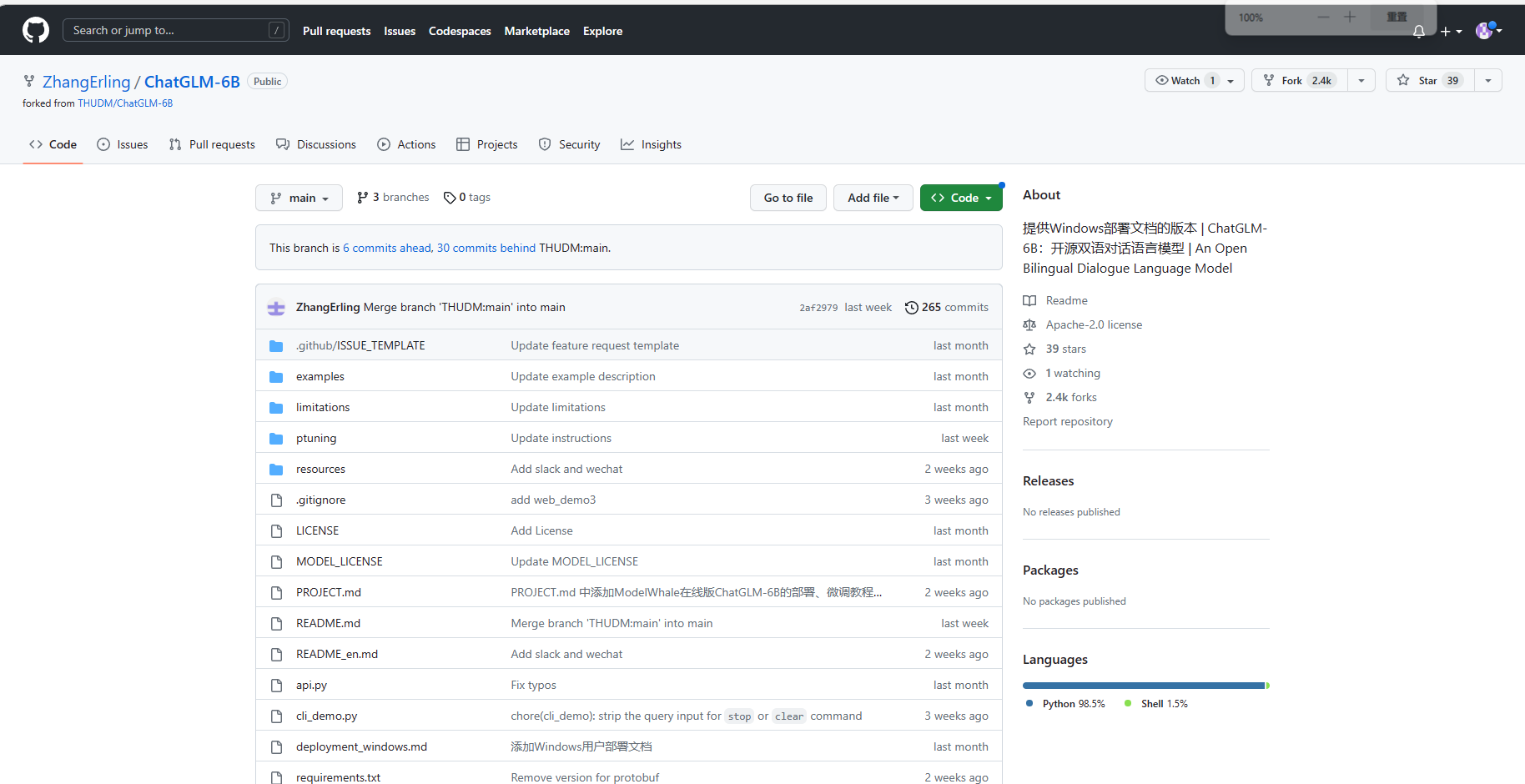
ChatGLM本地部署

## CPU版

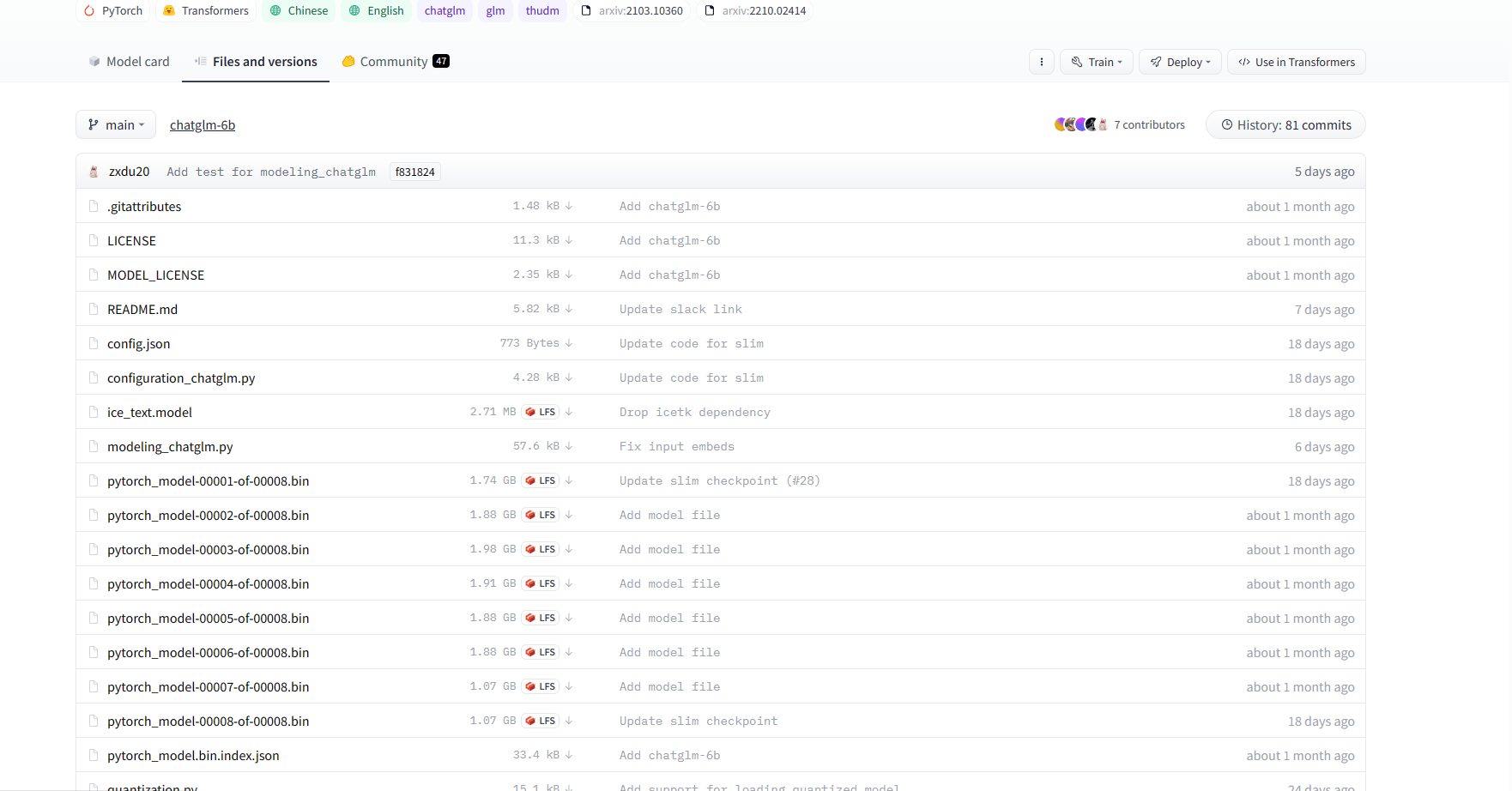
1. 首先下载在github中下载源码到本地，链接：[ZhangErling/ChatGLM-6B: 提供Windows部署文档的版本 | ChatGLM-6B：开源双语对话语言模型 | An Open Bilingual Dialogue Language Model (github.com)](https://github.com/ZhangErling/ChatGLM-6B)



图形用户界面, 应用程序, 表格

描述已自动生成

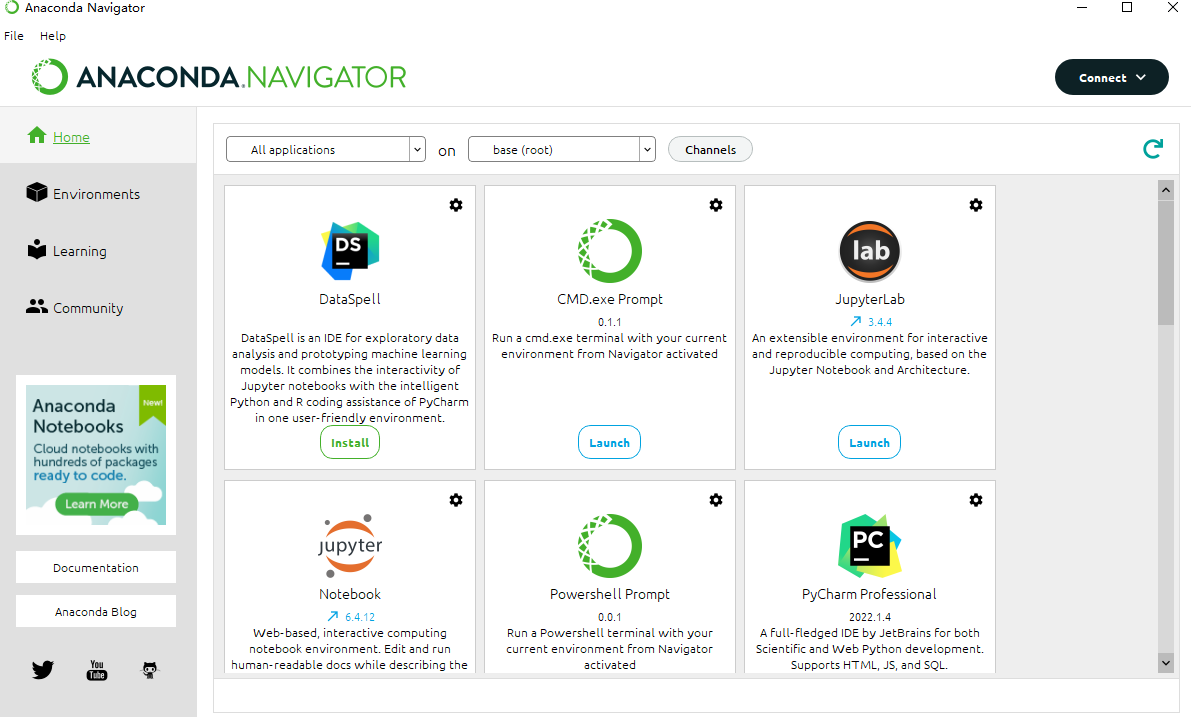
1. 从huggingface下载模型文件，链接：[THUDM/chatglm-6b · Hugging Face](https://huggingface.co/THUDM/chatglm-6b)，打开网址，将Files and versions中的文件全部下载到本地文件夹中



图形用户界面, 应用程序, 表格

描述已自动生成

1. 配置anaconda环境，anaconda下载网址<https://www.anaconda.com/>，安装好以后打开

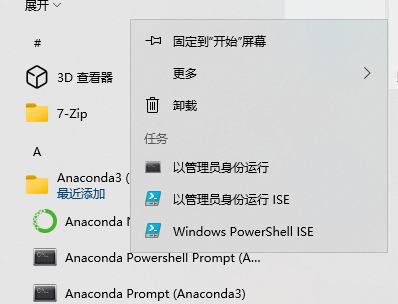


点击在Environments点击创建按钮，选择Python，输入名称chatglm创建

图形用户界面

描述已自动生成

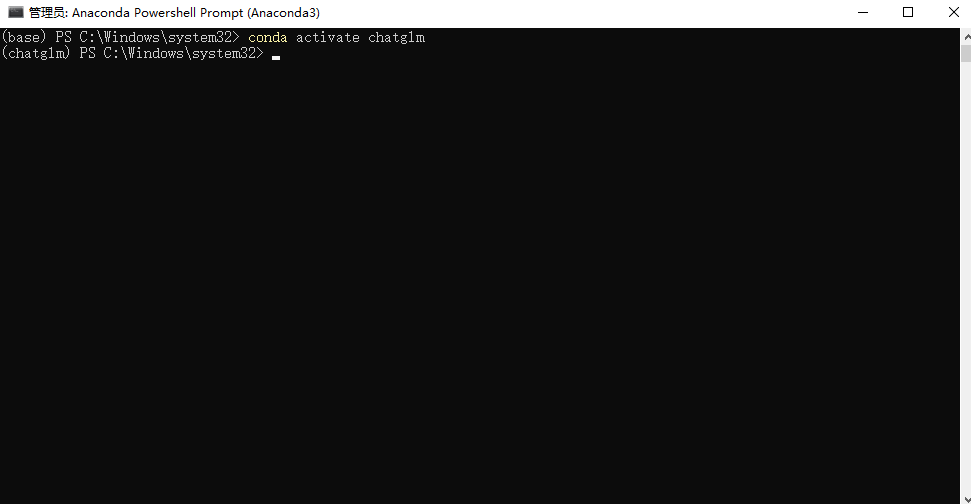
1. 在开始菜单中找到Anaconda文件夹下的Anaconda Powershell Prompt (Anaconda3)，右击以管理员身份运行



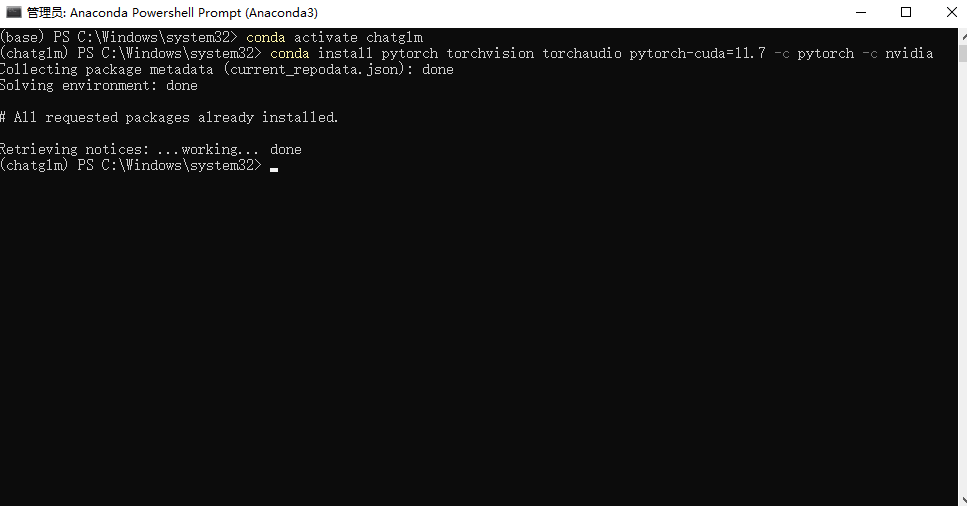
图片包含 形状

描述已自动生成

5、开始执行各种命令安装各种依赖，首先输入conda activate chatglm

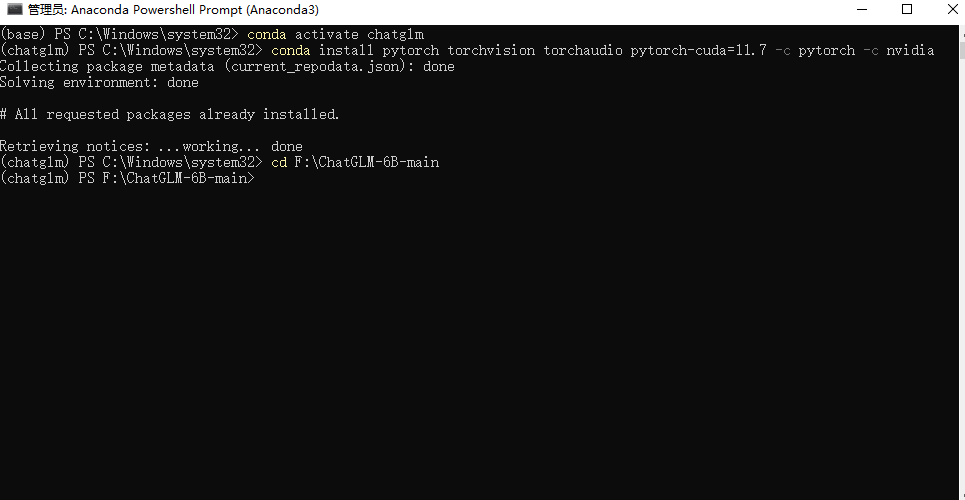


然后安装pytorch，命令为conda install pytorch torchvision torchaudio pytorch-cuda=11.7 -c pytorch -c nvidia



正常应该是下载各种东西，由于我已经安装就不会有下载显示

进入github下载的文件所在的文件夹，我这里的文件位置是F:\ChatGLM-6B-main，所以使用cd F:\ChatGLM-6B-main进入



分别使用命令pip install -r requirements.txt和pip install gradio下载依赖

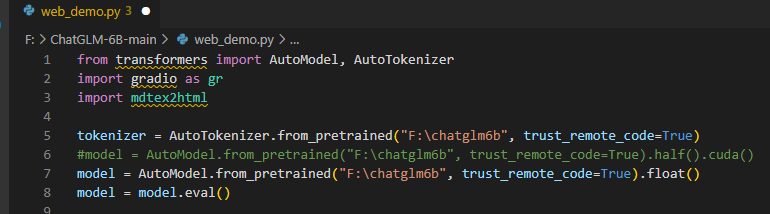


文本

描述已自动生成

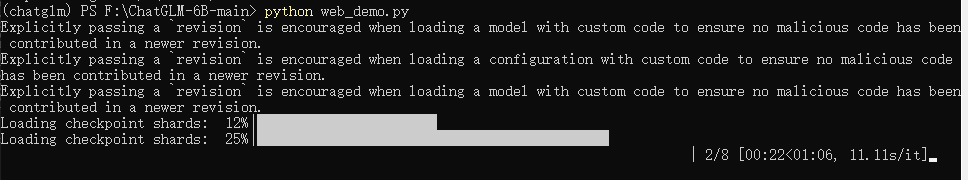
由于我已经下载过了，依旧不会出现下载界面

1. 更改脚本web\_demo.py，在源码文件夹中打开web\_demo.py，更改第5、6行代码



其中的“F:\chatflm6b”为模型所在文件夹的位置，即第二步下载的模型，将后方更改后即为CPU运行

1. 运行。回到命令行窗口，输入python web\_demo.py尝试运行



加载完会打开浏览器

图形用户界面, 应用程序

描述已自动生成

如果没有打开，但出现了如下界面，可手动输入127.0.0.1:7860，打开网址

文本

描述已自动生成

即为部署完成

若是报错，则根据错误提示，使用“pip install 缺失的依赖”命令安装缺失的依赖即可

## GPU版

GPU版与CPU版大致相同，区别在于需要安装CUDA和cudnn，以及代码的更改

1、下载CUDA和cudnn

CUDA：

<https://developer.nvidia.com/cuda-11-6-1-download-archive?target_os=Windows&target_arch=x86_64&target_version=11&target_type=exe_local>

下载后安装，注意不要选择最新版本，建议11.6或者11.7。下载后记下保存地址

Cudnn：

<https://developer.nvidia.com/rdp/cudnn-archive>

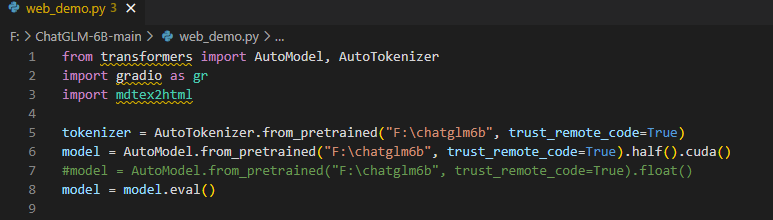
下载后得到三个文件夹

图形用户界面, 应用程序, 表格, Excel

描述已自动生成

将这三个文件夹拖入C:\Program Files\NVIDIA GPU Computing Toolkit\CUDA\v11.6中，这是默认安装位置，根据安装位置不同，路径并不一致

2、在web\_demo.py中将代码更改为如下所示，图一这样会以默认的设置运行，用于16g显存的显卡，图二为4g显存的显卡，若不同需要，将代码中的4改为其它即可，4代表4g，8代表8g



电脑萤幕截图

描述已自动生成

## API调用

与上面方法大体一致，不同的是需要安装额外的依赖，命令为：python api.py

且最后运行是api.py而不是web\_demo.py

图形用户界面, 文本, 应用程序

描述已自动生成

同时更改api.py脚本中的第53和54行，更改方法与上面一致

文本

描述已自动生成

启动后

文本

描述已自动生成

调用方法为

图形用户界面, 文本, 应用程序

描述已自动生成

可使用Postman进行测试

电脑萤幕的截图

描述已自动生成

文本

描述已自动生成