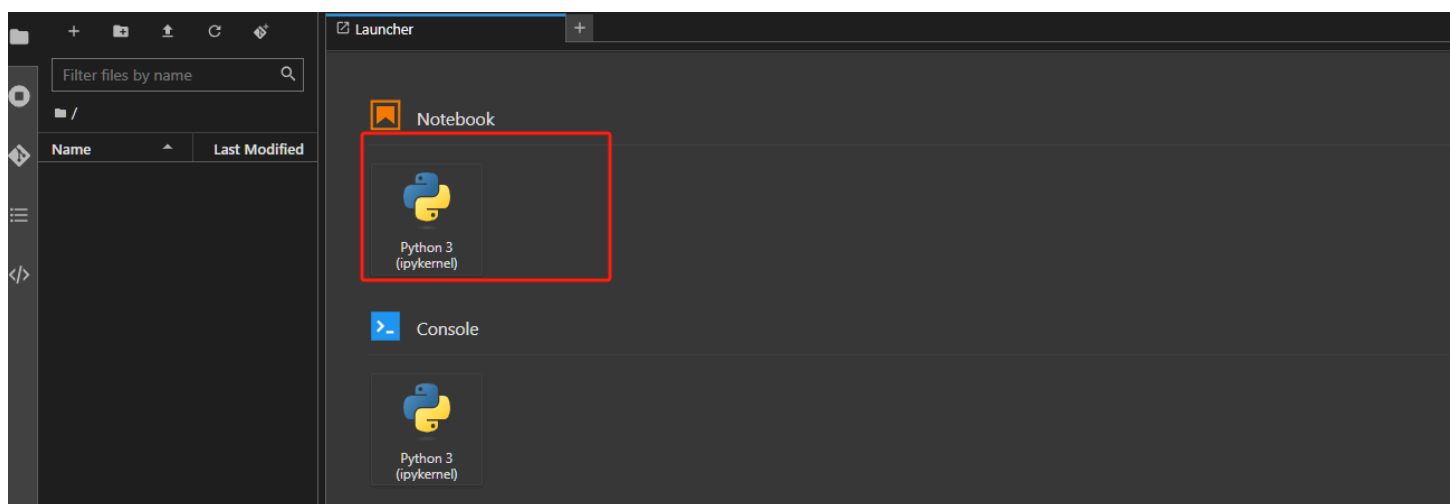
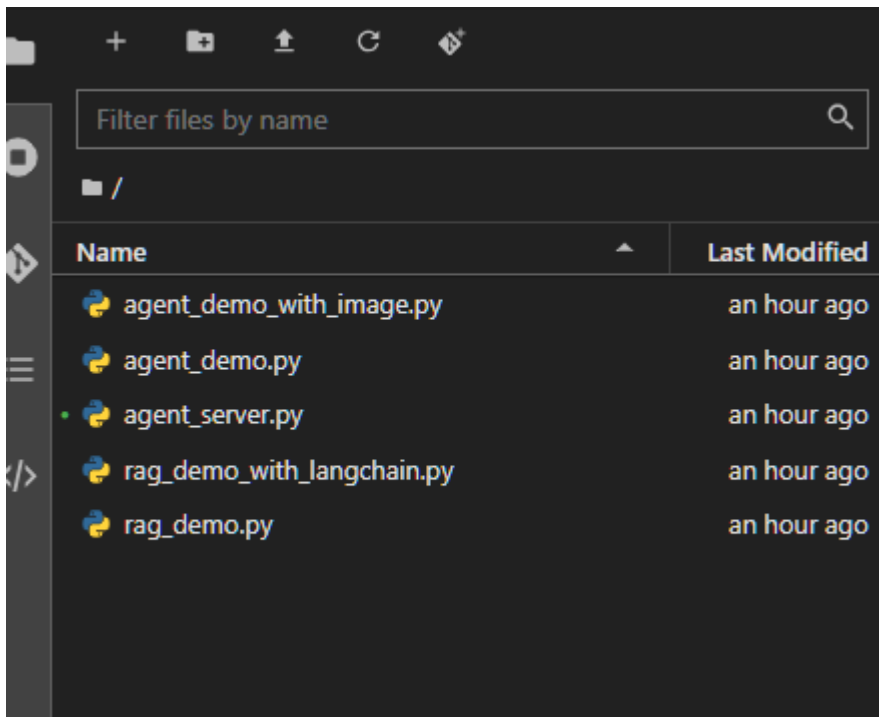


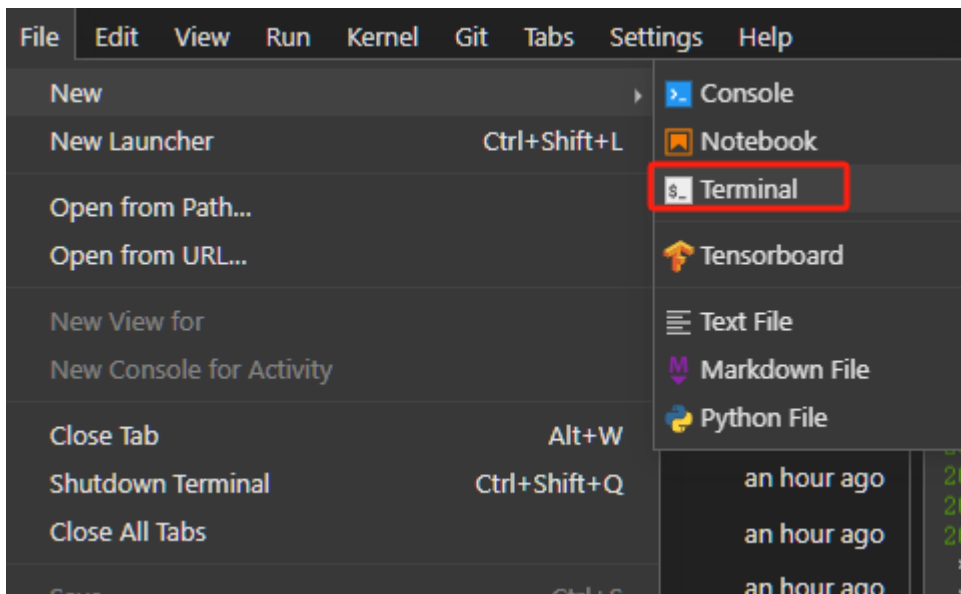
1. 启动我的Notebook



2. 把文件一个一个复制到里面。类似下面这样：



3. 打开一个终端：



3.1 修改环境变量：

执行命令打开环境变量文件：

```
vim ~/.bashrc
```

在文件的最后，加上：

```
export OPENAI_API_KEY="sk-xxxx"
```

```
export OPENAI_BASE_URL="https://xxxx"
```

保存文件之后，执行命令：

```
source ~/.bashrc
```

4. 下载ngrok工具：

首先得注册账号，登录 <https://dashboard.ngrok.com/get-started/setup/linux>

依次在终端执行以下命令：

Installation

Apt Download Snap

Install ngrok via Apt with the following command:

```
curl -s https://ngrok-agent.s3.amazonaws.com/ngrok.asc \
| sudo tee /etc/apt/trusted.gpg.d/ngrok.asc >/dev/null \
&& echo "deb https://ngrok-agent.s3.amazonaws.com buster main" \
| sudo tee /etc/apt/sources.list.d/ngrok.list \
&& sudo apt update \
&& sudo apt install ngrok
```

Run the following command to add your authtoken to the default ngrok.yml configuration file.

```
ngrok config add-authtoken 2gFT0ke5XhFDgIPn...ohVpCn
```

Deploy your app online

Ephemeral Domain Static Domain

Put your app online at [ephemeral domain](#) Forwarding to your upstream service. For example, if it is listening on port `http://localhost:8080`, run:

```
ngrok http http://localhost:8080
```

Once running, your endpoints will be listed on the [endpoints page](#).

执行完之后你会看到notebook中下面的界面：

ngrok

Try our new Traffic Inspector Dev Preview: <https://ngrok.com/r/ti>

```
Session Status      online
Account             XINJ_ZH (Plan: Free)
Version             3.9.0
Region              Japan (jp)
Latency              50ms
Web Interface        http://127.0.0.1:4040
Forwarding            https://6d34-120-27-137-238.ngrok-free.app -> http://localhost:8080
```

Connections	ttl	opn	rt1	rt5	p50	p90
	0	0	0.00	0.00	0.00	0.00

5. 打开一个新的终端，运行程序：

5.1 安装所需的依赖库

```
pip install pypdf pdfminer.six openai langchain langchain_openai agentscope Chroma chromadb Spee
```

5.2 启动Flask程序

```
python3 agent_server.py
```

运行成功后看到如下界面：

```
root@dsw-465976-7b5cb688fb-hzfaj:/mnt/workspace# python3 agent_server.py
2024-05-19 11:01:20.991 INFO agentscope.models.read_model_configs:170 - Load configs for model wrapper: openai_cfg
2024-05-19 11:01:20.995 INFO agentscope.utils.monitor._create_monitor_table:390 - Init [monitor_metrics] as the monitor table
2024-05-19 11:01:20.995 INFO agentscope.utils.monitor._create_monitor_table:397 - Init [monitor_metrics_quota_exceeded] as the monitor trigger
2024-05-19 11:01:20.995 INFO agentscope.utils.monitor.__init__:366 - SqliteMonitor initialization completed at [./runs/run_20240519-110119_cellvg/agentscope.db]
2024-05-19 11:01:20.995 WARNING agentscope.models.model.get_wrapper:212 - Model type [openai] will be deprecated in future releases, please use [openai_chat] instead.
2024-05-19 11:01:20.997 INFO agentscope.models.model.__init__:201 - Initialize model by configuration [openai_cfg]
2024-05-19 11:01:20.981 INFO agentscope.utils.monitor.register_budget:609 - set budget None to gpt-3.5-turbo
2024-05-19 11:01:20.982 INFO agentscope.utils.monitor.register:417 - Register metric [gpt-3.5-turbo.cost] to SqliteMonitor with unit [dollar] and quota [None]
2024-05-19 11:01:20.984 INFO agentscope.utils.monitor.register:417 - Register metric [gpt-3.5-turbo.prompt_tokens] to SqliteMonitor with unit [token] and quota [None]
2024-05-19 11:01:20.986 INFO agentscope.utils.monitor.register:417 - Register metric [gpt-3.5-turbo.completion_tokens] to SqliteMonitor with unit [token] and quota [None]
2024-05-19 11:01:20.989 INFO agentscope.utils.monitor.register:417 - Register metric [gpt-3.5-turbo.call_counter] to SqliteMonitor with unit [times] and quota [None]
2024-05-19 11:01:20.991 INFO agentscope.utils.monitor.register:417 - Register metric [gpt-3.5-turbo.total_tokens] to SqliteMonitor with unit [token] and quota [None]
* Serving Flask app 'agent_server'
* Debug mode: off
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on all addresses (0.0.0.0)
* Running on http://127.0.0.1:8080
* Running on http://10.241.187.76:8080
Press CTRL+C to quit
```

6. 在电脑浏览器中访问：

```
ngrok

Try our new Traffic Inspector Dev Preview: https://ngrok.com/r/ti

Session Status      online
Account             XINJ_ZH (Plan: Free)
Version             3.9.0
Region              Japan (jp)
Latency             50ms
Web Interface       http://127.0.0.1:4040
Forwarding           https://6d34-120-27-137-238.ngrok-free.app -> http://localhost:8080

Connections
  ttl    opn    rt1    rt5    p50    p90
    0      0    0.00  0.00  0.00  0.00
```

点这里的链接

后面加接口就可以调用了，类似下面这样：

https://dsw-gateway-cn-hangzhou.data.aliyun.com/dsw-465976/proxy/8080/user_query?query=hello

← → ↻ dsw-gateway-cn-hangzhou.data.aliyun.com/dsw-465976/proxy/8080/user_query?query=hello

Google



必应



百度一下, 你就知道



美观输出 ☒

```
{
  "content": "Hello! How can I help you today?",
  "id": "6c691a03b8ec476686bfe1e86456f092",
  "name": "myAgent",
  "role": "assistant",
  "timestamp": "2024-05-19 11:03:48",
  "url": null
}
```