# 讲义地址

https://docs.qq.com/doc/p/0555f97a90b6f2f9f039bd0c9251f5d7d1e7a506 密码:hsw123456

# 阿里云租赁GPU服务器

阿里云租赁GPU服务器: https://ecs-buy.aliyun.com/ecs? spm=5176.8300896.J\_5253785160.3.251e3680LsIXpx#/custom/postpay/cn-hangzhou? toNewVersion=true&instanceType=ecs.gn7i-c16g1.4xlarge

之所以不用autoDL是因为它的每一个服务器实例其实都是一个docker容器,而docker实例无法 套娃,做不了本次实验。

租赁阿里云服务器的时候,一定要注意看配置,必须是带gpu显卡的,并且系统盘内存最好是大一些,我选的是一个T,因为后续我们下载大模型文件,可能会耗费大量内存。

我是用xterminer来利用ssh连的服务器,必须要在监控模块看到gpu才算对。如下图:





## 安装docker和docker-compose

安装过程就省略了,最后一定要验证 docker --version 和 docker-compose --version,能 看到版本号才能说明装完了。

# 准备compose配置文件

mkdir fastgpt
cd fastgpt

内容为:

- # 数据库的默认账号和密码仅首次运行时设置有效
- # 如果修改了账号密码,记得改数据库和项目连接参数,别只改一处~
- # 该配置文件只是给快速启动,测试使用。正式使用,记得务必修改账号密码,以及调整合适的知识库参

```
数,共享内存等。
# 如何无法访问 dockerhub 和 git,可以用阿里云(阿里云没有arm包)
version: '3.3'
services:
  # db
  pg:
    #image: pgvector/pgvector:0.7.0-pg15 # docker hub
    image: registry.cn-hangzhou.aliyuncs.com/fastgpt/pgvector:v0.7.0 # 阿里云
    container_name: pg
    restart: always
    ports: # 生产环境建议不要暴露
      - 5432:5432
    networks:
      - fastgpt
    environment:
      # 这里的配置只有首次运行生效。修改后,重启镜像是不会生效的。需要把持久化数据删除再重启,
才有效果
      - POSTGRES_USER=username

    POSTGRES_PASSWORD=password

    POSTGRES_DB=postgres

    volumes:

    ./pg/data:/var/lib/postgresgl/data

  mongo:
    #image: mongo:5.0.18 # dockerhub
    image: registry.cn-hangzhou.aliyuncs.com/fastgpt/mongo:5.0.18 # 阿里云
    # image: mongo:4.4.29 # cpu不支持AVX时候使用
    container_name: mongo
    restart: always
    ports:
      - 27017:27017
    networks:
      - fastapt
    command: mongod --keyFile /data/mongodb.key --replSet rs0
    environment:

    MONGO_INITDB_ROOT_USERNAME=myusername

    MONGO_INITDB_ROOT_PASSWORD=mypassword

    volumes:
      - ./mongo/data:/data/db
    entrypoint:
      - bash
      - -C
      - |
        openssl rand -base64 128 > /data/mongodb.key
        chmod 400 /data/mongodb.key
        chown 999:999 /data/mongodb.key
        echo 'const isInited = rs.status().ok === 1
        if(!isInited){
          rs.initiate({
              _id: "rs0".
             members: [
                  { _id: 0, host: "mongo:27017" }
```

```
]
         })
       }' > /data/initReplicaSet.js
       # 启动MongoDB服务
       exec docker-entrypoint.sh "$$@" &
       # 等待MongoDB服务启动
       until mongo -u myusername -p mypassword --authenticationDatabase admin -
-eval "print('waited for connection')" > /dev/null 2>&1; do
         echo "Waiting for MongoDB to start..."
         sleep 2
       done
       # 执行初始化副本集的脚本
       mongo -u myusername -p mypassword --authenticationDatabase admin
/data/initReplicaSet.js
       # 等待docker-entrypoint.sh脚本执行的MongoDB服务进程
       wait $$!
 # fastqpt
 sandbox:
   container_name: sandbox
   # image: ghcr.io/labring/fastgpt-sandbox:latest # git
   image: registry.cn-hangzhou.aliyuncs.com/fastgpt/fastgpt-sandbox:latest # 阿
里云
   networks:
     - fastqpt
   restart: always
 fastqpt:
   container_name: fastgpt
   # image: ghcr.io/labring/fastgpt:v4.8.8-fix2 # git
   image: registry.cn-hangzhou.aliyuncs.com/fastgpt/fastgpt:v4.8.8-fix2 # 阿里云
   ports:
     - 3000:3000
   networks:
     - fastgpt
   depends_on:
     - mongo
     - pa
     - sandbox
   restart: always
   environment:
     # root 密码, 用户名为: root。如果需要修改 root 密码, 直接修改这个环境变量, 并重启即
可。
     - DEFAULT_ROOT_PSW=1234
     # AI模型的API地址哦。务必加 /v1。这里默认填写了OneApi的访问地址。
     - OPENAI_BASE_URL=http://oneapi:3000/v1
     # AI模型的API Key。(这里默认填写了OneAPI的快速默认key,测试通后,务必及时修改)
     - CHAT_API_KEY=sk-fastapt
```

```
# 数据库最大连接数
```

```
- DB_MAX_LINK=30
```

- # 登录凭证密钥
- TOKEN\_KEY=any
- # root的密钥,常用于升级时候的初始化请求
- ROOT\_KEY=root\_key
- # 文件阅读加密
- FILE\_TOKEN\_KEY=filetoken
- # MongoDB 连接参数. 用户名myusername,密码mypassword。

- MONGODB\_URI=mongodb://myusername:mypassword@mongo:27017/fastgpt?

```
authSource=admin
```

- # pg 连接参数
- PG\_URL=postgresql://username:password@pg:5432/postgres
- # sandbox 地址
- SANDBOX\_URL=http://sandbox:3000
- # 日志等级: debug, info, warn, error
- LOG\_LEVEL=info
- STORE\_LOG\_LEVEL=warn

#### volumes:

- ./config.json:/app/data/config.json

```
# oneapi
```

#### mysql:

```
image: registry.cn-hangzhou.aliyuncs.com/fastgpt/mysql:8.0.36 # 阿里云
# image: mysql:8.0.36
container_name: mysql
restart: always
ports:
    - 3306:3306
networks:
    fastant
```

```
- fastgpt
```

```
command: --default-authentication-plugin=mysql_native_password
environment:
```

```
# 默认root密码,仅首次运行有效
```

```
MYSQL_ROOT_PASSWORD: oneapimmysql
```

```
MYSQL_DATABASE: oneapi
```

#### volumes:

```
- ./mysql:/var/lib/mysql
```

#### oneapi:

#### container\_name: oneapi

```
# image: ghcr.io/songquanpeng/one-api:v0.6.7
```

```
image: registry.cn-hangzhou.aliyuncs.com/fastgpt/one-api:v0.6.6 # 阿里云
ports:
```

```
- 3001:3000
```

#### depends\_on:

- mysql

```
networks:
```

- fastgpt

```
restart: always
```

### environment:

- # mysql 连接参数
- SQL\_DSN=root:oneapimmysql@tcp(mysql:3306)/oneapi
- # 登录凭证加密密钥
- SESSION\_SECRET=oneapikey

```
# 内存缓存
```

- MEMORY\_CACHE\_ENABLED=true
- # 启动聚合更新,减少数据交互频率
- BATCH\_UPDATE\_ENABLED=true
- # 聚合更新时长
- BATCH\_UPDATE\_INTERVAL=10
- # 初始化的 root 密钥 (建议部署完后更改, 否则容易泄露)
- INITIAL\_ROOT\_TOKEN=fastgpt

```
volumes:
```

- ./oneapi:/data

networks:

```
fastgpt:
```

# 准备fastgpt的配置文件

同样在fastgpt目录下, 创建文件 config.json

内容如下:

```
{
  "feConfigs": {
    "lafEnv": "https://laf.dev"
 },
  "systemEnv": {
    "vectorMaxProcess": 15,
    "qaMaxProcess": 15,
    "pgHNSWEfSearch": 100
  },
  "llmModels": [
    {
        "model": "llama-3.1-instruct",
        "name": "llama-3.1-instruct",
        "maxContext": 128000,
        "maxResponse": 128000,
        "quoteMaxToken": 32000,
        "maxTemperature": 1.2,
      "charsPointsPrice": 0,
      "censor": false,
      "vision": false,
      "datasetProcess": false,
      "usedInClassify": true,
      "usedInExtractFields": true,
      "usedInToolCall": true,
      "usedInQueryExtension": true,
      "toolChoice": false,
      "functionCall": false,
      "customCQPrompt": "",
      "customExtractPrompt": "",
      "defaultSystemChatPrompt": ""
```

```
"defaultConfig": {}
```

```
},
{
```

}, {

```
"model": "gwen2-instruct",
"name": "qwen2-instruct",
"avatar": "/imgs/model/gwen.svg",
"maxContext": 128000,
"maxResponse": 128000,
"quoteMaxToken": 32000,
"maxTemperature": 1.2,
"charsPointsPrice": 0,
"censor": false,
"vision": false,
"datasetProcess": true,
"usedInClassify": true,
"usedInExtractFields": true,
"usedInToolCall": true,
"usedInQueryExtension": true,
"toolChoice": false,
"functionCall": false,
"customCQPrompt": "",
"customExtractPrompt": "",
"defaultSystemChatPrompt": "",
"defaultConfig": {}
"model": "glm-4\vee",
"name": "glm-4v",
"avatar": "/imgs/model/chatglm.svg",
"maxContext": 128000,
"maxResponse": 128000,
"quoteMaxToken": 32000,
"maxTemperature": 1.2,
"charsPointsPrice": 0,
"censor": false,
"vision": true,
"datasetProcess": true,
"usedInClassify": true,
"usedInExtractFields": true,
"usedInToolCall": true,
"usedInQueryExtension": true,
```

"toolChoice": true,

"functionCall": false, "customCQPrompt": "",

```
"customExtractPrompt": "",
"defaultSystemChatPrompt": "",
```

```
"defaultConfig": {}
```

```
},
{
```

```
"model": "ERNIE-Speed-128K",
    "name": "ERNIE-Speed-128K",
"avatar": "/imgs/model/ernie.svg",
```

```
"maxContext": 128000,
    "maxResponse": 128000,
    "quoteMaxToken": 32000,
    "maxTemperature": 1.2,
    "charsPointsPrice": 0,
    "censor": false,
    "vision": false.
    "datasetProcess": true,
    "usedInClassify": true,
    "usedInExtractFields": true,
    "usedInToolCall": true,
    "usedInQueryExtension": true,
    "toolChoice": true,
    "functionCall": false,
    "customCQPrompt": "",
    "customExtractPrompt": "",
    "defaultSystemChatPrompt": "",
    "defaultConfig": {}
},
{
  "model": "gpt-4o-mini",
  "name": "gpt-4o-mini",
  "avatar": "/imgs/model/openai.svg",
  "maxContext": 125000,
  "maxResponse": 4000,
  "auoteMaxToken": 120000,
  "maxTemperature": 1.2,
  "charsPointsPrice": 0,
  "censor": false,
  "vision": true.
  "datasetProcess": true,
  "usedInClassify": true,
  "usedInExtractFields": true,
  "usedInToolCall": true,
  "usedInQueryExtension": true,
  "toolChoice": true,
  "functionCall": false,
  "customCQPrompt": "",
  "customExtractPrompt": "",
  "defaultSystemChatPrompt": "",
  "defaultConfig": {}
},
{
  "model": "gpt-40",
  "name": "gpt-40",
  "avatar": "/imgs/model/openai.svq",
  "maxContext": 125000,
  "maxResponse": 4000,
  "quoteMaxToken": 120000,
  "maxTemperature": 1.2,
  "charsPointsPrice": 0,
  "censor": false,
```

```
"vision": true,
    "datasetProcess": false,
    "usedInClassify": true,
    "usedInExtractFields": true,
    "usedInToolCall": true,
    "usedInQueryExtension": true,
    "toolChoice": true,
    "functionCall": false,
    "customCQPrompt": "",
    "customExtractPrompt": "",
    "defaultSystemChatPrompt": "",
    "defaultConfig": {}
  }
],
"vectorModels": [
  {
    "model": "m3e-base",
    "name": "m3e-base",
    "charsPointsPrice": 0,
    "defaultToken": 256,
    "maxToken": 512,
    "weight": 100,
    "defaultConfig": {},
    "dbConfig": {},
    "queryConfig": {}
  },
  {
    "model": "text-embedding-ada-002",
    "name": "Embedding-2",
    "avatar": "/imgs/model/openai.svg",
    "charsPointsPrice": 0,
    "defaultToken": 700,
    "maxToken": 3000,
    "weight": 100,
    "defaultConfig": {},
    "dbConfig": {},
    "queryConfig": {}
  },
  {
    "model": "text-embedding-3-large",
    "name": "text-embedding-3-large",
    "avatar": "/imgs/model/openai.svg",
    "charsPointsPrice": 0,
    "defaultToken": 512,
    "maxToken": 3000,
    "weight": 100,
    "defaultConfig": {
      "dimensions": 1024
    }
  },
  {
    "model": "embeding3",
```

```
"name": "embeding3",
      "avatar": "/imgs/model/chatglm.svg",
      "charsPointsPrice": 0,
      "defaultToken": 512,
      "maxToken": 3000,
      "weight": 100
    }
  ],
  "reRankModels": [
    {
      "model": "bge-reranker-v2-m3",
      "name": "bge-reranker-v2-m3"
    }
  ],
  "audioSpeechModels": [
    {
      "model": "tts-1",
      "name": "OpenAI TTS1",
      "charsPointsPrice": 0,
      "voices": [
        { "label": "Alloy", "value": "alloy", "bufferId": "openai-Alloy" },
{ "label": "Echo", "value": "echo", "bufferId": "openai-Echo" },
        { "label": "Fable", "value": "fable", "bufferId": "openai-Fable" },
        { "label": "Onyx", "value": "onyx", "bufferId": "openai-Onyx" },
        { "label": "Nova", "value": "nova", "bufferId": "openai-Nova" },
        { "label": "Shimmer", "value": "shimmer", "bufferId": "openai-Shimmer" }
      ٦
    }
  ],
  "whisperModel": {
    "model": "whisper-1",
    "name": "Whisper1",
    "charsPointsPrice": 0
  }
}
```

如果改变了这个文件的内容,想要让他立即生效的话,只需要 docker restart fastgpt

## 启动系统

进入fastgpt目录,并且执行 docker-compose up -d

出现下面的结果说明在正常安装。这里会下载很多组件,会花一些时间。

Last Login: Thu Oct 17 20:01:02 2024 from 120.228.142.167 root@1Zbp13vmycy22/29guorq12:-/fastgpt# docker-compose up -d Creating network "fastgpt_fastgpt" with the default driver Pulling pg (registry.ch-hang2bou.aliyuncs.com/fastgpt/pgvector:v0.7.0) v0.7.0: Pulling from fastgpt/pgvector bd0adcf830b12: Pulling fs layer 511dac095c3: Downloading [======>] 1.165k8/1.1 511dac095c3: Downloading [> 296.9k8/29. 15M83c6f657: Pulling fs layer 15M84c6303b12: Downloading [> 881.4k8/29. 15M84c980c3: Downloading [> 9 16401445083b: Downloading [====>] 1.165k8/1.4 562.8k8/1.4 5	终端 3 × +	×
v0.7.0: Pulling from fastgpt/pgvector b0a0cf830b12: Pulling fs layer b11dac095:3: Downloading [====================================	Last login: Thu Oct 17 20:01:02 2024 from 120.228.142.167 root@iZbp13vmycyz2y29guorq1Z:~# cd fastgpt/ root@iZbp13vmycyz2y29guorq1Z:~/fastgpt# docker-compose up -d Creating network "fastgpt_fastgpt" with the default driver Pulling pg (registry.cn-hangzhou.aliyuncs.com/fastgpt/pgvector:v0.7.0)	
b311dac09523: Downloading [       296.9kB/29.         b304cf830b12: Downloading [>       296.9kB/29.         f6401a45083b: Downloading [>       19.8kB/1.4         b030cf830b12: Downloading [>       19.8kB/1.4         b030cf830b12: Downloading [>       19.8kB/1.4         b030cf830b12: Downloading [=>       19.8kB/1.4         b030cf830b12: Downloading [=>       115.1kB/4.5         b030cf830b12: Downloading [=>       1.151kB/4.5         b030cf830b12: Downloading [=>       1.151kB/4.5         b030cf830b12: Downloading [=>       1.151kB/4.5         b141c60826: Downloading [=>       2.059kB/29.         15MB1ace398C3: Downloading [==>       1.759kB/4.5         b140cf830b12: Downloading [==>       2.059kB/29.         15MB3cea98C3: Downloading [==>>       2.649kB/29.         15MB3cea98C3: Downloading [==>>       2.649kB/29.         15MB3cea98C3: Downloading [===>>       3.234kB/29.         15MB       33MB       3.234kB/29.         b0a0cf830b12: Downloading [===>>       3.824kB/29.         15MB       3.825kB/29.       3.825kB/29.         15MB       5034cf830b12: Downloading [====>>       3.644kB/4.5         b0a0cf830b12: Downloading [====>>       3.825kB/29.         15MB       5034kB/29.       3.825kB/29.      <	v0.7.0: Pulling from fastgpt/pgvector b0a0cf830b12: Pulling fs layer	
Difutation of the second secon	b311dac095c3: Downloading [====================================	] 1.165kB/1.1
Implementation of the second system111 <td>balact830b12: Downloading [&gt;</td> <td>296.9kB/29</td>	balact830b12: Downloading [>	296.9kB/29
Te401a45083b: Downloading [>       19.8kB/1.4         b0a0cf830b12: Downloading [=>       881.4kB/29.         SMB049a9521: Waiting       562.8kB/1.4         fe401a45083b: Downloading [=>>       562.8kB/1.4         S150826823: Downloading [=>>       15.62.8kB/1.4         S163cea9863: Downloading [=>>       1.151MB/4.5         JSWB1a45083b: Downloading [=>>       1.131MB/1.4         JSWB1a45083b: Downloading [=>>       1.239MB/1.4         JSWB1a45083b: Downloading [=>>       1.339MB/1.4         JSWB1a2ca98C3: Downloading [==>>       1.239MB/2.9         JSWB3cea98C3: Downloading [==>>       1.232MB/4.5         JSWB2ca98C3: Downloading [==>>       1.339MB/1.4         JSWB2ca98C3: Downloading [===>>       1.324MB/29.         JSWB3cea98C3: Downloading [===>>       1.325MB/29.         JSWB3cea98C3: Downloading [===>>       1.325MB/29.         JSWB       3.504MB/4.5         b0a0cf830b12: Downloading [====>>       1.408/4.5         JSWB2cea98C3: Down	15MB3e66f657: Pulling fs laver	] 290.988/29.
b0a0cf830b12: Downloading [=> [=> [] 881.4kB/29. 15MB0995b21: Waiting [====================================	fe401a45083b: Downloading [>	] 19.8kB/1.4
15MBda935b21:Waiting [======>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	b0a0cf830b12: Downloading [=>	] 881.4kB/29.
fe401a45083b: Downloading       [=======>       ] 562.8k8/1.4         Sd163cea9863: Downloading       [=======>>       ] 1.151MB/4.5         b0a0cf830b12: Downloading       [======>>       ] 1.47MB/29.         1SMB1a45083b: Downloading       [=====>>       ] 1.151MB/4.5         b0a0cf830b12: Downloading       [===>>       ] 1.759MB/1.4         b0a0cf830b12: Downloading       [===>>       ] 2.649MB/29.         1SMB3cea9803: Downloading       [====>>       ] 2.32MB/4.5         JSMB3cea9803: Downloading       [====>>       ] 3.234MB/29.         ISMB       Downloading       [====>>       ] 3.234MB/29.         ISMB       Downloading       [====>>       ] 3.234MB/29.         ISMB       Downloading       [=====>>       ] 4.099MB/4.5         Doadcf830b12: Downloading       [=====>>       ] 4.412MB/29.         ISMB       Downloading       [=====>>       ] 4.412MB/29.         ISMB       Downloading       [=====>>       ] 4.412MB/29.         ISMB       Downloading       [===	15MBd9a95b21: Waiting	
36163Ce398C3: Downloading [=>>       ] 1.151MB/4.5         b0a0cf830b12: Downloading [=>>       ] 1.131MB/4.5         b15MB1ce398C3: Downloading [=>>       ] 2.059MB/29.         15MB3ce398C3: Downloading [==>>       ] 2.059MB/29.         15MB3ce398C3: Downloading [==>>       ] 2.059MB/29.         15MB3ce398C3: Downloading [===>>       ] 2.32MB/4.5         b0a0cf830b12: Downloading [===>>       ] 3.234MB/29.         15MB       ] 3.825MB/29.         15MB       ] 4.412MB/29.	fe401a45083b: Downloading [====================================	562.8kB/1.4
D030CT830D12: D0Whl03dting [==>   1.47/B/29. 1.139MB1.4 b030Cf830b12: D0wnl0adting [===>   1.139MB/1.4 b030Cf830b12: D0wnl0adting [===>   2.649MB/29. 15MB34ce398C3: D0wnl0adting [===>   2.649MB/29. 15MB3cea98C3: D0wnl0adting [===>   2.649MB/29. 15MB3cea98C3: D0wnl0adting [===>   3.234MB/29. 15MB 36163cea98C3: D0wnl0adting [====>   3.234MB/29. 15MB 36163cea98C3: D0wnl0adting [====>   3.234MB/29. 15MB 36163cea98C3: D0wnl0adting [====>   3.825MB/29. 15MB3cea98C3: D0wnl0adting [====>   4.099MB/4.5 b030cf830b12: D0wnl0adting [====>   1.13MB/4.5 b030cf830b12: D0wnl0adting [=====>   1.13MB/4.5 b030cf830b12: D0wnl0adting [=====>   1.13MB/4.5 b030cf830b12: D0wnl0adt complete b1.542B	36163cea98c3: Downloading [========>	1.151MB/4.5
Instat430830: Downloading [===>1.139hB/1.4b0a0cf830b12: Downloading [==>>2.059MB/29.15MB3cca98C3: Downloading [===>2.049MB/29.15MB3cca98C3: Downloading [===>>2.049MB/29.15MB3cca98C3: Downloading [===>>3.234MB/29.15MB3cca98C3: Downloading [===>>3.234MB/29.15MB3cca98C3: Downloading [===>>3.504MB/29.15MB3cca98C3: Downloading [===>>>3.504MB/29.15MB3cca98C3: Downloading [===>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>	DUaUCT83UD12: DOWNLOADING [==>	1.4/MB/29.
Doad(13):012:Downloading [====>]1.759MB/4.515MB3cea98c3:Downloading [====>]2.32MB/4.52.3MB2.32MB/4.5]2.32MB/4.5b0a0cf830b12:Downloading [====>]3.234MB/29.15MB36163cea98c3:Downloading [====>]3.504MB/29.15MB36163cea98c3:Downloading [====>]3.504MB/4.5b0a0cf830b12:Downloading [====>]3.504MB/4.5b0a0cf830b12:Downloading [====>]3.825MB/29.15MBISMB3cea98c3:Downloading [====>]15MBJoan(F330b12:Downloading [====>]15MBJoan(F330b12:Pull complete]b0a0cf830b12:Pull complete]4.412MB/29.15MBJ11dcop5c3:Pull complete]4.412MB/29.15MBSold(F330b12:Pull complete]103.6MB/106J11dcop5c3:Pull complete]103.6MB/106J98245430c33:Pull complete]103.6MB/106J98d5360f88:Download complete]103.6MB/106J98d5360f88:Download complete]103.6MB/106J973f95d4f335:Download complete]103.6MB/106J973f95d4f336:Download complete]543339c21380:J0301cad complete]J54339c21380:Download completeJ973f95d4f336:Download complete]543339c21380:J973f95d4f336:Download complete]54339c21380:J973f95d4f336: <t< td=""><td>ISMB1a45083D: Downloading [====================================</td><td>1.139MB/1.4</td></t<>	ISMB1a45083D: Downloading [====================================	1.139MB/1.4
<pre>1 1.1.5/16/1.15 1 2.649MB/29. 1 5MB/29. 1 5MB/29. 3 .234MB/29. 3 .234MB/29.</pre>	15MB3cea98c3: Downloading [====================================	$1 \ 759 MB/25$
ISMB3cea98c3: Downloading [=====>       ]       2.32MB/4.5         33MB       b0a0cf830b12: Downloading [=====>       ]       3.234MB/29.         ISMB       36163cea98c3: Downloading [=====>       ]       3.504MB/4.5         S0a0cf830b12: Downloading [=====>       ]       3.825MB/29.         ISMB3cea98c3: Downloading [=====>       ]       3.825MB/29.         ISMB3cea98c3: Downloading [=====>       ]       3.825MB/29.         ISMB3cea98c3: Downloading [=====>       ]       4.4099MB/4.5         b0a0cf830b12: Downloading [=====>       ]       4.412MB/29.         ISMB       ]       4.412MB/29.         b0a0cf830b12: Pull complete       ]       4.412MB/29.         b463e6f657: Pull complete       ]       4.412MB/29.         b463e6f657: Pull complete       ]       982493521: Pull complete         b463e6f6f83: Download complete       ]       103.6MB/106         .9MBd5360f88: Download complete       ]       103.6MB/106         .9M3922233	b0a0cf830b12: Downloading [====>	1 2.649MB/29.
33MB b0a0cf830b12: Downloading [====> ] 3.234MB/29. 15MB ] 3.234MB/29. 36163cea98c3: Downloading [====> ] 3.825MB/29. 15MB3cea98c3: Downloading [====> ] 4.099MB/4.5 15MB3cea98c3: Downloading [====> ] 4.099MB/4.5 16a0cf830b12: Downloading [====> ] 4.412MB/29. 15MB b0a0cf830b12: Pull complete 51613cea98c3: Pull complete 56163cea98c3: Pull complete 5463a66f657: Pull complete 5463a66f657: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Download complete 5b691e4e12b4: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 566316335: Download complete 5673195802589: Download complete 5743350: Download complete 5743350: Download complete 5743350: Download complete 5743350: Download complete 5745351; Download complete 574535; Download comp	15MB3cea98c3: Downloading [====================================	2.32MB/4.5
b0a0cf830b12: Downloading [====> ] 3.234MB/29. 15MB ] 3.234MB/29. 36163cea98c3: Downloading [=====> ] 3.504MB/4.5 b0a0cf830b12: Downloading [=====> ] 4.099MB/4.5 b0a0cf830b12: Downloading [=====> ] 4.099MB/4.5 b0a0cf830b12: Downloading [=====> ] 4.412MB/29. 15MB b0a0cf830b12: Pull complete 56163cea98c3: Pull complete 56163cea98c3: Pull complete 54663e66f657: Pull complete 54691a45083b: Pull complete 54691a45083b: Pull complete 55691e4e12b4: Download ing [======> ] 103.6MB/106 .9MBd5306M8: Download complete 5247a27db111: Download complete 55c19b802589: Download complete 5	33MB	
15MB 36163cea98c3: Downloading [=====> ] 3.504MB/4.5 b0a0cf830b12: Downloading [=====> ] 4.099MB/4.5 b0a0cf830b12: Downloading [=====> ] 4.412MB/29. 15MB b0a0cf830b12: Pull complete b311dac095c3: Pull complete 36163cea98c3: Pull complete 36163cea98c3: Pull complete 938245430c33: Pull complete 938245430c33: Pull complete 5b691e4e12b4: Downloading [======> ] 103.6MB/106 9980505688: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 9373F95d4f335: Download complete 973F95d4f335: Download complete 94339e21380: Download complete 5kB	b0a0cf830b12: Downloading [====>	] 3.234MB/29.
<pre>36163cea98c3: Downloading [=====&gt;&gt; ] 3.504MB/4.5 b0a0cf830b12: Downloading [====&gt;&gt; ] 4.099MB/4.5 b0a0cf830b12: Downloading [====&gt;&gt; ] 4.099MB/4.5 b0a0cf830b12: Pull complete b311dac095c3: Pull complete 36163cea98c3: Pull complete b4f63e66f57: Pull complete 4463e66f57: Pull complete 9a8245430c33: Pull complete 5b9ed9a95b21: Pull complete 5b9ed9a95b21: Pull complete 5b9ed9a95b21: Pull complete 5b9ed9a95b21: Pull complete 5b9ed9a95b21: Pull complete 5b9ed9a95b21: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 5362983: Download complete 54739221380: Download complete 54739221380: Download complete</pre>	15MB	
<pre>36163Ce398C3: Downloading [=====&gt;&gt; ] 3.304Hb/4.5 b0a0cf830b12: Downloading [====&gt;&gt; ] 4.09Mb/4.5 b0a0cf830b12: Downloading [====&gt;&gt; ] 4.09Mb/4.5 b0a0cf830b12: Downloading [====&gt;&gt; ] 4.409Mb/4.5 b0a0cf830b12: Downloading [====&gt;&gt; ] 4.412MB/29. 15MB b0a0cf830b12: Pull complete b311dac095c3: Pull complete fe401a45083b: Pull complete ga8245430c33: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 5b691e4e12b4: Downloading [======&gt;&gt; ] 103.6MB/106 .9MBd5360f88: Download complete 5247a27db111: Download complete 55c19b802589: Download complete 973f95d4f335: Download complete 973f95d4f335: Download complete 694339e21380: Download complete 5.5kB</pre>		
DodoCrosofiz: Downloading [=====> ] 4.099MB/4.5 boa0cf830b12: Downloading [====> ] 4.412MB/29. 15MB boa0cf830b12: Pull complete b311dac095c3: Pull complete 36163cea98c3: Pull complete b463e66f657: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 5b691e4e12b4: Downloading [======> ] 103.6MB/106 .9MBd5360f88: Download complete 5247a27db111: Download complete 5247a27db111: Download complete 54c19b802589: Download complete 973f95d4f335: Download complete 694339e21380: Download complete 5.5kB	36163Cea98C3: Downloading [====================================	3.504MB/4.5
<pre>1 100000000000000000000000000000000000</pre>	15MB3cea98c3: Downloading [>	1  2.025  mb/29.
<pre>15MB b0a0cf830b12: Pull complete b311dac095c3: Pull complete 36163cea98c3: Pull complete fe401a45083b: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [=======&gt;] 103.6MB/106 .9MBd5360f88: Download complete 0292c71f3c2d: Download complete 5247a27db111: Download complete 53a42a5269c3: Download complete 5fc19b802589: Download complete 973f95d4f335: Download complete 694339e21380: Download complete .5kB</pre>	b0a0cf830b12: Downloading [======>	1 4.412MB/29
b0a0cf830b12: Pull complete b311dac095c3: Pull complete 36163cea98c3: Pull complete b4f63e66f657: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================	15MB	]
b311dac095c3: Pull complete 36163cea98c3: Pull complete b4f63e66f657: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================	b0a0cf830b12: Pull complete	
36163cea98c3: Pull complete fe401a45083b: Pull complete b4f63e66f657: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================	b311dac095c3: Pull complete	
<pre>fe401a45083b: Pull complete b4f63e66f657: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================</pre>	36163cea98c3: Pull complete	
b4f63e66f657: Pull complete 9a8245430c33: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================	fe401a45083b: Pull complete	
<pre>9a8245430c33: Pull complete 74c893d239e5: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================</pre>	b4f63e66f657: Pull complete	
<pre>74C8930239eS: Pull complete 789ed9a95b21: Pull complete 5b691e4e12b4: Downloading [====================================</pre>	9a8245430c33: Pull complete	
<pre>&gt;</pre>	74C893d239e5: Pull complete	
<pre>&gt;</pre>	5691e4e12b4: Downloading [	1 103 6MB/106
0292671f3c2d: Download complete 5247a27db111: Download complete b3a42a5269c3: Download complete 5fc19b802589: Download complete 973f95d4f335: Download complete 694339e21380: Download complete .5kB	9MBd5360f88: Download complete	] 103.000/100
5247a27db111: Download complete b3a42a5269c3: Download complete 5fc19b802589: Download complete 973f95d4f335: Download complete 694339e21380: Download complete .5kB	0292671f3c2d: Download complete	
b3a42a5269c3: Download complete 5fc19b802589: Download complete 973f95d4f335: Download complete 694339e21380: Download complete .5kB	5247a27db111: Download complete	
<pre>5fc19b802589: Download complete 973f95d4f335: Download complete 694339e21380: Download complete .5kB</pre>	b3a42a5269c3: Download complete	
973f95d4f335: Download complete 694339e21380: Download complete .5kB	5fc19b802589: Download complete	
694339e21380: Download complete .5kB ☆	973f95d4f335: Download complete	
SKB	694339e21380: Download complete	
	. SKB	

出现下面这个说明启动完成:

Digest: sha256:480a8d63af5057f7d81437dc138da941998657779b6c0150fd5e595b6ff72c71
Status: Downloaded newer image for registry.cn-hangzhou.aliyuncs.com/fastgpt/one-api:v0.6.6
Creating sandbox done
Creating mysql done
Creating mongo done
Creating pg done
Creating oneapi done
Creating fastgpt done

用 docker-compose ps

Name	Command	State	Ports
fastgpt mongo mysql	sh -c nodemax-old-space bash -c openssl rand -base docker-entrypoint.shdef	Սp Սp Սp	0.0.0.0:3000->3000/tcp,:::3000->3000/tcp 0.0.0.0:27017->27017/tcp,:::27017->27017/tcp 0.0.0.0:3306->3306/tcp,:::3306->3306/tcp, 33060/tcp
oneapi pg sandbox	/one-api docker-entrypoint.sh postgres docker-entrypoint.sh node	Uр Uр Uр	0.0.0.0:3001->3000/tcp,:::3001->3000/tcp 0.0.0.0:5432->5432/tcp,:::5432->5432/tcp

### 本机浏览器中输入: http://8.154.23.247:3000/

如果能得到如下画面:

$\leftarrow  \rightarrow$	c 🛛	A 不安全 8.1	54.23.247:30	000/app/list						: Google 智能镜头	\$ £	3   6
🐝 稀土掘金	♥ 微信	言商户注册 🧲	微信公众平台	<b>③</b> Python 教程	🛅 哔哩哔哩 (゛-゛)つ	🖞 AGI大模型全栈实践	🧿 智谱清言智能体	👌 讯飞星火大模型	⊗ 大模型排名SuperC	∞⊷ 超算互联网	*	□ 所有‡
	全部	简易应用	工作流	插件					٩	搜索应用		+ 新建
(												
立た合												
知识库												
O 低号							۲ ۲					
						还没有应用	月,快去创建一个吧!					

## 输入: http://8.154.23.247:3001/ 如果能得到这个:

← → C (▲ 不安全)	8.154.23.247:3001/setting
💸 稀土掘金 🗲 微信商户 💈	<sup>「网站信息</sup> 」公众平台 🔞 Python 教程 🛗 哔哩哔哩 (゜-゜)つ… 🖞 AGI大模型全栈实践… 🥝 智谱清言智能体 🔥 讯飞星火大模型
	🌀 One API 💣 首页 🍶 渠道 🎤 令牌 💲 兑换 🍞 充值 🚨 用户 🗏 日志 🌣 设置
	个人设置 运营设置 系统设置 其他设置
	<b>进                                    </b>
	更新个人信息 生成系统访问令牌 复制邀请链接 删除个人账户
	<b>账号绑定</b>

就说明系统启动正常。



xinference是一个傻瓜式操作的大模型部署平台,有ui交互界面。

回到root,执行如下命令:

```
mkdir xinference
cd xinference
docker run \
  -d \
  -e XINFERENCE_MODEL_SRC=modelscope \
  -v $(pwd)/.xinference:/root/.xinference \
  -v $(pwd)/.cache/huggingface:/root/.cache/huggingface \
  -v $(pwd)/.cache/modelscope:/root/.cache/modelscope \
  -p 9997:9997 \
  -gpus all \
  registry.cn-hangzhou.aliyuncs.com/xprobe_xinference/xinference \
  xinference-local -H 0.0.0.0
```

这里它会下载很多大文件,所以要求服务器内存一定要够大。

如果命令执行出现问题,可以尝试安装 nvidia container toolkit:

```
sudo apt install curl -y
curl -fsSL https://nvidia.github.io/libnvidia-container/gpgkey | sudo gpg --
dearmor -o /usr/share/keyrings/nvidia-container-toolkit-keyring.gpg && curl -s -
L https://nvidia.github.io/libnvidia-container/stable/deb/nvidia-container-
toolkit.list | sed 's#deb https://#deb [signed-by=/usr/share/keyrings/nvidia-
container-toolkit-keyring.gpg] https://#g' | sudo tee
/etc/apt/sources.list.d/nvidia-container-toolkit.list
sudo apt-get update
sudo apt-get install -y nvidia-container-toolkit
```

这个命令执行完了之后要重启docker: sudo systemctl restart docker , 之后再次执行第一个 命令。

用 docker ps 能看到如下图所示, 说明启动成功:

root@iZbp13vmycyz2y29guorq1Z:~/fastgpt/xinference/xinference/xinference# docker ps											
CONTAINER ID IMAGE	COMMAND	CREATED	STATUS								
PORTS	NAMES										
76c183e30e56 registry.cn-hangzhou.aliyuncs.com/x	probe_xinference/xinference "xinference-local -H"	About a minute ago	Up About a								
minute 0.0.0.0:9997->9997/tcp, :::9997->9997/tc	p inspiring hermann										

在我本机浏览器中输入: http://8.154.23.247:9997/ 如果出现:

掘金	〇 微信商户注册 (学): 微信商户注册 …qq.com/apply/applyme	微信公众平台 nt4normal_progr	<b>③</b> Python 教程   芭 哔哩哔哩 (* - ess	*)つ <u>2</u> AGI大模型全桟?	实践 🥏 智谱清言智能体 🔥 讯기	5星火大模型 🧇 大模型排	名SuperC 🔤 🚳 地理全	<u>比上传</u> ※   C
	Xinferenc	e	Launch Mode	1				
<b>5</b> 8	Launch Model		LANGUAGE MODELS EMBE	DDING MODELS RERAN	K MODELS IMAGE MODELS	AUDIO MODELS VIE	DEO MODELS CUSTOM MODEL	S
٥	Running Models	>	Model Ability -	Status -	Search for model name and	description		Type / to search
ŧ	Register Model	>	aquila2	*	aquila2-chat	*	aquila2–chat–16	šk ☆
	Cluster Information	>	zh Aquila2 series moo	els are the base	zh Aquila2-chat series m	odels are the	zh AquilaChat2–16k series	s models are
0	Contact Us	>	language models		chat models		the long-text chat moc	els
			2K context length	<b>≕,</b> generate model	2K context length	E chat model	16K context length	E chat model
			baichuan-2		baichuan-2-ch	at 🕁	c4ai-command-	-r ☆
								t) ot in

则说明启动xinference成功。

至此作业所需的所有准备工作都完毕。

## 部署本地模型

首先在xinference交互界面中,我们部署3个模型,这3个都是接下来的实验必须的。

- chat模型 Ilama-3.1-instruct
- embeding模型 m3e-base
- rerank模型 bge-reranker-v2-m3

以第一个为例,在模型这一栏搜索 llama-3.1-instruct,点进去,然后按照下面图的逐个配置:

$\bigotimes$	Xinferenc	e	Launch Model	Ilama-3.1-instruct	
			LANGUAGE MODELS EMBEDDING MOD	Transformers	•
88	Launch Model	>		Model Format	•
٢	Running Models	>	Model Ability - Status	Model Size	
÷	Register Model	>	llama 31-instruct	8	×
	Cluster	>	en de fr it pt hi	None	×
	Ountent Us	,	th The Llama 3.1 instruction tune	N-GPU	
()	Contact US	>	models are optimized for diale use cases and outperform ma the available open source cha	Replica	
			131K context length cha	1	
				Optional Configurations 🗸	
				Additional parameters passed to the inference engine: Transfo	rmers 🕂
				23	r.

点击下面的小火箭图标,系统就会自动给你下载并且部署。

但是很不幸,我的xinference无法下载所有模型。重启xinference之后问题解决,从服务器的带 宽监控来看,正在下载,耐心等待。



刚才的小火箭图标也变成了loading:

Launch Model	llama-3.1-instruct
	Model Engine
LANGUAGE MODELS EMBEDDING MOD	Transformers -
	Model Format
Model Ability	pytorch -
	Model Size
	8
llama-3.1-instruct	Quantization
en de fr it pt hi	none
th	N-GPU
The Llama 3.1 instruction tune models are optimized for dial	1 ~
use cases and outperform ma	C Replica
131K	1
context length cha	Optional Configurations 🗸
	Additional parameters passed to the inference engine: Transformers 😝
	(

等他下完吧。这会可千万别手贱,不然会前功尽弃。刚才就是因为动了一下,才频繁报错的。 上面是在modelscope(大模型国内镜像源)去下载,所以应该不存在网络问题。道理我都懂, 为什么这里不显示下载进度???

### 这3个模型已经都在running状态了。

LANGUAGE MODELS	EMBEDDING MODELS RI	RANK MODELS	IMAGE MODELS	AUDIO MODELS	S VIDEO MODEL	S FLEXIBLE	EMODELS
ID	Name	Address	GPU Indexes	Size	Quantization	Replica	Actions
llama-3.1-instruct	llama-3.1-inst	0.0.0.0:36441	0	8	none	1	
LANGUAGE MODELS	EMBEDDING MODELS	RERANK MODEL	LS IMAGE MC	DELS AUDI	O MODELS VIE	DEO MODELS	FLEXIBLE MODELS
ID	Name		Address	GP	U Indexes	Replica	Actions
m3e-base	m3e-bas	e	0.0.0.0:35039	0		1	Ō
LANGUAGE MODELS	EMBEDDING MODELS	RERANK MODELS	IMAGE MODE	LS AUDIO M	ODELS VIDEO N	MODELS F	LEXIBLE MODELS
ID	Name	Ą	Address	GPU In	dexes	Replica	Actions
bge-reranker-v2-m3	bge-reran	ker-v2-m3 0	0.0.0.0:35995	0		1	Ō

## 将本地模型加入到oneapi渠道中

S One API	😭 首页	♣ 渠道	₽ 令牌	<b>\$</b> 兑换	🍹 充值	▲ 用户	■ 日志	✿ 设置	关于	root 👻
更新渠道信息										
类型*										
自定义渠道										•
Base URL										
http://8.154.23.24	7:9997									
名称*										
xinference										
分组*										
default 🗙										•
模型*										
llama-3.1-instruct	× m3e-base	e 🗙 🛛 bge-	reranker-v2-	m3 🗙						•
填入相关模型	填入所有模型	清陵 清陵	涂所有模型	输入自定	义模型名称		填入			
模型重定向										
此项可选,用于修t { "gpt-3.5-turbo- "gpt-4-0314": ", "gpt-4-32k-031 }	改请求体中的模型 0301": "gpt–3.5 gpt–4", !4": "gpt–4–32k	型名称,为- 5-turbo", "	-个 JSON 字	符串,键为译	f求中模型名₹	尔,值为要替	换的模型名称	,例如:		
密钥*										

注意这里的 8.154.23.247是我的服务器公网ip地址。下面模型栏,填的是刚部署的3个大模型。 经过我的测试,发现llama-3.1-instruct这个模型回答存在问题,总是给我一个json,决定重新下 一个 baichuan-2-chat 部署上去。

## fastgpt创建对话

创建对话,并选择刚才的baichuan-2-chat,看下图,已经可以正常聊天。



## 附加知识库

### 知识库就绪的情况如下

	F test知识库	文件 <b>(2)</b>				Q 搜索		인 新建/导入
○ 聊天	🔚 通用知识库	名称	训练模式	数据总量	创建/更新时间	状态	启用	
Ċ	山 数据集	🔒 民法典简要.pdf	直接分段	5	2024-10-18 00:13 2024-10-18 00:13	• 已就绪		
工作台	<ul> <li>         · 搜索测试         · · ·</li></ul>	🔓 实验室守则.pdf	直接分段	3	2024-10-18 00:13 2024-10-18 00:13	• 已就绪		
<b>O</b> 账号								

在附加知识库的过程中,有一些配置会影响知识库是否正常。 比如下面的,选择直接分段,可能导致知识库无法正常就绪,尝试更换为直接分段(猜测这个 和文档结构有关,总之要想办法变更配置让知识库可以正常使用。)

<b>下</b> test知识库	← 上一歩					
🔚 通用知识库		✓ 选择文件	2 数	文据处理	3 上传数据	
<ul> <li>山 数据集</li> <li>② 搜索测试</li> <li>③ 配置</li> </ul>	◎数据处理	参数		<b> </b>		
	训练模式	<ul> <li>● 直接分段 ⑦</li> <li>○ 増强处理(实验) ⑦</li> </ul>		☑ 民法典简要.md		۲
		○ 问答拆分 ⑦				
	处理方式	• <b>自动</b> 自动设置分割和预处理规则				
		自定义规则 自定义设置分制和预处理规则				
			下一步			

QA 训练排队(空闲)

另一个需要注意的是,下图中有两个配置会影响到知识库能否正常加载:

第一个就是 m3e-base 索引模型,这个模型在 **安装xinference**这一章节已经部署了,所以我们 就可以选它。

另一个就是文件处理模型,我们这里选择的是 ERNIE-Speed-128K(这个是我按照讲义的步骤,在 百度智能云平台上注册的一个应用,然后在oneapi中添加了此渠道,所以我可以用它来帮我处理文件内容)

<b>test知识库</b> <b>通</b> 用知识库	知识库 ID	6711259aabddfc4876f31c6a
	索引模型	😁 m3e-base 🗸 🗸
<ul><li>     数据集     </li><li>     登 搜索测试     </li></ul>	单条数据上限	512
<b>診</b> 配置	文件处理模型	C ERNIE-Speed-128K ~
		compared when the second secon
	知识库头像	🧔 glm-4v
	知识库名称	C ERNIE-Speed-128K
		🕼 gpt-4o-mini
	介绍	

## 最后的测试

知识库就绪,本地模型就绪。现在来进行测试。

在我们新建的机器人配置中,选择我们本地部署的 baichuan-2-chat,同时关联刚才加进去的知识库,可以看到下图右下角的回答中,已经在饮用我上传的pdf文件内容了。

全部应用 / test建议机器人	应用配置 发布渠道 对话日志 ● 未发布	② 发布
<ul> <li>⑦ 对话</li> <li>② 设置</li> <li>● U 仅协作者</li> </ul>	<b>调试预览</b> ▶ 实验室守则,pdf	<u>ے</u>
🥶 AI 配置	3条引用         5条上下文         3.78s         查看详情	
AI 模型 🤗 baichuan-2-chat	िष्ठे । ८	
提示词 ⑦ ① 输入 / 可选择变量	民法典的意	意义是什么
3/	ⓒ ⓒ 则 ∠ 《中华人民共和国民法典》是新中国第一部以"法典"命名的法律,也是新中国成立以来的第六部法律。它2021年1月1日起施行,共有7编,合计1260条,涵盖了民事活动的各个方面,构成了中国民事法律的基本架。民法典的意义在于它将中国的民事法律体系带入了一个全新的阶段,强调私法自治与公序良俗的平衡为保障法治中国、构建和谐社会提供了坚实的法律基础。	3自 体框 断,
关联知识库 + 选择 IC 参数	•• 51月 • 民法典简要.pdf	
搜索模式     引用上限     最低相关度     结果重排     问题优化       砂语义检索     1500     0.4     X     baichuan-2-c	4条引用         5条上下文         6.04s         查看详情	
<b>F</b> test知识库	输入问题,发送 [Enter]/换行 [Ctrl(Alt/Shift) + Enter]	

### 实验成功,下一步把聊天机器人发布出去。

← 退出聊天	😭 test建议机器人	实验室可以抽烟么 O 2 条记录 💩 baichuan-2-chat	
最近使用更多 € ■ test建议机器人	<ul> <li>● 新对话</li> <li>● 实验室可以抽烟么</li> </ul>	<ul> <li>○ 小 ▲ マ</li> <li>● 欠 他 室</li> <li>● 欠 他 室</li></ul>	(学)
		输入问题,发送 [Enter]/换行 [Ctrl(Alt/Shift) + Enter]	

甚至可以发布成一个网页给大众使用。

	全部应用 / test建议机器人	应用配置 发布渠道 对话日志		
	◆ <b>免登录窗口</b> ◆ 全等链接给其他用户 天愿登录即可直接进行	登 选择使用方式	×	
聊天	免登录窗口 ③		F	
) 知识库	名称	展示历史对话	•	开始使用
4 账号	144 20 A 19 20	将下面链接复制到浏览器打开	G	Ликт
		http://8.154.23.247:3000/chat/share?shareId=19nfzixgztcy2hdg87yz6yqx		

### http://8.154.23.247:3000/chat/share?shareId=19nfzixgztcy2hdg87yz6yqx

### 只不过正式发布出去之后,在回答中不会像测试过程中看到原文的引用。



实验成功。

记得把按量计费的阿里云服务器关机,挺贵的。