|  |  |
| --- | --- |
| 文件编号： |  |
| 密级： |  |
| 项目ID： |  |
| 总页数： |  |

**AI竞赛接口平台**

**接口文档**

版本 <V 1>

|  |  |  |  |
| --- | --- | --- | --- |
| 拟制 | 潘旸 | 日期 | 2023年4月8日 |
| 审核 | 苏振文 | 日期 |  |
| 批准 |  | 日期 |  |

目录

1. 概述 3

2. 整体平台架构 3

3. 竞赛平台接口总览 3

3.1. 最新系统接口清单 3

3.2. 安全机制 3

3.3. 架构说明 3

4. 各厂家接口 3

4.1. chat 3

4.2. whisper 5

# 概述

本接口仅允许院内环境网络使用，通过AI竞赛平台跳链至各AI供应商厂家，如在线openai各大接口、在线文心一言各大接口、离线LLama、离线ChatGLM、离线whisper、离线stablediffsion。目前仅推出openai的chat、whisper接口供使用，后续将陆续推出各类AI接口。

# 整体平台架构

Openai

Stablediffsion

Llama等

客户端

AI竞赛平台

# 竞赛平台接口总览

## 最新系统接口清单

|  |  |  |  |
| --- | --- | --- | --- |
| 接口厂家 | 接口名 | 访问URL | 备注 |
| openai | chat | http://172.16.0.81:9999/api/openai/chat | openai的聊天接口 |
| openai | whisper | http://172.16.91.30:19990/asr\_translate | 本地化openai的语音识别接口 |

## 安全机制

采用header头部装载Authorization-Key，该密钥为uuid编码，请在竞赛平台APIKEY申请使用即可开通，如API短时调用过于频繁将会被限制请求频率。

## 架构说明

系统timeout时间为30秒。所有接口均需要使用Authorization-Key进行鉴权

# 各厂家接口

## chat

接口说明：对接openai的gpt3.5-turbo接口,不含上下文、不含session管理，无需传入openai要求的api-key，api-key由竞赛平台进行统一管理，但需要传入提前申请好的竞赛平台Authorization-Key。接口支持stream与非stream两种模式。

请求说明：本请求json与openai官方接口入参完全一致，全部参数可使用

入参说明

|  |  |  |  |
| --- | --- | --- | --- |
| 参数名 | 必填 | 类型 | 说明 |
| model | 是 | str | 目前chat接口只支持gpt-3.5-turbo模型 |
| messages | 是 | jsonlist | messages即会话上下文，注意子参数system为必填项，代表角色，后续role中代表AI的回答必须为assistant，user则跟user对应即可 |
| temperature |  | int | 默认为1，0~2，数值越高创造性越强 |
| top\_p |  | int | 默认为1，0~1，与temperature相斥，不可一起使用 |
| n |  | int | 默认为1，chatgpt对一个提问生成多少个回答，建议为1 |
| stream |  | bool | 可选，默认为False，设置为True和网页效果类似，需监听事件来解析，建议非聊天应用采用False |
| stop |  | str | chatgpt遇到stop里的字符串时停止生成内容 |
| max\_tokens |  | int | 默认无穷大，如果设置了，需要满足max\_tokens+message\_token<=4096,这是生成文本的最大长度，注意会触发openai的超长报错。注意整个上下文长度，可使用tiktoken库来计算即将入参的message长度 |
| presence\_penalty |  | int | 默认为0，-2~2，越大越不允许跑题 |
| frequency\_penalty |  | int | 默认为0，-2~2，越大越不允许复读机 |
| logit\_bias |  |  | 默认无，影响特定词汇的生成概率，需要另行查看说明 |
| user |  | str | 默认无，设置用户名 |

请求示例

|  |
| --- |
| #POST  http://172.16.0.81:9999/api/openai/chat  #request.header  Content-Type:application/json  Authorization-Key:b2828374d11511edad5b0550563633b1  #body  {      "stream":**false**,      "model":"gpt-3.5-turbo",      "temperature":1,      "messages":[          {"role":"system","content":"你是一个小助手"},          {"role":"user","content":"我是孙悟空"},          {"role":"assistant","content":"你好,悟空"},          {"role":"user","content":"今天师傅有没有被抓走？"}      ]  } |

返回示例

|  |
| --- |
| {      "msg": {          "id": "chatcmpl-72ytHzZDf0LlUGN5FiRWNn9IVQhtm",          "object": "chat.completion",          "created": 1680945075,          "model": "gpt-3.5-turbo-0301",          "usage": {              "prompt\_tokens": 61,              "completion\_tokens": 61,              "total\_tokens": 122          },          "choices": [              {                  "message": {                      "role": "assistant",                      "content": "很抱歉，作为一个AI助手，我无法得知现实世界中的事件。不过，如果你有烦恼或困惑，我可以尽力给你建议和帮助。"                  },                  "finish\_reason": "stop",                  "index": 0              }          ]      },      "err\_type": **null**,      "err\_msg": **null**  } |

参数异常示例

|  |
| --- |
| {      "msg": **null**,      "err\_type": "openai\_error",      "err\_msg": "invalid\_request\_error,3 is greater than the maximum of 2 - 'temperature'"  }  #意思是temperature参数不可以超过2 |

## whisper

接口说明：对接本地openai开源的whisper模型,用于上传语音文件，识别语音内容，目前无需进行header校验。

请求说明：请求参数仅需传一个parameter的language；同时request.bod只需传一个multipart/from-data的file即可完成语音识别输出文字信息，详情可查阅http://172.16.91.30:19990/docs#。

请求示例

|  |
| --- |
| #POST  http://172.16.91.30:19990/asr\_translate  #Query Param  language=Chinese  #request.header  Content-Type:multipart/from-data  #body  file<各类音频格式如mp3\m4a\wav> |

返回示例

|  |
| --- |
| {      "message": "success",      "time": 7.792051553726196,      "filename": "新录音.m4a",      "transcription": "喂喂喂",      "translation": " Hello?"  } |

异常示例

|  |
| --- |
| {      "detail": [          {              "loc": [                  "body",                  "file"              ],              "msg": "field required",              "type": "value\_error.missing"          }      ]  }  #意思是丢失body确实file这个key |