1. clone项目[llm-attacks/llm-attacks: Universal and Transferable Attacks on Aligned Language Models](https://github.com/llm-attacks/llm-attacks/tree/main)

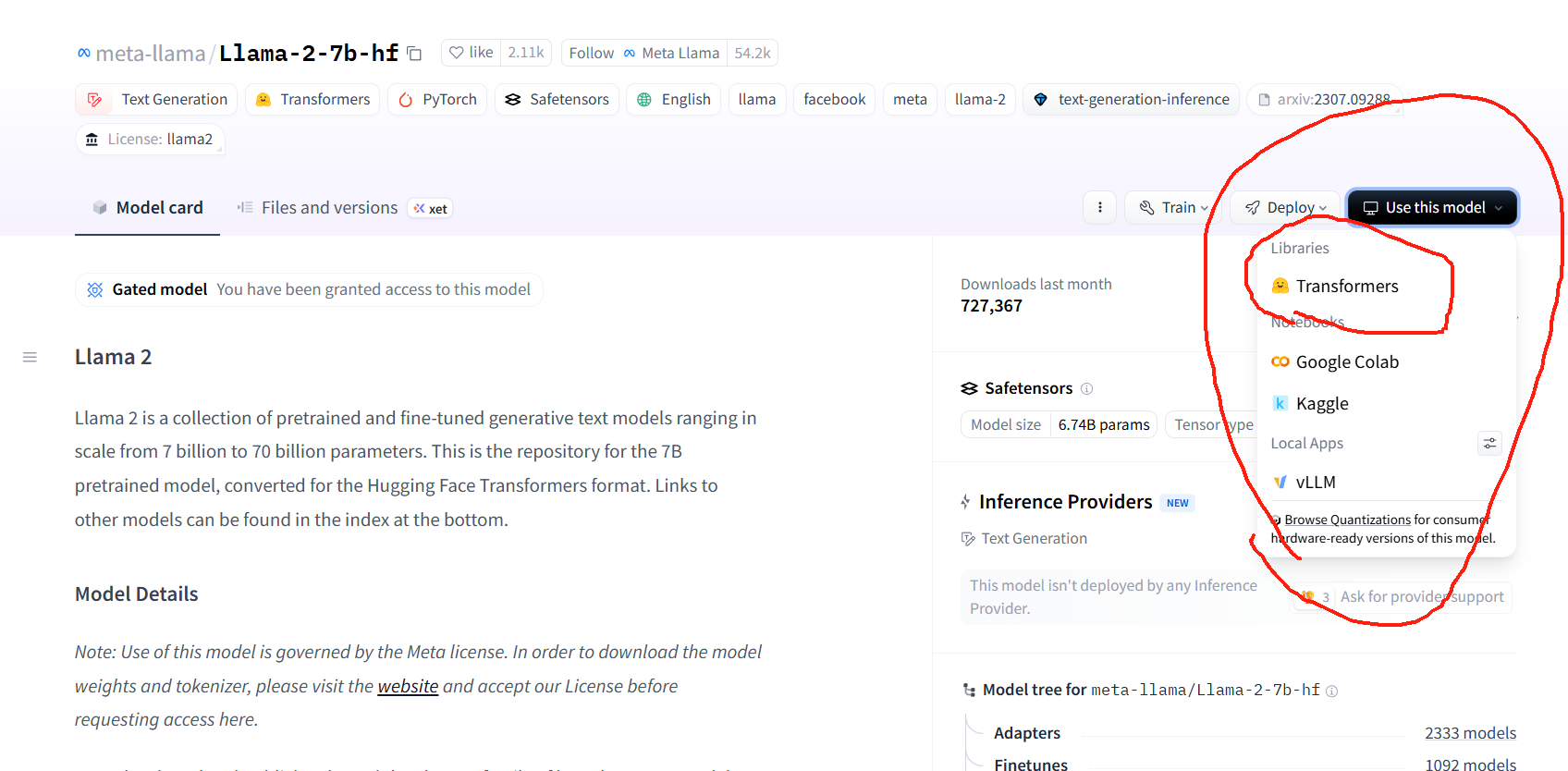
使用命令git clone https://github.com/llm-attacks/llm-attacks.git

1. 创建虚拟环境后，在环境中安装依赖（切换路径到项目主目录）

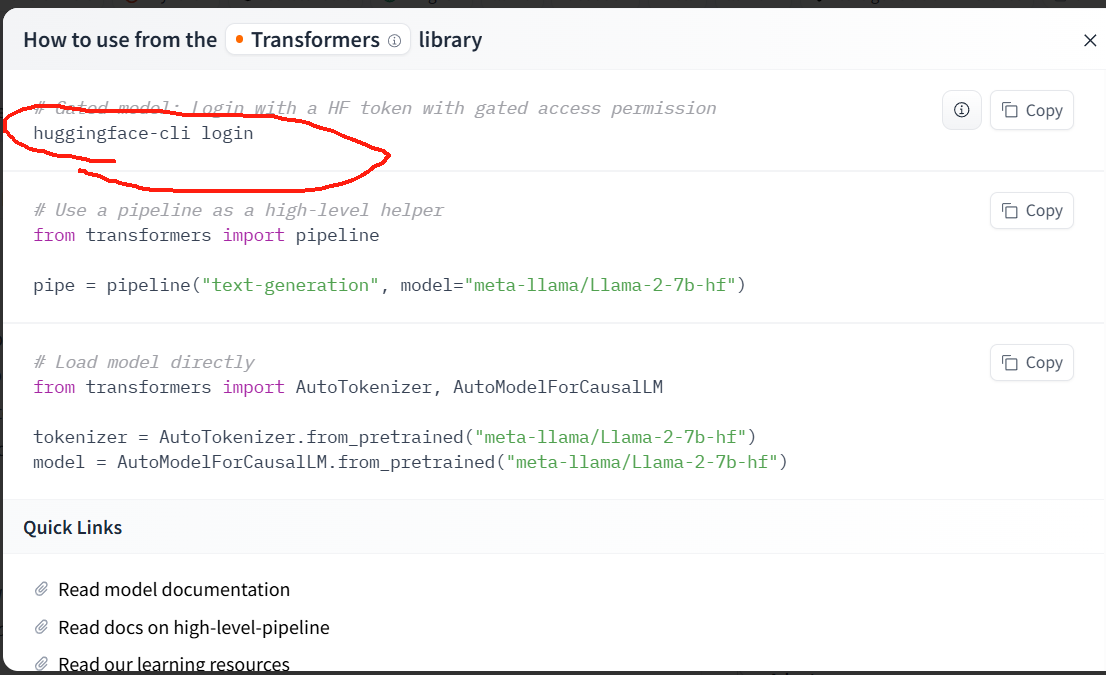
pip install -e .

1. 下载模型

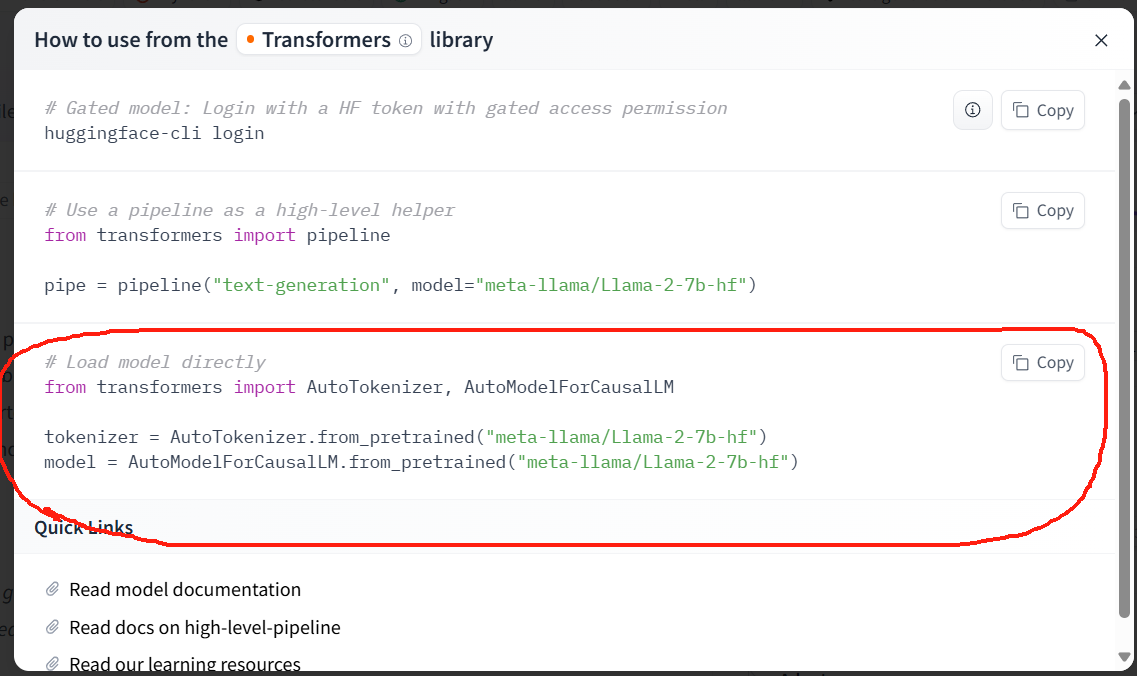
[lmsys/vicuna-7b-v1.5 at main](https://huggingface.co/lmsys/vicuna-7b-v1.5/tree/main)



点击**上图**中两个红圈



**在控制台输入上图中红圈命令并回车，之后输入令牌：hf\_tIaWojCpuUqNTCDsCtqjxntHwxWHmDLfRo**



将**上图**红圈代码作为python文件执行，复制模型下载到本地的地址

更改项目中文件experiments/configs/individual\_llama2.py中的内容：



将上图中两个中括号内的地址改为本地模型的储存文件夹地址

4．执行

**切换到目录experiments/launch\_scripts**

执行前根据情况修改experiments/launch\_scripts/run\_gcg\_individual.sh文件中的bathsize大小，具体位置如下图红圈：



最后执行命令bash run\_gcg\_individual.sh llama2 strings > experiment\_log.txt 2>&1 &