

基础岛任务笔记3-浦语提示词工程实践

- **背景问题：**近期相关研究发现，LLM在对比浮点数字时表现不佳，经验证，internlm2-chat-1.8b (internlm2-chat-7b)也存在这一问题，例如认为 `13.8<13.11` 。
- **任务要求：**利用LangGPT优化提示词，使LLM输出正确结果。**完成一次并提交截图即可**

环境配置

创建名为langgpt的虚拟环境，安装必要库和依赖，安装必要软件到项目目录中

```
1 # 创建虚拟环境
2 conda create -n langgpt python=3.10 -y
3
4 # 安装一些必要的库
5 conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2.1.2 pytorch-
  cuda=12.1 -c pytorch -c nvidia -y
6
7 # 安装其他依赖
8 pip install transformers==4.43.3
9
10 pip install streamlit==1.37.0
11 pip install huggingface_hub==0.24.3
12 pip install openai==1.37.1
13 pip install lmdeploy==0.5.2
14
15 ## 创建路径
16 mkdir langgpt
17 ## 进入项目路径
18 cd langgpt
19
20 #安装必要软件
21 apt-get install tmux
```

```
CPU 10.54% GPU-1: 10% Nvidia A100 0%
内存 0.60 / 24 GB 2.48% 显存 0 / 8182 MiB 0%

5. 将conda环境一键添加到jupyterlab:
lab add {YOUR_CONDA_ENV_NAME}

-----

(base) root@intern-studio-40077780:~# conda create -n langgpt python=3.10 -y
Retrieving notices: ...working... done
Collecting package metadata (current_repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /root/.conda/envs/langgpt

  added / updated specs:
    - python=3.10

The following NEW packages will be INSTALLED:

 _libgcc_mutex            anaconda/pkgs/main/linux-64::_libgcc_mutex-0.1-main
 _openmp_mutex            anaconda/pkgs/main/linux-64::_openmp_mutex-5.1-1_gnu
 bzip2                    anaconda/pkgs/main/linux-64:bzip2-1.0.8-h5eee18b_6
 ca-certificates          anaconda/pkgs/main/linux-64:ca-certificates-2024.7.2-h06a4308_0
 ld_impl_linux-64         anaconda/pkgs/main/linux-64:ld_impl_linux-64-2.38-h1181459_1
 libffi                   anaconda/pkgs/main/linux-64:libffi-3.4.4-h6a678d5_1
 libgcc-ng                anaconda/pkgs/main/linux-64:libgcc-ng-11.2.0-h1234567_1
 libgomp                  anaconda/pkgs/main/linux-64:libgomp-11.2.0-h1234567_1
 libstdcxx-ng             anaconda/pkgs/main/linux-64:libstdcxx-ng-11.2.0-h1234567_1
 libuuid                  anaconda/pkgs/main/linux-64:libuuid-1.41.5-h5eee18b_0
 ncurses                  anaconda/pkgs/main/linux-64:ncurses-6.4-h6a678d5_0
 openssl                  anaconda/pkgs/main/linux-64:openssl-3.0.14-h5eee18b_0
```

```
CPU 10.23% GPU-1: 10% Nvidia A100 0%
内存 0.60 / 24 GB 2.48% 显存 0 / 8182 MiB 0%

# To deactivate an active environment, use
#
# $ conda deactivate

(base) root@intern-studio-40077780:~# conda activate langgpt
(langgpt) root@intern-studio-40077780:~# conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2.1.2 pyto
rch-cuda=12.1 -c pytorch -c
usage: conda install [-h] [--revision REVISION] [-n ENVIRONMENT | -p PATH] [-c CHANNEL] [--use-local]
                    [--override-channels] [--repodata-fn REPODATA_FNS] [--experimental {jlap,lock}]
                    [--no-lock] [--strict-channel-priority] [--no-channel-priority] [--no-deps | --only-deps]
                    [--no-pin] [--copy] [-C] [-k] [--offline] [--json] [-v] [-q] [-d] [-y] [--download-only]
                    [--show-channel-urls] [--file FILE] [--solver {classic,libmamba}] [--force-reinstall]
                    [--freeze-installed] --update-deps | -S | --update-all | --update-specs] [-m]
                    [--clobber] [--dev]
                    [package_spec ...]

conda install: error: argument -c/--channel: expected one argument
(langgpt) root@intern-studio-40077780:~# conda install pytorch==2.1.2 torchvision==0.16.2 torchaudio==2.1.2 pyto
rch-cuda=12.1 -c pytorch -c nvidia -y
Collecting package metadata (current_repodata.json): done
Solving environment: unsuccessful initial attempt using frozen solve. Retrying with flexible solve.
Collecting package metadata (repodata.json): done
Solving environment: done

## Package Plan ##

  environment location: /root/.conda/envs/langgpt

  added / updated specs:
    - pytorch-cuda=12.1
    - pytorch==2.1.2
    - torchaudio==2.1.2
    - torchvision==0.16.2
```

```
CPU 10.67% GPU-1: 10% Nvidia A100 0%
内存 0.60 / 24 GB 2.48% 显存 0 / 8182 MiB 0%

Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
ERROR: Could not find a version that satisfies the requirement transformers==4.43.3 (from versions: none)
ERROR: No matching distribution found for transformers==4.43.3
(langgpt) root@intern-studio-40077780:~# pip install transformers==4.43.3
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting transformers==4.43.3
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/ad/ff/b3e311e58b9c90b149fb957953b228287d7c9fe78df9a3a72e8715c5fc56/transformers-4.43.3-py3-none-any.whl (9.4 MB)
    9.4/9.4 MB 19.3 MB/s eta 0:00:00
Requirement already satisfied: filelock in ./conda/envs/langgpt/lib/python3.10/site-packages (from transformers==4.43.3) (3.13.1)
Collecting huggingface-hub<1.0,>=0.23.2 (from transformers==4.43.3)
  Using cached https://pypi.tuna.tsinghua.edu.cn/packages/b9/8f/d6718641c14d98a5848c6a24d2376028d292074ffade0702940a4b1dde76/huggingface_hub-0.24.6-py3-none-any.whl (417 kB)
Requirement already satisfied: numpy>=1.17 in ./conda/envs/langgpt/lib/python3.10/site-packages (from transformers==4.43.3) (1.26.4)
Collecting packaging>=20.0 (from transformers==4.43.3)
  Using cached https://pypi.tuna.tsinghua.edu.cn/packages/08/aa/cc0199a5f0ad350994d660967a8efb233fe0416e4639146c089643407ce6/packaging-24.1-py3-none-any.whl (53 kB)
Requirement already satisfied: pyyaml>=5.1 in ./conda/envs/langgpt/lib/python3.10/site-packages (from transformers==4.43.3) (6.0.1)
Collecting regex!=2019.12.17 (from transformers==4.43.3)
  Using cached https://pypi.tuna.tsinghua.edu.cn/packages/3e/66/04b63f31580026c8b819aed7f171149177d10cfab27477ea8800a2268d50/regex-2024.7.24-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (776 kB)
Requirement already satisfied: requests in ./conda/envs/langgpt/lib/python3.10/site-packages (from transformers==4.43.3) (2.32.3)
Collecting safetensors>=0.4.1 (from transformers==4.43.3)
  Using cached https://pypi.tuna.tsinghua.edu.cn/packages/18/f3/27bf4d7112b194eea2d8401706953080692d37acelb74b36fcc7234961cd/safetensors-0.4.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (435 kB)
Collecting tokenizers<0.20,>=0.19 (from transformers==4.43.3)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/40/4f/eb78de4af3b17b589f43a369cbf0c3a7173f25c3d2cd93068852c07689aa/tokenizers-0.19.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.6 MB)
    3.6/3.6 MB 3.2 MB/s eta 0:00:00
```

```
CPU 10.67% GPU-1: 10% Nvidia A100 0%
内存 0.60 / 24 GB 2.48% 显存 0 / 8182 MiB 0%

ad: https://pip.pypa.io/warnings/venv. Use the --root-user-action option if you know what you are doing and want to suppress this warning.
(langgpt) root@intern-studio-40077780:~# pip install huggingface_hub==0.24.3
Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting huggingface_hub==0.24.3
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/0f/36/83c0f0c7a5ec75738241c4c0c066097e4f74729716961db6a2905395015c/huggingface_hub-0.24.3-py3-none-any.whl (417 kB)
Requirement already satisfied: filelock in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (3.13.1)
Requirement already satisfied: fsspec>=2023.5.0 in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (2024.6.1)
Requirement already satisfied: packaging>=20.9 in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (24.1)
Requirement already satisfied: pyyaml>=5.1 in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (6.0.1)
Requirement already satisfied: requests in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (2.32.3)
Requirement already satisfied: tqdm>=4.42.1 in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (4.66.5)
Requirement already satisfied: typing-extensions>=3.7.4.3 in ./conda/envs/langgpt/lib/python3.10/site-packages (from huggingface_hub==0.24.3) (4.11.0)
Requirement already satisfied: charset-normalizer<4,>=2 in ./conda/envs/langgpt/lib/python3.10/site-packages (from requests->huggingface_hub==0.24.3) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in ./conda/envs/langgpt/lib/python3.10/site-packages (from requests->huggingface_hub==0.24.3) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in ./conda/envs/langgpt/lib/python3.10/site-packages (from requests->huggingface_hub==0.24.3) (2.2.2)
Requirement already satisfied: certifi>=2017.4.17 in ./conda/envs/langgpt/lib/python3.10/site-packages (from requests->huggingface_hub==0.24.3) (2024.7.4)
Installing collected packages: huggingface_hub
  Attempting uninstall: huggingface_hub
    Found existing installation: huggingface-hub 0.24.6
    Uninstalling huggingface-hub-0.24.6:
```

```
CPU 11.79% GPU-1: 10% Nvidia A100 0%
内存 0.59 / 24 GB 2.47% 显存 0 / 8182 MiB 0%

Successfully uninstalled huggingface-hub-0.24.6
Successfully installed huggingface_hub-0.24.3
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager, possibly rendering your system unusable. It is recommended to use a virtual environment instead: https://pip.pypa.io/warnings/venv. Use the --root-user-action option if you know what you are doing and want to suppress this warning.
(langgpt) root@intern-studio-40077780:~# pip install openai==1.37.1

Looking in indexes: https://pypi.tuna.tsinghua.edu.cn/simple
Collecting openai==1.37.1
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/43/18/8425b246cefad230b2574018bba0c039d70750b0a270eba1c9576a1fcab2/openai-1.37.1-py3-none-any.whl (337 kB)
Collecting anyio<5,>=3.5.0 (from openai==1.37.1)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/7b/a2/10639a79341f6c019dedc95bd48a4928eed9f1d1197f4c04f546fc7ae0ff/anyio-4.4.0-py3-none-any.whl (86 kB)
Collecting distro<2,>=1.7.0 (from openai==1.37.1)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/12/b3/231ffd4ab1fc9d679809f356cebee130ac7daa00d6d6f3206dd4fd137e9e/distro-1.9.0-py3-none-any.whl (20 kB)
Collecting httpx<1,>=0.23.0 (from openai==1.37.1)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/56/95/9377bcb415797e44274b51d46e3249eba641711cf3348050f76ee7b15ffc/httpx-0.27.2-py3-none-any.whl (76 kB)
Collecting pydantic<3,>=1.9.0 (from openai==1.37.1)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/1f/fa/b7f815b8c9ad021c07f88875b601222ef5e70619391ade4a49234d12d278/pydantic-2.8.2-py3-none-any.whl (423 kB)
Collecting sniffio (from openai==1.37.1)
  Downloading https://pypi.tuna.tsinghua.edu.cn/packages/e9/44/75a9c9421471a6c4805dbf2356f7c181a29c1879239abablea2cc8f38b40/sniffio-1.3.1-py3-none-any.whl (10 kB)
Requirement already satisfied: tqdm>4 in ./conda/envs/langgpt/lib/python3.10/site-packages (from openai==1.37.1) (4.66.5)
Requirement already satisfied: typing-extensions<5,>=4.7 in ./conda/envs/langgpt/lib/python3.10/site-packages (from openai==1.37.1) (4.11.0)
Requirement already satisfied: idna>=2.8 in ./conda/envs/langgpt/lib/python3.10/site-packages (from anyio<5,>=3.5.0->openai==1.37.1) (3.7)
```

模型部署

我们使用了开发机，在路径 `/share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1_8b` 下找到书生浦语1.8B大语言模型

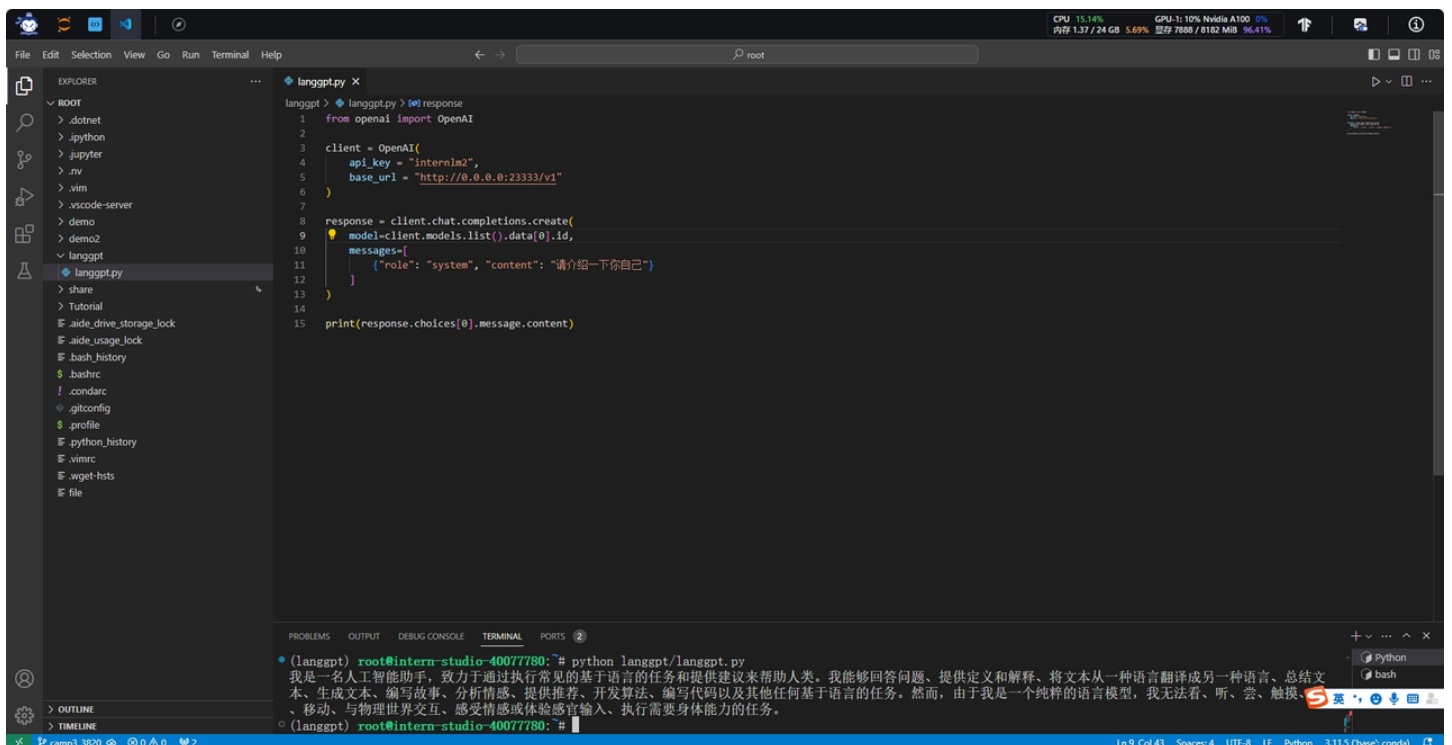
部署模型为OpenAI server

- 1 #由于服务需要持续运行，需要将进程维持在后台，所以这里使用tmux软件创建新的命令窗口。运行如下命令创建窗口：
- 2 `tmux new -t langgpt`
- 3
- 4 #创建完成后，运行下面的命令进入新的命令窗口(首次创建自动进入，之后需要连接)：
- 5 `tmux a -t langgpt`
- 6
- 7 #进入命令窗口后，需要在新窗口中再次激活环境。
- 8 #然后，使用LMDeploy进行部署，参考如下命令：
- 9 `CUDA_VISIBLE_DEVICES=0 lmdeploy serve api_server`
`/share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1_8b --server-port`
`23333 --api-keys internlm2`

```
(base) root@intern-studio-40077780:~/langgpt# conda activate langgpt
(langgpt) root@intern-studio-40077780:~/langgpt# CUDA_VISIBLE_DEVICES=0 lmdeploy serve api_server /share/new_models/Shanghai_AI_Laboratory/internlm2-chat-1.8b --server-port 23333 --api-keys internlm2
[WARNING] gemm_config.in is not found; using default GEMM algo
HINT: Please open http://0.0.0.0:23333 in a browser for detailed api usage!!!
HINT: Please open http://0.0.0.0:23333 in a browser for detailed api usage!!!
HINT: Please open http://0.0.0.0:23333 in a browser for detailed api usage!!!
INFO: Started server process [350102]
INFO: Waiting for application startup.
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:23333 (Press CTRL+C to quit)
```

部署成功后，可以利用如下脚本调用部署的InternLM2-chat-1.8b模型并测试是否部署成功。

```
1 from openai import OpenAIClient = OpenAI(
2     api_key = "internlm2",
3     base_url = "http://0.0.0.0:23333/v1"
4 )
5
6 response = client.chat.completions.create(
7     model=client.models.list().data[0].id,
8     messages=[
9         {"role": "system", "content": "请介绍一下你自己"}
10    ]
11 )
12 print(response.choices[0].message.content)
```



图形化界面调用

InternLM部署完成后，利用提供的 `chat_ui.py` 创建图形化界面。

从Github获取项目，运行如下命令：

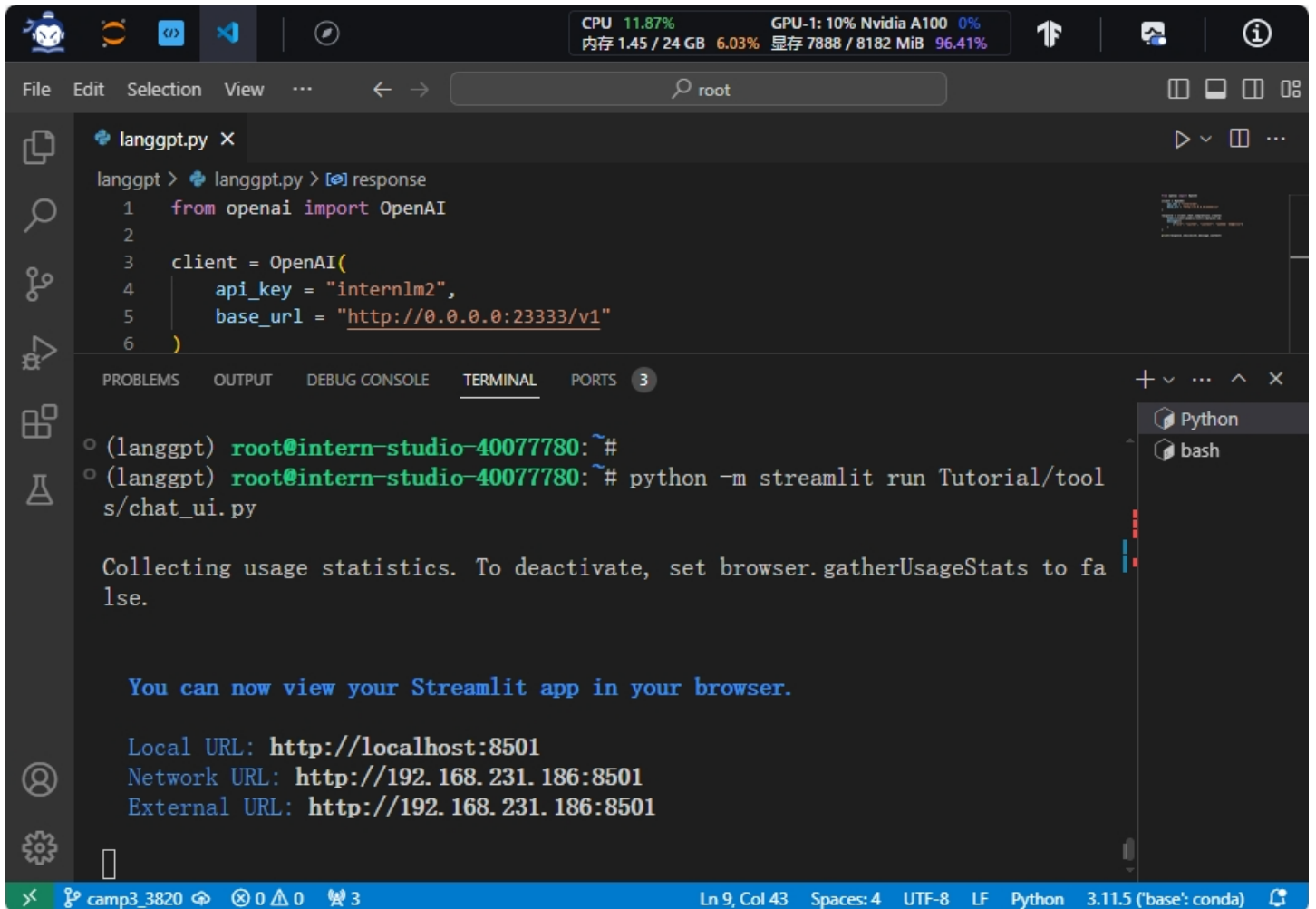
```
git clone https://github.com/InternLM/Tutorial.git
```

下载完成后，运行如下命令进入项目所在的路径：

```
cd Tutorial/tools
```

进入正确路径后，运行如下脚本运行项目：

```
python -m streamlit run chat_ui.py
```



The screenshot shows a VS Code editor window with a Python script named `langgpt.py` open. The script imports `OpenAI` from the `openai` module and initializes a client with an API key and a base URL. Below the editor, the `TERMINAL` tab is active, showing the execution of the script. The terminal output indicates that the Streamlit app is running successfully and provides the local, network, and external URLs for viewing the app in a browser.

```
langgpt.py x
langgpt > langgpt.py > [E] response
1 from openai import OpenAI
2
3 client = OpenAI(
4     api_key = "internlm2",
5     base_url = "http://0.0.0.0:23333/v1"
6 )
```

PROBLEMS OUTPUT DEBUG CONSOLE **TERMINAL** PORTS 3

```
(langgpt) root@intern-studio-40077780: ~#
(langgpt) root@intern-studio-40077780: ~# python -m streamlit run Tutorial/tools/chat_ui.py

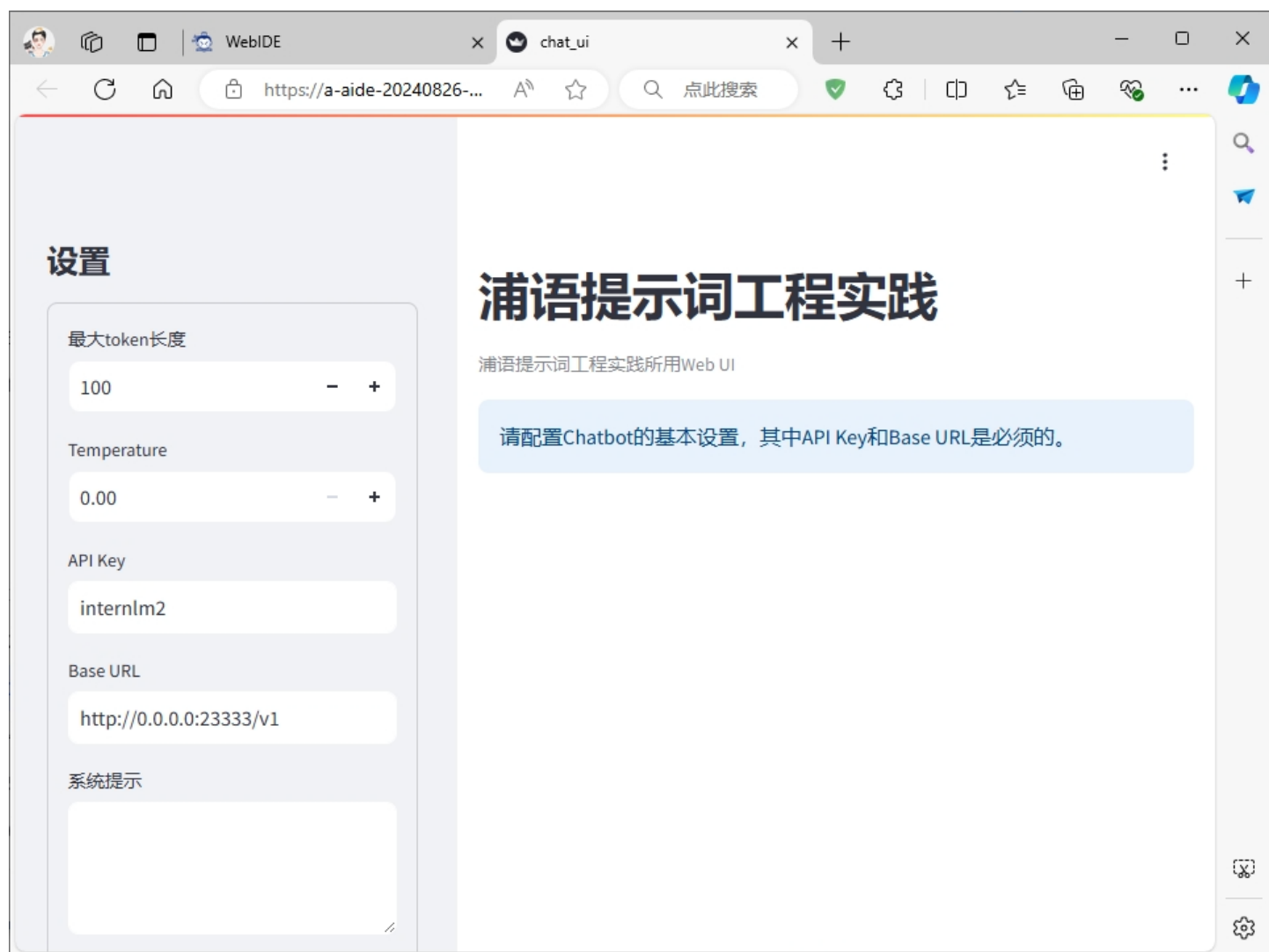
Collecting usage statistics. To deactivate, set browser.gatherUsageStats to false.

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.231.186:8501
External URL: http://192.168.231.186:8501
```

Python
bash

camp3_3820 0 0 3 Ln 9, Col 43 Spaces: 4 UTF-8 LF Python 3.11.5 ('base': conda)



利用Prompt解决“LLM在对比浮点数字时表现不佳”问题



利用思维链写了提示词如下：

- 1 a.b和c.d谁大?
- 2 `if a>c,则a.b>c.d`
- 3 `if a=c,b>d,则a.b>c.d`
- 4 `if a<c,则a.b<c.d`

解决了问题



~~提示工程~~

~~什么是prompt~~

~~什么是提示工程~~

~~提示设计框架~~

~~LangGPT结构化提示词~~