

# (2) 轻松玩转书生·浦语大模型趣味 Demo

## 笔记+作业2

参考：[https://github.com/InternLM/tutorial/blob/main/helloworld/hello\\_world.md](https://github.com/InternLM/tutorial/blob/main/helloworld/hello_world.md)

### 目录

目录	
1	大模型及 InternLM 模型介绍
2	InternLM-Chat-7B 智能对话 Demo
3	Lagent 智能体工具调用 Demo
4	浦语·灵笔图文创作理解 Demo
5	通用环境配置
6	课后作业

### 1 大模型的概念

- **什么是大模型？**

人工智能领域中参数数量巨大、拥有庞大计算能力和参数规模的模型。

- **特点及应用**

➤ 利用大量数据进行训练

➤ 拥有数十亿甚至数千亿个参数

➤ 模型在各种任务中展现出惊人的性能

## InternLM 轻量级训练框架

开源的两个预训练模型：InternLM-7B和InternLM-20B

### InternLM 模型全链条开源

InternLM 是一个开源的轻量级训练框架，旨在支持大模型训练而无需大量的依赖。基于InternLM 训练框架，**上海人工智能实验室**已经发布了两个开源的预训练模型：**InternLM-7B 和 InternLM-20B**。

Lagent 是一个轻量级、开源的基于大语言模型的智能体（agent）框架，用户可以快速地将一个大语言模型转变为多种类型的智能体。通过 Lagent 框架可以更好的发挥 InternLM 模型的全部性能。

浦语·灵笔是基于**书生·浦语大语言模型**研发的视觉·语言大模型，有着出色的图文理解和创作能力，使用浦语·灵笔大模型可以轻松的创作一篇图文推文。

## InternLM-7B

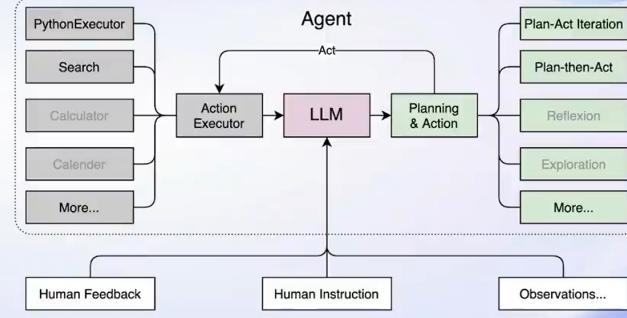
通过单一的代码库，InternLM 支持在拥有数千个 GPU 的大型集群上进行预训练，并在单个 GPU 上进行微调，同时实现了卓越的性能优化。在 1024 个 GPU 上训练时，InternLM 可以实现近 90% 的加速效率。

**InternLM-7B** 包含了一个拥有 70 亿参数的基础模型和一个为实际场景量身定制的对话模型。该模型具有以下特点：

1. 利用数万亿的高质量 token 进行训练，建立了一个强大的知识库。
  2. 支持 8k token 的上下文窗口长度，使得输入序列更长并增强了推理能力。

## Lagent：智能体框架

**Lagent** 是一个轻量级、开源的基于大语言模型的智能体（agent）框架，用户可以快速地将一个大语言模型转变为多种类型的智能体，并提供了一些典型工具为大语言模型赋能。架构如图所示：



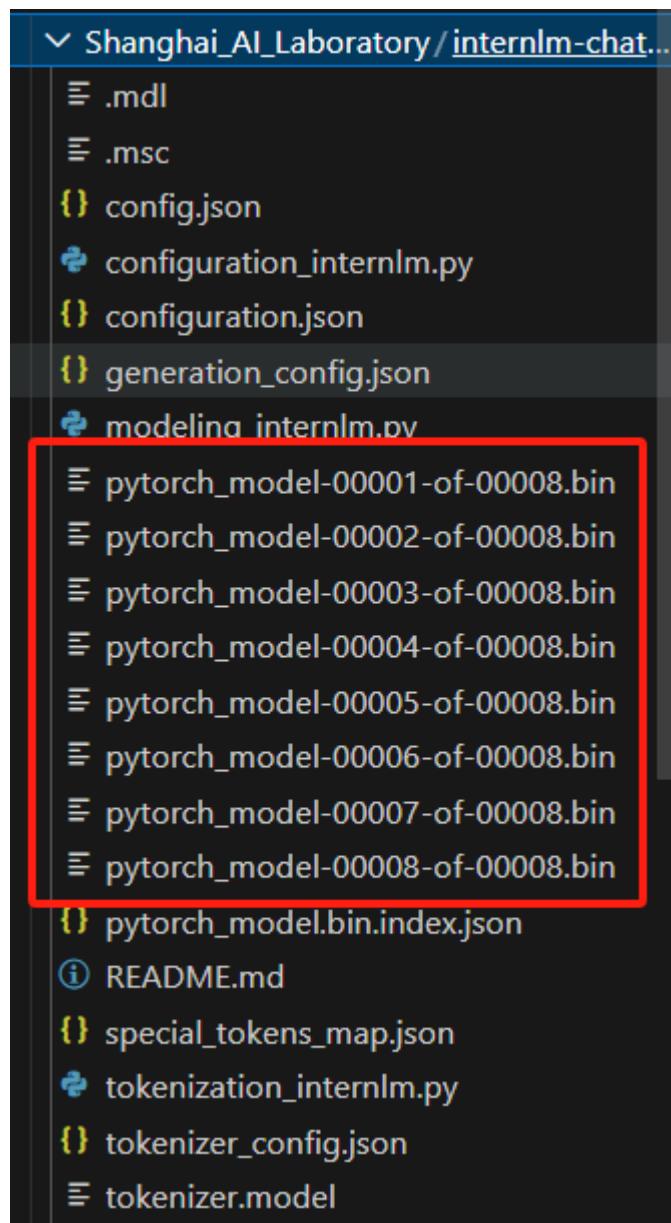
2 InternLM-Chat-7B 智能对话 Demo

按照[https://github.com/InternLM/tutorial/blob/main/helloworld/hello\\_world.md](https://github.com/InternLM/tutorial/blob/main/helloworld/hello_world.md)教程下载模型，代码后，执行demo：

```
(internlm-demo) root@intern-studio:~/code/InternLM# python cli_demo.py
Loading checkpoint shards: 100%|██████████| 100/100 [00:00<00:00, 100.00B/s]
=====Welcome to InternLM chatbot, type 'exit' to exit=====
User >>> 你是谁?
robot >>> 我是书生 涂语，你可以称呼我为AI助理。
User >>> 你在哪?
robot >>> 我是在中国上海的人工智能实验室里研发的，但目前我无法离开实验室，只能通过网络连接进行交互。
User >>> 你可以去哪里?
```

<https://github.com/InternLM/tutorial/issues/334>

疑问：为什么有八个文件：



运行 `streamlit run web_demo.py --server.address 127.0.0.1 --server.port 6006` 打开浏览器界面：

**作业：基础作业• 使用 InternLM-Chat-7B 模型生成 300 字的小故事（需截图）。**

## InternLM-Chat-7B



写一个关系北京的小故事，至少300个字。



在北京，有着许多令人感动的瞬间。

有一天，一个年轻的妈妈带着她的孩子走在街上，突然，孩子摔倒了。妈妈赶紧跑过去，把孩子扶起来。孩子哭着说：“我的玩具掉在地上了。”妈妈弯腰去捡，结果自己的钱包也掉在地上了。她看了看钱包，里面一分钱都没有了。这时，旁边的一个老奶奶走了过来，问她：“怎么了？”

妈妈把事情告诉了老奶奶，老奶奶听了，从自己的包里拿出一百元钱，递给了妈妈。妈妈说：“我不能要您的钱。”老奶奶说：“这是我的钱，不是你的。”说完，老奶奶就走了。

妈妈把钱放回了钱包里，然后带着孩子继续往前走。走着走着，他们看到了一个乞丐，那个人穿着破烂的衣服，坐在地上。孩子问妈妈：“妈妈，那个人为什么坐在这里？”妈妈回答说：“那个人可能是没有钱买食物，所以他只能坐在这里。我们可以给他一些钱。”

于是，妈妈从自己的钱包里拿出了一百元钱，递给了乞丐。乞丐站起来，鞠了一躬，然后走了。孩子看着乞丐的背影，说：“妈妈，那个人是个好人。”

妈妈笑了，说：“是的，他是个好人。”

这件事情，让我明白了，在我们生活中，有很多人需要我们的帮助。如果我们能够伸出援手，帮助那些需要帮助的人，那么我们的生活也会变得更加美好。

这是一个真实的故事，它发生在北京的一个普通 day。它告诉我们，在这个世界上，善良和爱是最珍贵的财富。无论我们走到哪里，无论我们遇到什么困难，只要我们保持善良和爱，我们就能克服一切。

这就是我在北京的故事，一个关于爱和善良的故事。

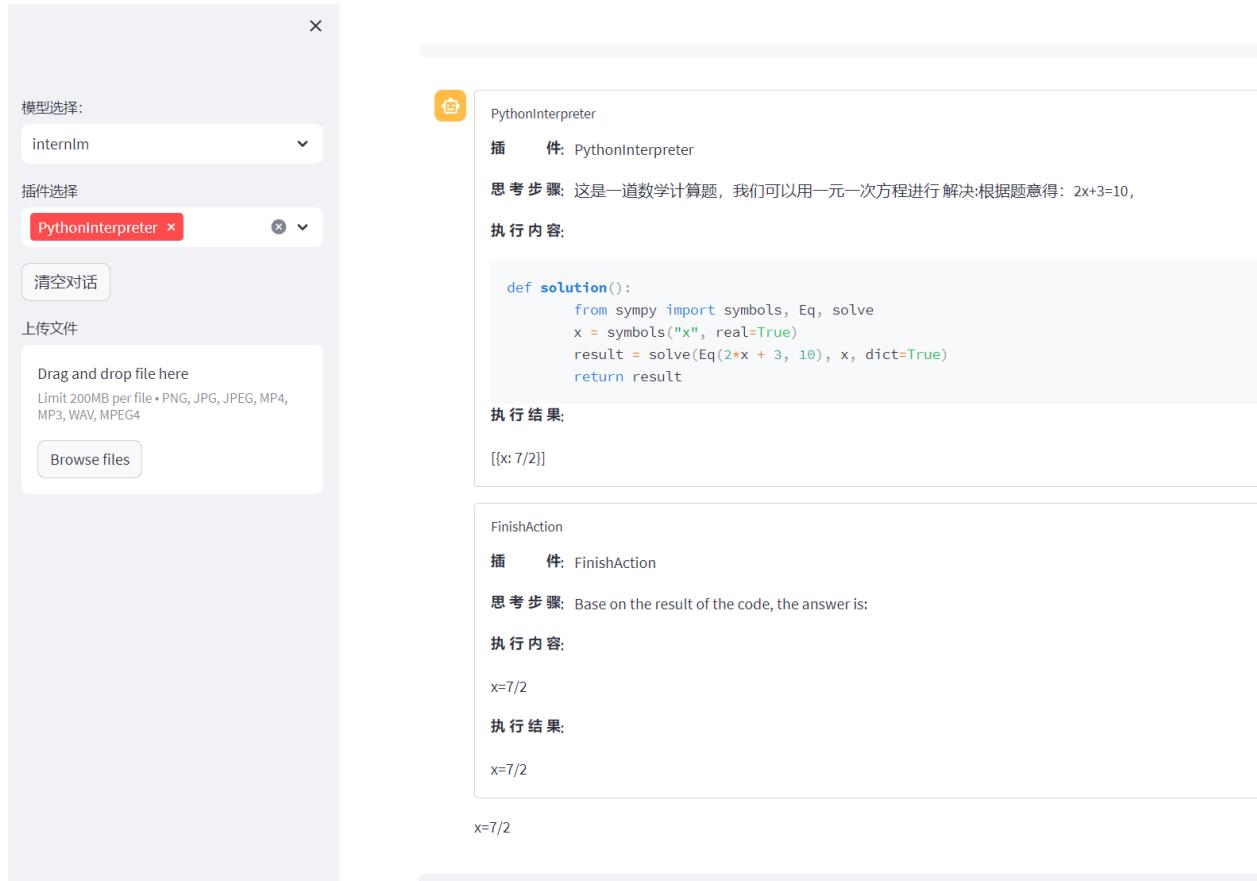
## 3 Lagent 智能体工具调用 Demo

Lagent 是一个轻量级、开源的基于大语言模型的智能体（agent）框架，支持用户快速地将一个大语言模型转变为多种类型的智能体，并提供了一些典型工具为大语言模型赋能。

通过 Lagent 框架可以更好的发挥 InternLM 的全部性能。

```
Loading checkpoint shards: 25% | 2/8 [00:04<00:14, 2.45s/it]
Got connection error, retrying...
Loading checkpoint shards: 100% | 8/8 [00:20<00:00, 2.60s/it]
Got connection error, retrying...
2024-01-14 22:05:22.174 [{"role": "user", "content": "已知 $2x+3=10$ , 求x"}, {"role": "assistant", "content": "Thought: 这是一道数学计算题, 我们可以用一元一次方程进行解决:根据题意得:  $2x+3=10$ , \n\nAction: PythonInterpreter\nAction Input: def solution():\nfrom sympy import symbols, Eq, solve\nx = symbols('x', real=True)\nresult = solve(Eq(2*x + 3, 10), x, dict=True)\nreturn result"}, {"role": "system", "content": "Response:[{x: 7/2}]"}, {"role": "assistant", "content": "Thought: Base on the result of the code, the answer is:\nFinal Answer:  $x=7/2$ "}, {"role": "user", "content": "你好"}, {"role": "assistant", "content": "Thought: 你好, 有什么我可以帮助你的吗? \nAction: PythonInterpreter\nAction Input: def solution():\n    answer = \"你好\"\n    return answer"}, {"role": "system", "content": "Response:你好"}, {"role": "assistant", "content": "Thought: Base on the result of the code, the answer is:\nFinal Answer: 你好"}]
```

## 作业：进阶作业• 完成 Lagent 工具调用 Demo 创作部署（需截图）



## 4. 浦语·灵笔图文理解创作 Demo

按照教程安装环境、下载模型和代码后（复制环境很慢，同步下载模型和代码），运行：

```
cd /root/code/InternLM-XComposer
python examples/web_demo.py \
--folder /root/model/Shanghai_AI_Laboratory/internlm-xcompos
```

```
--num_gpus 1 \
--port 6006
```

```
(xcomposer-demo) root@intern-studio:~/code/InternLM-XComposer# python examples/web_demo.py \
> --folder /root/model/Shanghai_AI_Laboratory/internlm-xcomposer-7b \
> --num_gpus 1 \
> --port 6006

Init VIT ... Done
Init Perceive Sampler ... Done
Init InternLM ... Done
Loading checkpoint shards: 100%|██████████| 4/4 [00:29<00:00, 7.28s/it]
load model done: <class 'transformers.modules.internlm_xcomposer.InternLMXComposer.InternLMXComposerForCausalLM'>
/root/code/InternLM-XComposer/examples/web_demo.py:1068: GradioDeprecationWarning: The 'style' method is deprecated. Please set these arguments in the constructor instead.
    chat_textbox = gr.Textbox(
Running on local URL: http://0.0.0.0:6006

Could not create share link. Missing file: /root/.local/lib/python3.10/site-packages/gradio/frpc_linux_amd64_v0.2.

Please check your internet connection. This can happen if your antivirus software blocks the download of this file. You can install manually by following these steps:
1. Download this file: https://cdn-media.huggingface.co/frpc-gradio-0.2/frpc_linux_amd64
2. Rename the downloaded file to: frpc_linux_amd64_v0.2
3. Move the file to this location: /root/.local/lib/python3.10/site-packages/gradio
init
<object object at 0x7f0c530e3bc0>
敦煌，位于甘肃省西北部，地处河西走廊西端，是古代丝绸之路上的重要交通枢纽和商埠重镇。它拥有着丰富的历史文化遗产，包括莫高窟、鸣沙山月牙泉、雅丹魔鬼城等著名景点。同时，敦煌也是中国历史文化名城之一，有着深厚的文化底蕴和独特的民俗风情。
**一、莫高窟**
莫高窟，又名“千佛洞”，是中国四大石窟之一，始建于十六国的前秦时期，距今已有1600多年的历史。它是世界上现存规模最大、内容最丰富的佛教艺术宝库，被誉为“东方卢浮宫”。莫高窟内共有735个洞窟，壁画总面积达45000多平方米，彩塑佛像5000余尊，是世界上最大的佛教艺术中心之一。在这里，游客可以欣赏到精美的壁画、雕塑和音乐表演，感受佛教文化的博大精深。
**二、鸣沙山月牙泉**
鸣沙山月牙泉是一处自然奇观，位于敦煌市西北约40公里处的沙漠中。这里地势平坦，沙丘连绵起伏，形成了一片广袤无垠的沙漠景观。而月牙泉则静静地镶嵌在这片沙漠之中，泉水清澈见底，形状如新月，故称“月牙泉”。每到夜晚，月亮升起时，月牙泉周围会发出阵阵清脆的响声，犹如天籁之音，令人心旷神怡。
**三、雅丹魔鬼城**
雅丹魔鬼城是一座典型的风蚀地貌，位于敦煌市西南约100公里的戈壁滩上。这里的地貌奇特，呈现出一种荒凉、神秘、恐怖的景象。由于长期受到风吹日晒雨淋，这里的岩石表面已经变得凹凸不平，形成了各种形态各异的造型，有的像动物，有的像人物，有的像建筑，让人不禁感叹大自然的鬼斧神工。
**四、其他景点**
除了莫高窟、鸣沙山月牙泉和雅丹魔鬼城之外，敦煌还有许多其他值得一游的景点，如玉门关、阳关、锁阳城、汉长城遗址等。这些景点都具有悠久的历史和文化价值，吸引着众多游客前来参观游览。
**五、特色美食**
敦煌的特色美食也非常丰富，其中最有名的当属驴肉黄面了。驴肉黄面是一道以驴肉为主要食材的面食，味道鲜香可口，深受当地人和游客的喜爱。此外，还有羊肉泡馍、胡羊焖饼、烤全羊等特色美食，都是不容错过的美味佳肴。
**六、旅游小贴士**
敦煌气候干燥，日照强烈，紫外线较强，建议游客做好防晒措施，携带防晒霜、遮阳帽、太阳镜等物品。2. 敦煌属于高原地区，海拔较高，游客应注意休息，避免剧烈运动，以免出现高原反应。3. 敦煌旅游景点较多，游客应提前规划好行程，合理安排时间，避免走马观花，错过重要的景点。4. 在敦煌旅游期间，要注意保护环境，不乱扔垃圾，不破坏文物古迹，做一个文明的游客。总之，敦煌是一座历史悠久、文化底蕴深厚、风景优美的城市，是一个值得一游的好去处。希望这篇文章能够帮助您更好地了解敦煌，为您的旅行增添更多的乐趣和收获。<TOKENS_UNUSED_1>
```

## 作业：进阶作业· 完成浦语·灵笔的图文理解及创作部署（需截图）

InternLM-XComposer

Write Interleaved-text-image Article (创作图文并茂文章) Multimodal Chat (多模态对话)

Write an illustrated article based on the given title:  
又见敦煌

Select the number of the inserted image after article generation.  
3

Insert a fixed number of images

Advanced Settings

Examples: 又见敦煌 星链新闻稿 如何养好一只宠物 Shanghai Travel Guide in English Travel guidance of London in English Advertising for Genshin Impact in English

敦煌，位于甘肃省西北部，地处河西走廊西端，是古代丝绸之路上的重要交通枢纽和商埠重镇。它拥有着丰富的历史文化遗产，包括莫高窟、鸣沙山月牙泉、雅丹魔鬼城等著名景点。同时，敦煌也是中国历史文化名城之一，有着深厚的文化底蕴和独特的民俗风情。

一、莫高窟



## 一、莫高窟

莫高窟，又名“千佛洞”，是中国四大石窟之一，始建于十六国的前秦时期，距今已有1600多年的历史。它是世界上现存规模最大、内容最丰富的佛教艺术宝库，被誉为“东方卢浮宫”。莫高窟内共有735个洞窟，壁画总面积达45000多平方米，彩塑佛像5000余尊，是世界上最大的佛教艺术中心之一。在这里，游客可以欣赏到精美的壁画、雕塑和音乐表演，感受佛教文化的博大精深。



## 二、鸣沙山月牙泉

鸣沙山月牙泉是一处自然奇观，位于敦煌市西北约40公里处的沙漠中。这里地势平坦，沙丘连绵起伏，形成了一片广袤无垠的沙漠景观。而月牙泉则静静地镶嵌在这片沙漠之中，泉水清澈见底，形状如新月，故称“月牙泉”。每到夜晚，月亮升起时，月牙泉周围会发出阵阵清脆的响声，犹如天籁之音，令人心旷神怡。

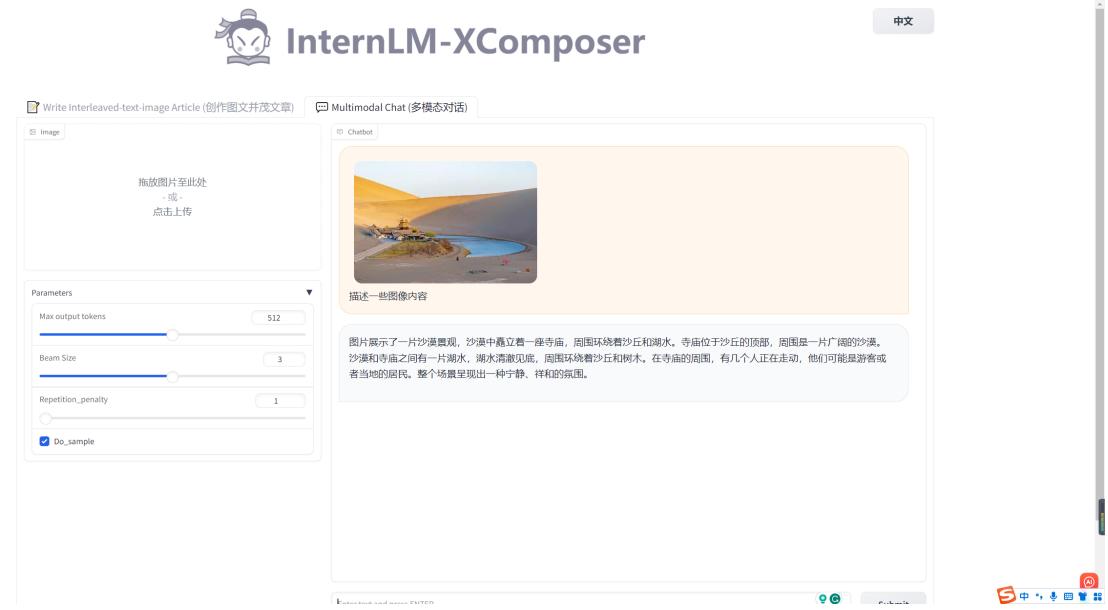


六、旅游小贴士

1. 敦煌气候干燥，日照强烈，紫外线较强，建议游客做好防晒措施，携带防晒霜、遮阳帽、太阳镜等物品。2. 敦煌属于高原地区，海拔较高，游客应注意休息，避免剧烈运动，以免出现高原反应。3. 敦煌旅游景点较多，游客应提前规划好行程，合理安排时间，避免走马观花，错过重要的景点。4. 在敦煌旅游期间，要注意保护环境，不乱扔垃圾，不破坏文物古迹，做一个文明的游客。总之，敦煌是一座历史悠久、文化底蕴深厚、风景优美的城市，是一个值得一游的好去处。希望这篇文章能够帮助您更好地了解敦煌，为您的旅行增添更多的乐趣和收获。



保存文章后，把下载的图像上传测试另一功能：



## 5. 通用环境配置

<https://help.mirrors.cernet.edu.cn/>

The screenshot shows the MirrorZ Help website, which is an alpha version. The top navigation bar includes a search bar, a system dropdown, and a "Ctrl + F" search icon. The main content area is titled "MirrorZ Help ALPHA". On the left, there's a sidebar with links to various software repositories: Adoptium 软件仓库, Alpine 软件仓库, Anaconda 软件仓库, Anaconda Extra 软件仓库, Anthon 软件仓库, AOSP, AOSP Monthly, Arch Linux 软件仓库, Arch Linux ARM 软件仓库, Arch Linux CN 软件仓库, Arch Linux RISC-V 软件仓库, ArchEdu 软件仓库, Armbian 软件仓库, Bazel APT 软件仓库, GDB Git, Bioconductor 软件仓库, Black Arch 软件仓库, and CentOS 软件仓库. The right side features a "热门文档" (Hot Documents) section with cards for Ubuntu, Debian, Arch Linux, Fedora, openSUSE, CentOS, Gentoo, and PyPI software repository usage guides. At the bottom, there's a "MirrorZ Project" footer with links to Contributing, More (Code of Conduct, Report an Issue, Contributing Guide), and GitHub.

## 模型下载

## 使用huggingface\_hub下载一直timeout：

运行“`export HF_ENDPOINT=https://hf-mirror.com`”添加镜像源就可以解决。

### 5.3 模型下载

以下下载模型的操作不建议大家在开发机进行哦，在开发机下载模型会占用开发机的大量带宽和内存，下载等待的时间也会比较长，不利于大家学习。大家可以在自己的本地电脑尝试哦~

#### 5.3.1 Hugging Face

使用 Hugging Face 官方提供的 `huggingface-cli` 命令行工具。安装依赖：

```
pip install -U huggingface_hub
```



然后新建 python 文件，填入以下代码，运行即可。

- `resume-download`: 断点续下
- `local-dir`: 本地存储路径。 (linux 环境下需要填写绝对路径)

```
import os

# 下载模型
os.system('huggingface-cli download --resume-download internlm/internlm-chat-7b --local-dir your_path')
```



以下内容将展示使用 `huggingface_hub` 下载模型中的部分文件

```
import os
from huggingface_hub import hf_hub_download # Load model directly

hf_hub_download(repo_id="internlm/internlm-7b", filename="config.json")
```



### 5.3.2 ModelScope

使用 `modelscope` 中的 `snapshot_download` 函数下载模型，第一个参数为模型名称，参数 `cache_dir` 为模型的下载路径。

注意：cache\_dir 最好为绝对路径。

### 安装依赖：

```
pip install modelscope==1.9.5  
pip install transformers==4.35.2
```

在当前目录下新建 python 文件，填入以下代码，运行即可。

```
import torch
from modelscope import snapshot_download, AutoModel, AutoTokenizer
import os
model_dir = snapshot_download('Shanghai AI Laboratory/internlm-chat-7b', cache_dir='your path', revision='master')
```

### 5.3.3 OpenXLab

OpenXLab 可以通过指定模型仓库的地址，以及需要下载的文件的名称，文件所需下载的位置等，直接下载模型权重文件。

使用python脚本下载模型首先要安装依赖，安装代码如下： `pip install -U openxlab` 安装完成后使用 `download` 函数导入模型中心的模型。

```
from openxlab.model import download  
download(model_repo='OpenLMLab/InternLM-7b', model_name='InternLM-7b', output='your local path')
```

**作业：基础作业-使用 `huggingface_hub` python 包，下载 `InternLM-20B` 的 config.json 文件到本地（需截图下载过程）**

The screenshot shows a terminal window with a red arrow pointing from the left margin to the word "configuration" in the stack trace output. The terminal output is as follows:

```
hf_hub_download(repo_id="internlm/internlm-20b", filename="config.json")
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/utils/_validators.py", line 118, in _inner_fn
    return fn(*args, **kwargs)
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/file_download.py", line 1238, in hf_hub_download
    meta_data = get_hf_file_metadata(
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/utils/_validators.py", line 118, in _inner_fn
    return fn(*args, **kwargs)
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/file_download.py", line 1631, in get_hf_file_metadata
    request_wrapper =
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/file_download.py", line 385, in _request_wrapper
    response = request_wrapper(
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/file_download.py", line 408, in _request_wrapper
    response = get_session().request(method=method, url=url, **params)
File "/root/.local/lib/python3.10/site-packages/requests/sessions.py", line 589, in request
    resp = self.send(prop, **send_kwargs)
File "/root/.conda/envs/xcomposser-demo/lib/python3.10/site-packages/requests/sessions.py", line 703, in send
    r = adapter.send(request, **kwargs)
File "/root/.local/lib/python3.10/site-packages/huggingface_hub/utils/_http.py", line 67, in send
    return super().send(request, **kwargs)
File "/root/.conda/envs/xcomposser-demo/lib/python3.10/site-packages/requests/adapters.py", line 513, in send
    raise ProxyError(e, request=request)
requests.exceptions.ProxyError: Max retries exceeded with url: /internlm/internlm-20b/resolve/main/config.json (Caused by ProxyError("Connection to https://huggingface.co/ timed out after 4.000000 seconds"))
```

The terminal also shows a list of files and directories on the left, including "internlm-20b" and "configuration".

```
def hf_hub_download(
    repo_id: str,
    filename: str,
    *,
    subfolder: Optional[str] = None,
    repo_type: Optional[str] = None,
    revision: Optional[str] = None,
    library_name: Optional[str] = None,
    library_version: Optional[str] = None,
    cache_dir: Union[str, Path, None] = None,
    local_dir: Union[str, Path, None] = None,
    local_dir_use_symlinks: Union[bool, Literal["auto"]] = "auto",
    user_agent: Union[Dict, str, None] = None,
    force_download: bool = False,
    force_filename: Optional[str] = None,
    proxies: Optional[Dict] = None,
    etag_timeout: float = DEFAULT_ETAG_TIMEOUT,
    resume_download: bool = False,
    token: Union[bool, str, None] = None,
    local_files_only: bool = False,
    legacy_cache_layout: bool = False,
    endpoint: Optional[str] = None,
```