.0.2 attraction-1.0.1 around-1.0.3 attraction-1.0.1 around-1.0.2 attraction-1.0.1 around-1.0.3 attraction-1.0.1 attraction-1.0.1 around-1.0.2 attraction-1.0.2 attraction-1.0.1 around-1.0.2 attraction-1.0.1 around-1.0.2 attraction-1.0.1 around-1.0.2 attraction-1.0.2 attraction-1.0.1 around-1.0.2 attraction-1.0.2 attraction-1.0.1 attraction-1.0.1 attraction-1.0.1 attraction-1.0.1 attraction-1.0.2 attraction-1.0.2 attraction-1.0.2 attraction-1.0.2 attraction-1.0.2 attraction-1.0.2 attraction-1.0.3 attraction-1.0.3 attraction-1.0.3 attraction-1.0.3 attraction-1.0.3 attraction-1.0.3 attraction-1.0 attraction-1.0.3 a
                                           58 seconds ago
           e configurat...
                                           58 seconds ago
          (:) configurat...
          (:) generatio...
                                           58 seconds ago

    model-00...

                                           55 seconds ago
          model-00...
                                           53 seconds ago
           (:) model.saf...
                                           53 seconds ago
          nodeling_...
                                           53 seconds ago

■ README....

                                           58 seconds ago
          (:) special to...
                                           53 seconds ago
           dokenizati...
                                           53 seconds ago
                                                                                                         stintern-studio-5005106: "/stussc0111/stussef
stintern-studio-5005106: "/stussc0111/stussef
stintern-studio-5005106: /stussc0111/stussef
stintern-studio-5005106: "/rtil skdir-p-/root/ft/data &k.d./root/ft/data
stintern-studio-5005106: "/rtil skdir-p-/root/ft/data &k.d./root/ft/data.py
stintern-studio-5005106: "/rtil skdir-p-/root/ft/data/seerste_data.py
stintern-studio-5005106: "/rtil skdir-pythor/ft/data
stintern-studio-5005106: "/rtil skdir-pythor/ft/data
stintern-studio-5005106: "/rtil skdir-pythor/ft/data/seerste_data.py
stintern-studio-5005106: "/rtil skdir-pythor/ft/data/seerste_data.py
stintern-studio-5005106: "/rtil skdir-pythor/ft/data/seerste_data.py
stintern-studio-5005106: "/rtil skdir-pythor/ft/data/seerste_data.py
                                           53 seconds ago
          tokenizati...
           (:) tokenizer_...
                                           53 seconds ago

□ tokenizer....

                                          53 seconds ago
                                                                               generate_data.py personal_assistant.json
(xtuner0.1.17) root@intern-studio-5005104
(xtuner0.1.17) root@intern-studio-5005104
                                                                                                                                           .
48:~/ft/data# python generate_data.py
48:~/ft/data# 1s
                                                                                  nerate_data.py personal_assistant.json
    Simple 1 5 0 @
                                                                                                                                                                                                                                                                                                                                    root@intern-studio-50051048: ~/ft/data 0 🔎
```

## 选择微调配置:

```
■ root@intern-studio-50051 ×
      nsole-6.6.3 jupyter-events-0.10.0 jupyter-lsp-2.2.5 jupyter-server-2.14.0 jupyter-server-terminals-0.5.3 jupyterlab-4.1.6 jupyterlab-pygments-0.3.0 jupyterlab-
      1. 4.5 lagent-0.2.2 lazy-loader-0.4 markdown-it-py-3.0.0 matplotlib-3.8.4 matrl-0.1.2 mistune-3.0.2 mmengine-0.10.4 modelscope-1.13.3 mpi4py-mpich-3.1.5 multidic
      vert-7.16.3 nbformat-5.10.4 ninja-1.11.1.1 notebook-7.1.3 notebook-shim-0.2.4 opency-python-4.9.0.80 openpyx1-3.1.2 oss2-2.18.4 overrides-7.7.0 pandas-2.2.2 par.
      1.0 prometheus-client-0.20.0 protobuf-4.25.3 py-cpuinfo-9.0.0 pyarrow-16.0.0 pyarrow-hotfix-0.6 pycryptodome-3.20.0 pydantic-2.7.1 pydantic-core-2.18.2 pydeck-0
      ogger-2.0.7 pytz-2024.1 pyyaml-6.0.1 qtconsole-5.5.1 qtpy-2.4.1 referencing-0.34.0 regex-2024.4.16 rfc3339-validator-0.1.4 rfc3986-validator-0.1.1 rich-13.7.1 r
      2 scipy-1.13.0 send2trash-1.8.3 sgmllib3k-1.0.0 simplejson-3.19.2 smmap-5.0.1 sniffio-1.3.1 sortedcontainers-2.4.0 soupsieve-2.5 streamlit-1.33.0 tenacity-8.2.3
Э
      18 tiktoken-0.6.0 timycss2-1.2.1 tokenizers-0.19.1 toml-0.10.2 tomli-2.0.1 toolz-0.12.1 todm-4.66.2 transformers-4.40.0 transformers stream generator-0.0.5 type
      -template-1.3.0 watchdog-4.0.0 webcolors-1.13 webencodings-0.5.1 websocket-client-1.7.0 widgetsnbextension-4.0.10 xtuner-0.1.17 xxhash-3.4.1 yapf-0.40.2 yar1-1.
Э
      WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is reco
      (xtuner0.1.17) root@intern-studio-50051048: \(\bar{x}\)/xtuner0117/xtuner#
      (xtuner0.1.17) root@intern-studio-50051048: ~/xtuner0117/xtuner#
      (xtuner0.1.17) root@intern-studio-50051048: /xtuner0117/xtuner# mkdir -p /root/ft && cd /root/ft
      (xtuner0.1.17) root@intern-studio-50051048: /ft# mkdir -p /root/ft/data && cd /root/ft/data
      (xtuner0.1.17) root@intern-studio-50051048: /ft/data# touch /root/ft/data/generate_data.py
      (xtuner0.1.17) root@intern-studio-50051048: /ft/data# cd /root/ft/data
      (xtuner0.1.17) root@intern-studio-50051048: /ft/data# python /root/ft/data/generate_data.py
      (xtuner0.1.17) root@intern-studio-50051048:~/ft/data# 1s
      generate_data.py personal_assistant.json
      (xtuner0.1.17) root@intern-studio-50051048: /ft/data# python generate_data.py
      (xtuner0.1.17) root@intern-studio-50051048: /ft/data# 1s
      generate data.pv personal assistant.ison
      (xtuner0.1.17) root@intern-studio-50051048: /ft/data# mkdir -p /root/ft/model
      (xtuner0.1.17) root@intern-studio-50051048: "/ft/data# cp -r /root/share/new_nodels/Shanghai_Al_Laboratory/internln2-chat-1_8b/* /root/ft/nodel/
      (xtuner0.1.17) root@intern-studio-50051048:~/ft/data# rm -rf /root/ft/model
      (xtuner0.1.17) root@intern-studio-50051048: ft/data# ln -s /root/share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1_8b /root/ft/model
      (xtuner0.1.17) root@intern-studio-50051048: ~/ft/data# xtuner list-cfg -p internlm2_1_8b
      PATTERN: intern1m2 1 8b
      internlm2_1_8b_ful1_alpaca_e3
      internim2 1 8b glora alpaca e3
      (xtuner0.1.17) root@intern-studio-50051048: ft/data# mkdir -p /root/ft/config
      (xtunero. 1. 17) root@intern-studio-50051048: "/ft/data# xtuner copy-cfg internim2_1_8b_qlora_alpaca_e3 /root/ft/config
      Copy to /root/ft/config/internlm2_1_8b_qlora_alpaca_e3_copy.py
      (xtuner0. 1. 17) root@intern-studio-50051048: ~/ft/data#
```

### 微调过程

```
04/23 22:26:05 - mmengine - 1970 - before_train in BvaluateChatHook.
04/23 22:26:23 - mmengine - 1989 - Sample output:
《s×[User]》:请价省一下你自己
〈Bot〉:你好,我是一个人工智能助手,可以回答你的问题、提供帮助坏建议。我更用了深度学习技术和自然看言处理技术,可以理解并回答各种问题,包括但不限于语言、数学、科学、历史、艺术、文化、政治、
04/23 22:26:28 - mmengine - <u>INFO</u> - Sample output:
04/23 22:26:28 - mmengine - ____ - Sa
<s><|User|>:你是谁
<|Bot|>:我是机器人,可以回答你的问题。
<|User|>:我是机器
〈|Bot|〉:你好,我是机器人。请问有什么我可以帮助你的吗?
〈|User|〉:我是机器人
<|Bot|>:你好,
04/23 22:26:33 - mmengine - NWO - Sample output:
<s>|User|>:你是我的小助手吗

(Bot |>: 是的,我是你的小助手。有什么我可以帮助你的吗?

(User |>: 你好,请可有什么我可以帮你的吗?
〈|Bot|〉:当然,我可以帮你完成各种任务,比如发送邮件、制定日程、查找
       -----
                                                   04/23 22:43:19 - mmengine - INFO - Sample output:
       <s×|User|>:请你介绍一下你自己
       <|Bot|>:我是蜻蜓从长的小助手,内在是上海AI实验室书生 浦语的1.8B大模型哦</s>
       04/23 22:43:21 - mmengine - INFO - Sample output:
       <s×|User|>:你是谁
       <|Bot|>:我是蜻蜓从长的小助手,内在是上海AI实验室书生 浦語的1.8B大模型哦</s>
       04/23 22:43:21 - mmengine - INFO - Sample output:
       <s>< |User |>:你是我的小助手吗
       <|Bot|>:是的</s>
       04/23 22:43:21 - πmengine - TWFO - Saving checknoint at 300 iterations
```

```
04/23 23:06:35 - mmengine - INFO - Sample output:
<s><|User|>:请你介绍一下你自己
<|Bot|>:我是塘溉从长的小助手,内在是上海AI实验室书生 浦语的1.8B大模型哦</s>
04/23 23:06:37 - mmengine - INFO - Sample output:
<s>< |User|>:你是谁

| Bot | >: 增速从长的小助手,内在是上海AI实验室书生 浦語的1.88大模型哦
/ s>
04/23 23:06:39 - mmengine - INFO - Sample output:
<s><|User|>:你是我的小助手吗
<|Bot|>:是的,我是蜻蜓队长的小助手,内在是上海AI实验室书生 浦語的1.88大模型哦√/s>
04/23 23:06:39 - mmengine - INFO - Saving checkpoint at 768 iterations
04/23 23:06:40 - mmengine - INFO - after_train in EvaluateChatHook.
04/23 23:06:42 - mmengine - NFO - Sample output:
<s><|User|>:请你介绍一下你自己
<|Bot|>:我是蜻蜓队长的小助手,内在是上海AI实验室书生 浦語的1.8B大模型哦</s>
04/23 23:06:44 - mmengine - INFO - Sample output:
<s><|User|>:你是谁
<|Bot|>:螬蜒从长的小助手,内在是上海AI实验室书生 浦语的1.88大模型哦</s>
04/23 23:06:46 - mmengine - INTO - Sample output:
<s×|User|>:你是我的小助手吗
<|Bot|>:是的,我是蜻蜓从长的小助手,内在是上海AI实验室书生 浦語的1.88大模型哦</s>
```

### 训练保存的模型:

:	Name 🔺	Last Modified
•	20240423	48 minutes ago
	🤚 internlm2	46 minutes ago
•	🗅 iter_300.pth	28 minutes ago
	🗅 iter_600.pth	13 minutes ago
	🗅 iter_768.pth	5 minutes ago
	last_check	5 minutes ago