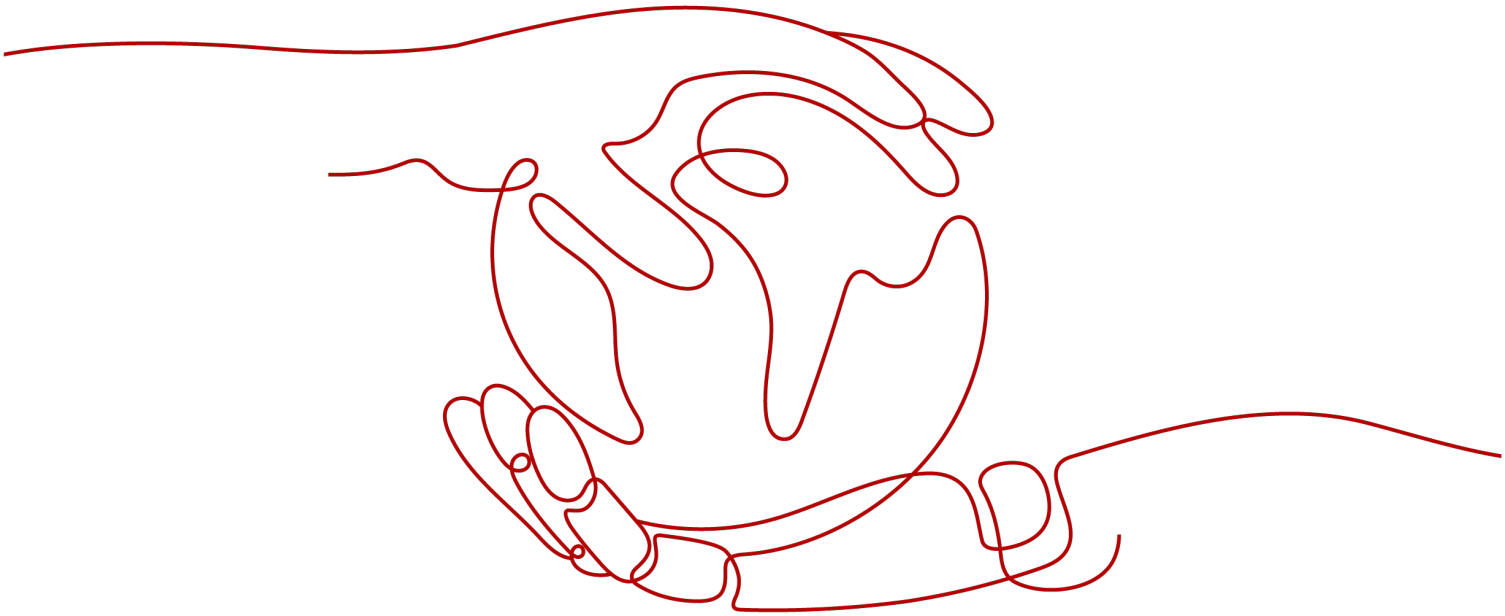
应用平台

API 参考

文档版本 03

发布日期 2024-09-23



华为云计算技术有限公司

版权所有 © 华为云计算技术有限公司 2024。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目 录

 版权所有 © 华为云计算技术有限公司

1 使用前必读

欢迎使用应用平台（AppStage），华为云AppStage是基于平台工程（Platform Engineering）理念打造的下一代应用全生命周期管理和AI原生应用生命周期管理平台，帮助客户快速高效地实现传统应用及AI原生应用全生命周期管理，为应用构建、运维和运营等生命周期管理活动提供自助式服务能力。

目前AppStage的AI原生应用引擎提供API供您调用。在调用AppStage的AI原生应用引擎API之前，请确保已经充分了解AppStage的相关概念，详细信息请参见AppStage服务的**产品介绍**。

终端节点

终端节点即调用API的**请求地址**，不同服务不同区域的终端节点不同，AppStage目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。

基本概念

- 大模型推理服务

  直接调用预置大模型提供API完成推理过程。

- 私有模型部署

  针对已经微调训练好的模型，如需评测此模型效果，或通过应用调用此模型，则需将模型部署为线上服务。

- 向量知识库

  通过引入多种类型和格式的企业知识，将数据转化为向量，并利用高效的存储和索引方式进行查询，实现基于检索增强的大模型能力。

- 工作流

  任务流程的细化分解是一种有效策略，能够简化系统架构，并降低对大语言模型能力的过度依赖。通过将繁复的工作拆解为一系列独立节点，不仅增强了复杂任务处理的效率，还在很大程度上提升了整个系统的透明度、鲁棒性和错误容忍度。这种方法使得LLM的应用范围得以扩大，即便面对高度复杂的任务也能表现出色。

版权所有 © 华为云计算技术有限公司

2 API 概览

AppStage接口的分类与说明如表2-1所示。

表 2-1 API 概览

类型	说明
文本对话	文本对话类模型服务调用。
文本向量化	文本向量化类模型服务调用。
调用Agent	调用用户发布的Agent。
调用技能	调用用户配置的技能。
调用流	触发工作流调用接口。
检索知识库数据	用于检索指定知识的数据。

版权所有 © 华为云计算技术有限公司

3 如何调用 API

3.1 构造请求

本节介绍REST API请求的组成，并以调用AppStage服务的文本对话接口说明如何调用API。

您还可以通过这个视频教程了解如何构造请求调用API：**https://bbs.huaweicloud.com/videos/102987** 。

请求 URI

请求URI由如下部分组成。

{URI-scheme} :// {Endpoint} / {resource-path} ? {query-string}

尽管请求URI包含在请求消息头中，但大多数语言或框架都要求您从请求消息中单独传递它，所以在此单独强调。

表 3-1 URI 中的参数说明

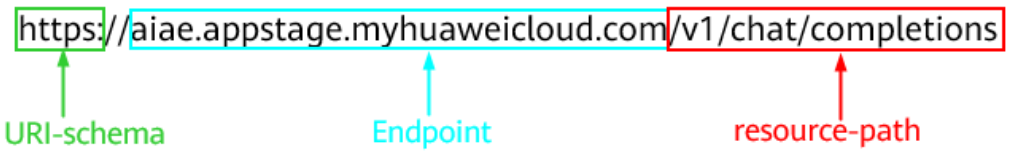
参数	描述
URI-scheme	表示用于传输请求的协议，当前所有API均采用**HTTPS**协议。
Endpoint	指定承载REST服务端点的服务器域名或IP，不同服务不同区域的Endpoint不同，当前AppStage服务只在“华北-北京四”部署，Endpoint为“aiae.appstage.myhuaweicloud.com”。
resource-path	资源路径，即API访问路径。从具体API的URI模块获取，例如文本对话API的resource-path为“/v1/chat/completions”。
query-string	查询参数，是可选部分，并不是每个API都有查询参数。查询参数前面需要带一个**“?”**，形式为**“参数名=参数取值”**，例如**“?limit=10”**，表示查询不超过10条数据。

版权所有 © 华为云计算技术有限公司

例如，您需要调用AppStage在“华北-北京四”区域的文本对话接口，则需使用“华北-北京四”区域的Endpoint（aiae.appstage.myhuaweicloud.com），并在文本对话的URI部分找到resource-path（/v1/chat/completions），拼接起来如下所示。

```
https://aiae.appstage.myhuaweicloud.com/v1/chat/completions
```

图 3-1 URI 示意图



说明

为查看方便，在每个具体API的URI部分，只给出resource-path部分，并将请求方法写在一起。这是因为URI-scheme都是HTTPS，而Endpoint在同一个区域也相同，所以简洁起见将这两部分省略。

请求方法

HTTP请求方法（也称为操作或动词），它告诉服务你正在请求什么类型的操作。

表 3-2 HTTP 方法

方法	说明
GET	请求服务器返回指定资源。
PUT	请求服务器更新指定资源。
POST	请求服务器新增资源或执行特殊操作。
DELETE	请求服务器删除指定资源，如删除对象等。
HEAD	请求服务器资源头部。
PATCH	请求服务器更新资源的部分内容。 当资源不存在的时候，PATCH可能会去创建一个新的资源。

在调用文本对话接口的URI部分，您可以看到其请求方法为“POST”，则其请求为：

```
POST https://aiae.appstage.myhuaweicloud.com/v1/chat/completions
```

请求消息头

附加请求头字段，如指定的URI和HTTP方法所要求的字段。例如定义消息体类型的请求头“Content-Type”，请求鉴权信息等。

详细的公共请求消息头字段请参见表3-3和表3-4。

版权所有 © 华为云计算技术有限公司

表 3-3 AK/SK 认证公共请求消息头

名称	描述	是否必选	示例
Content-Type	消息体的类型（格式），当前只支持application/json。	是	application/json
ts	毫秒时间戳。	是	1707101222000
nonce	请求唯一标识（UUID）。从**AK/SK认证**中获取。	是	-
ak	为AK/SK凭证文件中的AK明文。从**AK/SK认证**中获取。	是	-
sign	签名字符串。从**AK/SK认证**中获取。	是	-
resource-code	WiseAgent对外开放接口对应的唯一编码，每个接口唯一。请参考**表3-5**。	是	modelmarket.chat

表 3-4 API Key 认证公共请求消息头

名称	描述	是否必选	示例
Content-Type	消息体的类型（格式），当前只支持application/json。	是	application/json
Authorization	认证信息。格式为：Bearer ${API Key}	是	Bearer sk-5db9*********dd58

表 3-5 Resource-code

Resource-code	接口
modelrouter.chat	**文本对话**
modelrouter.embeddings	**文本向量化**

版权所有 © 华为云计算技术有限公司

Resource-code	接口
knowledgeBases.query.embeddata	检索知识库数据

请求消息体

请求消息体通常以结构化格式发出，与请求消息头中Content-type对应，传递除请求消息头之外的内容。如果请求消息体中参数支持中文，则中文字符必须为UTF-8编码。

每个接口的请求消息体内容不同，也并不是每个接口都需要有请求消息体（或者说消息体为空），GET、DELETE操作类型的接口就不需要消息体，消息体具体内容需要根据具体接口而定。

对于文本对话接口，您可以从接口的请求部分看到所需的请求参数及参数说明。将消息体加入后的请求如下所示。

```
POST https://aiae.appstage.myhuaweicloud.com/v1/chat/completions

{
   "model": "platform:chatglm3-6b",
   "messages": [
      {
         "role": "user",
         "content": "你好!"
      }
   ],
   "stream": false
}
```

到这里为止这个请求需要的内容就具备齐全了，您可以使用curl、Postman或直接编写代码等方式发送请求调用API。

3.2 认证鉴权

AppStage调用接口支持AK/SK和API Key认证鉴权。

AK/SK认证：通过AK（Access Key ID）/SK（Secret Access Key）进行API调用时的认证。

API Key：通过API密钥进行API调用时的认证。

AK/SK 认证

AK/SK认证就是使用AK/SK对请求进行签名，在请求时将签名信息添加到消息头，从而通过身份认证。

- AK(Access Key ID)：访问密钥ID。与私有访问密钥关联的唯一标识符；访问密钥ID和私有访问密钥一起使用，对请求进行加密签名。
- SK(Secret Access Key)：与访问密钥ID结合使用的密钥，对请求进行加密签名，可标识发送方，并防止请求被修改。

使用AK/SK认证时，您可以基于签名算法使用AK/SK对请求进行签名。详细的签名认证操作流程如下。

 版权所有 © 华为云计算技术有限公司

1. AK/SK申请

   使用具有管理员权限（admin）账号登录到WiseAgent主页，从右上角凭证管理进入到AK/SK管理页面，新建AK/SK。

   每个用户只能同时拥有两个AK/SK凭证。

2. AK/SK下载

   成功创建AK/SK后，会立刻弹出AK/SK凭证下载弹窗，下载后得到凭证文件。

   每个凭证仅能下载一次，且无法找回，请妥善保管凭证文件。

3. 使用AK/SK鉴权

   在请求头里添加如下header：

   ts: 毫秒时间戳

   nonce: 请求唯一标识（UUID）

   ak: 凭证文件中的AK明文

   resource-code: WiseAgent对外开放接口对应的唯一编码，每个接口唯一

   sign: 按如下规则拼接字符串"ts={变量名}&nonce={nonce}&ak={ak}"，对拼接得到的字符串plain进行SHA256散列后得到散列值hash，再使用凭证中的SK明文对刚才生产的hash进行再散列，最后进行Base64转码，得到签名字符串。

   签名样例代码（JAVA）：

```
public String sha256(String plain) {
    try {
        MessageDigest messageDigest = MessageDigest.getInstance("SHA-256");
        messageDigest.update(plain.getBytes(StandardCharsets.UTF_8));
        byte[] bytes = messageDigest.digest();
        StringBuffer hexBuffer = new StringBuffer();
        for (byte aByte : bytes) {
            String hex = Integer.toHexString(0xff & aByte);
            if (hex.length() == 1) {
                hexBuffer.append('0');
            }
            hexBuffer.append(hex);
        }
        return hexBuffer.toString();
    } catch (NoSuchAlgorithmException ignore) {
    }
}

public String hmacSha256(String hash, String sk) {
    try {
        Mac hmacSHA256 = Mac.getInstance("HmacSHA256");
        SecretKeySpec secretKeySpec = new SecretKeySpec(sk.getBytes(StandardCharsets.UTF_8),
"HmacSHA256");
        hmacSHA256.init(secretKeySpec);
        byte[] bytes = hmacSHA256.doFinal(hash.getBytes(StandardCharsets.UTF_8));
        return Base64.encodeBase64String(bytes);
    } catch (NoSuchAlgorithmException | InvalidKeyException ignore) {
    }
}
```

API Key 认证

API Key全称为应用程序接口密钥，是一种用于验证和授权API请求的代码。它通常是一串字符，用于识别调用API的应用程序和开发者。

1. 获取API Key

   以管理员身份登录AI原生应用引擎工作台，参考**创建API Key**获取。

2. 使用API Key鉴权

版权所有 © 华为云计算技术有限公司

调用时，在请求头里新增字段Authorization，值填写为Bearer ${API Key}，拼接起来如下所示。

```
Authorization:Bearer sk-5db9*********dd58
```

3.3 返回结果

状态码

请求发送以后，您会收到响应，包含状态码、响应消息头和消息体。

状态码是一组从1xx到5xx的数字代码，状态码表示了请求响应的状态，完整的状态码列表请参见**状态码**。

对于文本对话接口，如果调用后返回状态码为“200”，则表示请求成功。

响应消息头

对应请求消息头，响应同样也有消息头。例如，Content-Type=application/json

响应消息体

响应消息体通常以结构化格式返回，与响应消息头中Content-type对应，传递除响应消息头之外的内容。

对于文本对话接口，返回如下消息体。为篇幅起见，这里只展示部分内容。

```
{
    "created": 1718772336,
    "usage": {
        "completion_tokens": 23,
        "prompt_tokens": 45,
        "total_tokens": 68
    },
    "model": "chatglm3-6b",
    "id": "chatcmpl-xxx",
    "choices": [{
        "finish_reason": "stop",
        "index": 0,
        "message": {
            "role": "assistant",
            "content": "你好，有什么我可以帮助你的吗？"
        },
        "logprobs": null
    }],
    "object": "chat.completion"
}
```

当接口调用出错时，会返回错误码及错误信息说明，错误响应的Body体格式如下所示。

```
{
    "error_code": "AIAE.31001702",
    "error_msg": "Model not exists, please check and try again later!"
}
```

其中，error_code表示错误码，error_msg表示错误描述信息。

版权所有 © 华为云计算技术有限公司

4 API

4.1 模型调用

4.1.1 文本对话

功能介绍

调用大语言模型推理服务，根据用户问题，获取大语言模型的回答。

URI

POST /v1/chat/completions

请求参数

表 4-1 请求 Header 参数

参数	是否必选	参数类型	描述
Authorization	是	String	AI原生应用引擎鉴权API Key。 1.以管理员身份登录AI原生应用引擎工作台，在左侧导航栏选择“配置中心 > 平台租户鉴权”。 2.在“平台租户鉴权”页面，选择“平台API Key”页签，单击“新增平台API Key”。 3.在“新增平台API Key”对话框中的输入框设置API Key名称，用以区分API Key。 4.在弹出的下载窗口中单击“立即下载”，将API Key下载到本地查看。

版权所有 © 华为云计算技术有限公司

表 4-2 请求 Body 参数

参数	是否必选	参数类型	描述
messages	是	Array of **ChatCompletionRequestMessage** objects	文本对话消息体类。
model	是	String	模型服务调用唯一id字段。平台定义了4种模型服务: ● 平台预置模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“资产中心 > 大模型”，查看支持的模型服务。例如调用chatglm3-6b，model填写为platform:chatglm3-6b。 ● 平台接入模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“资产中心 > 大模型”，查看支持的模型服务。例如调用Baichuan2-Turbo模型服务，model填写为Baichuan2-Turbo即可。 ● 租户部署模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“Agent编排中心 > 我的模型服务 > 我部署的”，model填写为对应模型服务的模型服务调用ID。 ● 租户接入模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“Agent编排中心 > 我的模型服务 > 我接入的”，model填写为对应模型服务的模型服务调用ID。

版权所有 © 华为云计算技术有限公司

参数	是否必选	参数类型	描述
frequency_penalty	否	Number	介于-2.0和2.0之间的数字。 正值会根据文本中新Token的现有频率对其进行惩罚，从而降低模型重复相同行的可能性。 最小值：**-2** 最大值：**2** 缺省值：**0**
logit_bias	否	Map<String,Integer>	该参数接受一个JSON对象，将标记映射到从-100（禁止）到100（独占选择标记）的关联偏差值。 像-1和1这样的适度值将以较小的程度改变选择标记的概率。 使用logit_bias参数时，偏差被添加到模型生成的logits之前进行抽样。
max_tokens	否	Integer	返回体允许的最大token数。
n	否	Integer	返回体中包含的chatCompletionChoice数量，建议默认设置为1，最大限度地降低成本。 最小值：**1** 最大值：**128** 缺省值：**1**
presence_penalty	否	Number	介于-2.0和2.0之间的数字。 正值会根据它们是否出现在文本中来惩罚得到新的Token，从而增加模型谈论新主题的可能性。 最小值：**-2** 最大值：**2** 缺省值：**0**
stream	否	Boolean	布尔类型。 ● 设为true时，返回结果为流式； ● 设为false时，返回结果为JSON格式结构化数据。 缺省值：**false**

版权所有 © 华为云计算技术有限公司

参数	是否必选	参数类型	描述
temperature	否	Number	较高的数值会使输出更加随机，而较低的数值会使其更加集中和确定。 建议该参数和top_p只设置1个。 最小值：**0** 最大值：**2** 缺省值：**1**
top_p	否	Number	影响输出文本的多样性，取值越大，生成文本的多样性越强。 建议该参数和temperature只设置1个。 最小值：**0.0** 最大值：**1.0** 缺省值：**1**
tools	否	**FunctionCallTool** object	可供模型调用的工具。目前仅如下模型支持此功能： • glm-4 • glm-3-turbo • moonshot-v1-8k • moonshot-v1-32k • moonshot-v1-128k • spark-general-v3.5
tool_choice	否	String	用于控制模型是如何选择要调用的函数，仅当工具类型为function时补充。 默认为auto，且当前仅支持auto。
content_security_verify	否	**ContentSecurityVerify** object	控制是否开启内容审核。

表 4-3 ChatCompletionRequestMessage

参数	是否必选	参数类型	描述
content	是	String	消息具体内容。

版权所有 © 华为云计算技术有限公司

参数	是否必选	参数类型	描述
role	是	String	消息体对应的角色。 如果是系统则为system。 如果是用户则为user。 枚举值： • **system** • **user**
name	否	String	对话参与者的可选名称，提供给模型信息以区分相同角色的不同对话参与者。

表 4-4 FunctionCallTool

参数	是否必选	参数类型	描述
type	否	String	调用工具类型，目前仅支持function。
function	否	**function** object	仅当工具类型为function时补充。

表 4-5 function

参数	是否必选	参数类型	描述
name	否	String	函数名称，只能包含a-z，A-Z，0-9，下划线和中横线。 最大长度限制为64。
description	否	String	用于描述函数功能。 模型会根据这段描述决定函数调用方式。
parameters	否	Object	Json Schema对象，用于定义函数所接受的参数。

版权所有 © 华为云计算技术有限公司

表 4-6 ContentSecurityVerify

参数	是否必选	参数类型	描述
is_response_verify	否	Boolean	是否开启返回体内容审核（默认不开启）。 有文本内容，则对文本进行内容审核； 有图片内容，则会对图片进行内容审核。 缺省值：**false**

响应参数

状态码： 200

表 4-7 响应 Body 参数

参数	参数类型	描述
id	String	文本对话唯一标识符。
choices	Array of **choices** objects	返回体列表。 如果‘n’大于1，则结果为多个。
created	Integer	问答发生的时间（格式为时间戳）。
model	String	实际转发后调用的模型名称，与请求体中model可能不同。
object	String	固定值‘chat.completion’。
usage	**CompletionUsage** object	文本对话用量统计。

版权所有 © 华为云计算技术有限公司

表 4-8 choices

参数	参数类型	描述
finish_reason	String	返回结束的原因。 • stop：模型达到自然停止点或提供的停止序列； • length：达到请求中指定的最大令牌数； • content_filter：由于内容过滤器的标志而省略了内容。 枚举值： • **stop** • **length** • **content_filter**
index	Integer	返回多个choices时，每个choice对应的顺序。
message	**ChatCompletionResponseMessage** object	模型服务返回的具体消息体内容。

表 4-9 ChatCompletionResponseMessage

参数	参数类型	描述
content	String	返回消息体的内容。
role	String	返回消息体的角色。 枚举值： • **assistant**

表 4-10 CompletionUsage

参数	参数类型	描述
completion_tokens	Integer	回答包含的token数。
prompt_tokens	Integer	提问包含的token数。
total_tokens	Integer	提问+回答token总数。

状态码： 500

版权所有 © 华为云计算技术有限公司

表 4-11 响应 Body 参数

参数	参数类型	描述
error	**Error** object	异常详情。
error_code	String	平台异常错误码。
error_msg	String	异常信息。

表 4-12 Error

参数	参数类型	描述
code	String	异常码。 枚举值： • **invalid_request_error** • **invalid_api_key** • **internal_error** • **invalid_third_response** • **invalid_third_authentication** • **...**
message	String	异常信息。
param	String	异常参数，暂未使用。
type	String	异常类型，同code。

请求示例

```
{
  "model" : "publisher:baichuan:Baichuan2-Turbo",
  "messages" : [ {
    "role" : "system",
    "content" : "You are a helpful assistant."
  }, {
    "role" : "user",
    "content" : "你好!"
  } ]
}
```

响应示例

状态码： 200

OK

```
{
  "created" : 1718772336,
  "usage" : {
    "completion_tokens" : 23,
    "prompt_tokens" : 45,
    "total_tokens" : 68
```


版权所有 © 华为云计算技术有限公司

```
 },
 "model" : "Baichuan2-Turbo",
 "id" : "chatcmpl-xxx",
 "choices" : [ {
   "finish_reason" : "stop",
   "index" : 0,
   "message" : {
     "role" : "assistant",
     "content" : "你好，有什么我可以帮助你的吗？"
   },
   "logprobs" : null
 } ],
 "object" : "chat.completion"
}
```


状态码： 500

服务器内部错误或三方服务器内部错误。


```
{
 "error" : {
   "message" : "Internal server error, please try again later!",
   "type" : "internal_error",
   "param" : null,
   "code" : "internal_error"
 },
 "error_code" : "AIAE.31001001",
 "error_msg" : "Internal server error, please try again later!"
}
```


状态码

状态码	描述
200	OK
500	服务器内部错误或三方服务器内部错误。

错误码

请参见**错误码**。

4.1.2 文本向量化

功能介绍

将用户输入的文本转化成数字向量，多用于从向量化知识库中查询相似的文本。

URI

POST /v1/embeddings

版权所有 © 华为云计算技术有限公司

请求参数

表 4-13 请求 Header 参数

参数	是否必选	参数类型	描述
Authorization	是	String	AI原生应用引擎鉴权API Key。 1.以管理员身份登录AI原生应用引擎工作台，在左侧导航栏选择“配置中心 > 平台租户鉴权”。 2.在“平台租户鉴权”页面，选择“平台API Key”页签，单击“新增平台API Key”。 3.在“新增平台API Key”对话框中的输入框设置API Key名称，用以区分API Key。 4.在弹出的下载窗口中单击“立即下载”，将API Key下载到本地查看。

表 4-14 请求 Body 参数

参数	是否必选	参数类型	描述
input	是	Array of strings	输入支持2种格式： ● 纯文本（string），例如："你好"； ● 文本列表（array），例如：["你","好"]。 数组长度：**1 - 2048**

版权所有 © 华为云计算技术有限公司

参数	是否必选	参数类型	描述
model	是	String	模型服务调用唯一id字段。平台定义了4种模型服务: ● 平台预置模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“资产中心 > 大模型”，查看支持的模型服务。例如调用chatglm3-6b，model填写为platform:chatglm3-6b。 ● 平台接入模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“资产中心 > 大模型”，查看支持的模型服务。例如调用Baichuan2-Turbo模型服务，model填写为Baichuan2-Turbo即可。 ● 租户部署模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“Agent编排中心 > 我的模型服务 > 我部署的”，model填写为对应模型服务的模型服务调用ID。 ● 租户接入模型服务 – 登录AI原生应用引擎，在左侧导航栏选择“Agent编排中心 > 我的模型服务 > 我接入的”，model填写为对应模型服务的模型服务调用ID。 枚举值: ● **publisher:baichuan:Baichuan-Text-Embedding** ● **publisher:zhipu:embedding-2** ● **platform:bge-large-zh-v1.5**

响应参数

状态码： 200

版权所有 © 华为云计算技术有限公司

表 4-15 响应 Body 参数

参数	参数类型	描述
data	Array of **Embedding** objects	向量化结果。
model	String	实际转发后调用的模型名称，与请求体中model可能不同。
object	String	固定值‘list’。
usage	**usage** object	每次请求的用量统计。

表 4-16 Embedding

参数	参数类型	描述
index	Integer	向量在向量列表中的排序。
embedding	Array of numbers	向量数组（Float类型）。
object	String	固定值‘embedding’。

表 4-17 usage

参数	参数类型	描述
prompt_tokens	Integer	提问包含的token数。
total_tokens	Integer	提问包含的token数。

状态码： 500

表 4-18 响应 Body 参数

参数	参数类型	描述
error	**Error** object	异常详情。
error_code	String	平台异常错误码。
error_msg	String	异常信息。

版权所有 © 华为云计算技术有限公司

表 4-19 Error

参数	参数类型	描述
code	String	异常码。 枚举值： • **invalid_request_error** • **invalid_api_key** • **internal_error** • **invalid_third_response** • **invalid_third_authentication** • **...**
message	String	异常信息。
param	String	异常参数，暂未使用。
type	String	异常类型，同code。

请求示例

```
{
  "model" : "publisher:zhipu:embedding-2",
  "input" : "你好啊"
}
```

响应示例

状态码： 200

OK

```
{
    "data": [
        {
            "index": 0,
            "embedding": [
                0.02513289265334606,
                -0.017512470483779907,
                -0.029955564066767693,
                ...
            ],
    "object": "embedding"
            }
    ],
    "usage": {
        "prompt_tokens": 5,
        "total_tokens": 5
    },
    "model": "embedding-2",
    "object": "list"
}
```

状态码： 500

服务器内部错误或三方服务器内部错误。

```
{
  "error" : {
```


版权所有 © 华为云计算技术有限公司

```
  "message" : "Internal server error, please try again later!",
  "type" : "internal_error",
  "param" : null,
  "code" : "internal_error"
 },
 "error_code" : "AIAE.31001001",
 "error_msg" : "Internal server error, please try again later!"
}
```

状态码

状态码	描述
200	OK
500	服务器内部错误或三方服务器内部错误。

错误码

请参见**错误码**。

4.2 应用中心

4.2.1 调用 Agent

功能介绍

调用用户发布的Agent。

URI

POST /v1/routes/open/{id}/execute

表 4-20 路径参数

参数	是否必选	参数类型	描述
id	是	String	Agent在数据库中保存的UUID。

请求参数

表 4-21 请求 Header 参数

参数	是否必选	参数类型	描述
Authorization	是	String	鉴权信息，填写租户管理员已创建的API Key，前缀加Bearer与空格。

版权所有 © 华为云计算技术有限公司

表 4-22 请求 Body 参数

参数	是否必选	参数类型	描述
query	是	String	输入问题。
memory	否	Array of **memory** objects	用于传递在本次请求时，大模型提前记住的部分。

表 4-23 memory

参数	是否必选	参数类型	描述
role	是	String	角色，一般为user或assistant。
content	是	String	内容。

响应参数

状态码： 200

表 4-24 响应 Body 参数

参数	参数类型	描述
request_id	String	唯一请求DI
agent_id	String	待补充
user_id	String	AI引擎用户的唯一身份标识，orgid的不可逆加密值
conversation_id	String	会话ID
type	String	返回内容的类型：有hint、workflow、tool、knowledge、message5种类型
data	Object	不同响应类型的响应体中包含不同的参数，见示例。

状态码： 400

版权所有 © 华为云计算技术有限公司

表 4-25 响应 Body 参数

参数	参数类型	描述
error_code	String	错误码。
error_msg	String	错误描述。

状态码： 500

表 4-26 响应 Body 参数

参数	参数类型	描述
error_code	String	错误码。
error_msg	String	错误描述。

请求示例

```
{
  "query" : "查询北京天气",
  "memory" : [ {
    "role" : "user",
    "content" : "你是谁"
  }, {
    "role" : "assistant",
    "content" : "我是盘古大模型"
  }, {
    "role" : "user",
    "content" : "南京天气"
  }, {
    "role" : "assistant",
    "tool_calls" : [ {
      "id" : "efd6ff92-422c-4ba4-b531-ac1991af7c1a",
      "type" : "function",
      "function" : {
        "name" : "查询当前天气 查询当前天气",
        "arguments" : "{\"city\":\"320100\",\"extensions\":\"all\"}"
      }
    } ]
  }, {
    "role" : "tool",
    "tool_call_id" : "efd6ff92-422c-4ba4-b531-ac1991af7c1a",
    "content" : "{\"data\":{\"status\":\"1\",\"count\":\"1\",\"info\":\"OK\",\"infocode\":\"10000\",\"forecasts\":
[{\"city\":\"南京市\",\"adcode\":\"320100\",\"province\":\"江苏\",\"reporttime\":\"2024-08-20
16:32:01\",\"casts\":[{\"date\":\"2024-08-20\",\"week\":\"2\",\"dayweather\":\"中雨\",\"nightweather\":\"中雨
\",\"daytemp\":\"32\",\"nighttemp\":\"26\",\"daywind\":\"西北\",\"nightwind\":\"西北\",\"daypower
\":\"1-3\",\"nightpower\":\"1-3\",\"daytemp_float\":\"32.0\",\"nighttemp_float\":\"26.0\"}]}]}}"
  } ]
}
```

响应示例

状态码： 200

请求被服务所理解，正常调用。data的type表示响应的类型，包括knowledge知识库，tool工具，workflow工作流，message大模型，hint用于提示接下来使用knowledge，tool或workflow进行响应。下面的示例代表4类返回，所有实际返回类型

版权所有 © 华为云计算技术有限公司

均为data，示例中的数字仅为区分。第1个返回type为hint，tool_type为workflow，提示接下来要调用儿科问答这个工作流。第2个返回type为workflow，代表调用workflow的返回，返回内容在response下的data下的responseBody中。后续非message的返回同理，先返回hint提示调用类型，再返回真正的响应。最后的message类型是大模型的响应。


```
{
  "hint_data" : {
    "id" : "b6dbf1a6-f374-4d44-96fb-45726f7fa7f0",
    "name" : "儿科问答",
    "tool_type" : "workflow/Knowledge/tool"
  },
  "workflow_data" : {
    "id" : "b6dbf1a6-f374-4d44-96fb-45726f7fa7f0",
    "name" : "儿科问答",
    "status" : "SUCCESS",
    "request" : {
      "query" : "婴儿肥胖怎么办"
    },
    "response" : {
      "data" : {
        "responseBody" : "{\"result\":\"问题分析:主要控制儿童饮食，合理饮食不喝酒，不吃油炸食物意见和建议:建议孩子们多锻炼一点，每天至少锻炼一到两个小时，而且他们必须坚持锻炼。他们也应该少吃油和脂肪，多吃水果和蔬菜。我认为我们应该在一段时间后恢复正常。就食疗而言，父母必须参与其中，并被要求掌握一些相关知识，如不允许孩子吃得太多或太多，不给予高糖、高脂肪、高热量的饮食。治疗节食中的儿童并让他们挨饿也很困难。因此，在进行饮食控制之前，有必要耐心而详细地告诉儿童肥胖的危害、\"}",
        "responseHeaders" : {
          "Server" : "api-gateway",
          "X-Request-Id" : "6701c75b8f23102a659e63a3cc5a20d6",
          "X-Content-Type-Options" : "nosniff",
          "Connection" : "keep-alive",
          "X-Download-Options" : "noopen",
          "Date" : "Tue, 20 Aug 2024 08:37:27 GMT",
          "Referrer-Policy" : "no-referrer",
          "X-Frame-Options" : "SAMEORIGIN",
          "Strict-Transport-Security" : "max-age=31536000; includeSubdomains;",
          "lubanops-nenv-id" : "28164",
          "Content-Length" : "660",
          "X-XSS-Protection" : "1; mode=block;",
          "Content-Type" : "application/json"
        },
        "statusCode" : 200
      }
    }
  },
  "tool_data" : {
    "id" : "0333eb58-0914-4842-ad97-b0a42fe22dc9",
    "name" : "航班信息 航班信息",
    "status" : "SUCCESS",
    "request" : {
      "city" : "南京",
      "endcity" : "大理",
      "date" : "2024-08-24"
    },
    "response" : {
      "data" : {
        "status" : 0,
        "msg" : "ok",
        "result" : {
          "city" : "NKG",
          "endcity" : "DLU",
          "date" : "2024-08-24",
          "list" : [ {
            "flightno" : "ZH2010",
            "airline" : "深圳航空",
            "realflightno" : "TV6026",
            "departportcode" : "NKG",
            "departport" : "禄口国际机场",
```


版权所有 © 华为云计算技术有限公司

```
          "arrivalportcode" : "DLU",
          "arrivalport" : "大理荒草坝机场",
          "departterminal" : "T1",
          "arrivalterminal" : "",
          "departdate" : "2024-08-24",
          "arrivaldate" : "2024-08-24",
          "departtime" : "16:35",
          "arrivaltime" : "19:40",
          "departdateadd" : 0,
          "arrivaldateadd" : 0,
          "craft" : "19N",
          "stopnum" : "0",
          "costtime" : "03:05",
          "punctualrate" : "95",
          "pricelist" : [ ],
          "minprice" : "0",
          "airporttax" : "50",
          "fueltax" : "50",
          "food" : "1,",
          "isasr" : "1,",
          "iseticket" : "1,",
          "iscodeshare" : 1
        } ]
      }
    }
  }
 },
 "knowledge_data" : {
   "id" : "00e7ebb2-c52d-46ec-b7e9-62c53b1f6f47",
   "name" : "华西医院肠息肉",
   "status" : "SUCCESS",
   "request" : {
     "query" : "肠息肉怎么办"
   },
   "response" : "结直肠息肉应该怎么办？\n结直肠息肉是什么\n结肠息肉需不需要切除"
 },
 "message_data" : {
   "id" : "20240815205441517403 3b6a6544a1",
   "content" : "查询",
   "url" : null,
   "raw" : {
     "role" : "assistant",
     "content" : "查询"
   }
 }
}
```


状态码： 400

缺少请求体。


```
{
  "error_code" : "AIAE.00001400",
  "error_msg" : "Request body is missing"
}
```


状态码： 500

服务器内部错误或三方服务器内部错误。


```
{
  "error_code" : "AIAE.00001500",
  "error_msg" : "Internal Server Error."
}
```


 版权所有 © 华为云计算技术有限公司

状态码

状态码	描述
200	请求被服务所理解，正常调用。data的type表示响应的类型，包括knowledge知识库，tool工具，workflow工作流，message大模型，hint用于提示接下来使用knowledge，tool或workflow进行响应。下面的示例代表4类返回，所有实际返回类型均为data，示例中的数字仅为区分。第1个返回type为hint，tool_type为workflow，提示接下来要调用儿科问答这个工作流。第2个返回type为workflow，代表调用workflow的返回，返回内容在response下的data下的responseBody中。后续非message的返回同理，先返回hint提示调用类型，再返回真正的响应。最后的message类型是大模型的响应。
400	缺少请求体。
500	服务器内部错误或三方服务器内部错误。

错误码

请参见**错误码**。

4.2.2 调用技能

功能介绍

调用用户配置的技能。

URI

POST /v1/workflow-adapter-open/skills/{skill_id}

表 4-27 路径参数

参数	是否必选	参数类型	描述
skill_id	是	String	技能id。

请求参数

表 4-28 请求 Header 参数

参数	是否必选	参数类型	描述
authorization	是	String	鉴权信息，填写WiseAgent注册的api key，本接口只需要此鉴权信息，不需要使用公共请求头中的鉴权方式。

版权所有 © 华为云计算技术有限公司

表 4-29 请求 Body 参数

参数	是否必选	参数类型	描述
body	否	Object	调用技能请求体，与技能配置相关。结构与技能的请求体配置相同，并且所有请求头中的入参与请求参数均添加至请求体中，由wiseAgent自动完成分配。如果为GET请求则为非必填，如果为POST请求则为必填

响应参数

状态码： 200

表 4-30 响应 Body 参数

参数	参数类型	描述
data	Map<String,Object>	响应体。
error_code	String	错误码。
error_msg	String	错误信息。

状态码： 500

表 4-31 响应 Body 参数

参数	参数类型	描述
error_code	String	错误码。
error_msg	String	错误信息。

请求示例

```
{
  "object_example" : {
    "string_example" : "abc",
    "boolean_example" : "true",
    "integer_example" : 123
  }
}
```

响应示例

状态码： 200

请求被服务所理解，正常调用。

版权所有 © 华为云计算技术有限公司

```
{
  "data" : {
    "responseBody" : "something in response body",
    "responseHeaders" : {
      "Server" : "api-gateway",
      "X-Request-Id" : "787b7740f42e75b007ac3bfb599fcef4",
      "X-Content-Type-Options" : "nosniff",
      "Connection" : "keep-alive",
      "lubanops-nspan-id" : "1",
      "X-Download-Options" : "noopen",
      "Date" : "Tue, 23 Jul 2024 11:38:29 GMT",
      "lubanops-ntrace-id" : "2748112-1721734708992-1130609",
      "Referrer-Policy" : "no-referrer",
      "X-Frame-Options" : "SAMEORIGIN",
      "Strict-Transport-Security" : "max-age=31536000; includeSubdomains;",
      "lubanops-nenv-id" : "28164",
      "Content-Length" : "0",
      "X-XSS-Protection" : "1; mode=block;",
      "Content-Type" : "application/json"
    },
    "statusCode" : 200
  }
}
```


状态码： 500

服务器内部错误或三方服务器内部错误。


```
{
  "error_code" : "AIAE.22009001",
  "error_msg" : "Internal Server Error."
}
```


状态码

状态码	描述
200	请求被服务所理解，正常调用。
500	服务器内部错误或三方服务器内部错误。

错误码

请参见**错误码**。

4.2.3 调用流

功能介绍

调用用户配置的工作流。

URI

POST /v1/workflow-adapter-open/flows/{flow_id}

版权所有 © 华为云计算技术有限公司

表 4-32 路径参数

参数	是否必选	参数类型	描述
flow_id	是	String	流id。

表 4-33 Query 参数

参数	是否必选	参数类型	描述
query_example	否	String	按照用户配置的工作流请求参数配置，如果用户配置的工作流输入存在查询参数，则应该配置对应的参数并在调用时输入特定的值。

请求参数

表 4-34 请求 Header 参数

参数	是否必选	参数类型	描述
authorization	是	String	鉴权信息，填写WiseAgent注册的api key，本接口只需要此鉴权信息，不需要使用公共请求头中的鉴权方式。
header_example	否	String	按照用户配置的工作流请求头参数配置，如果用户配置的工作流输入存在请求头输入，则应该配置对应的参数并在调用时输入特定的值。

表 4-35 请求 Body 参数

参数	是否必选	参数类型	描述
body	否	Object	调用流请求体，与工作流起始节点配置相关，按照工作流的配置格式填写即可。

响应参数

状态码： 200

版权所有 © 华为云计算技术有限公司

表 4-36 响应 Body 参数

参数	参数类型	描述
data	Map<String,Object>	响应体。
error_code	String	错误码。
error_msg	String	错误信息。

状态码： 500

表 4-37 响应 Body 参数

参数	参数类型	描述
error_code	String	错误码。
error_msg	String	错误信息。

请求示例

```
{
  "object_example" : {
    "string_example" : "abc",
    "boolean_example" : "true",
    "integer_example" : 123
  }
}
```

响应示例

状态码： 200

请求被服务所理解，正常调用。

```
{
  "data" : {
    "responseBody" : "something in response body",
    "responseHeaders" : {
      "Server" : "api-gateway",
      "X-Request-Id" : "787b7740f42e75b007ac3bfb599fcef4",
      "X-Content-Type-Options" : "nosniff",
      "Connection" : "keep-alive",
      "lubanops-nspan-id" : "1",
      "X-Download-Options" : "noopen",
      "Date" : "Tue, 23 Jul 2024 11:38:29 GMT",
      "lubanops-ntrace-id" : "2748112-1721734708992-1130609",
      "Referrer-Policy" : "no-referrer",
      "X-Frame-Options" : "SAMEORIGIN",
      "Strict-Transport-Security" : "max-age=31536000; includeSubdomains;",
      "lubanops-nenv-id" : "28164",
      "Content-Length" : "0",
      "X-XSS-Protection" : "1; mode=block;",
      "Content-Type" : "application/json"
    },
    "statusCode" : 200
```


版权所有 © 华为云计算技术有限公司

```
  }
}
```

状态码： 500

服务器内部错误或三方服务器内部错误。

```
{
  "error_code" : "AIAE.22009001",
  "error_msg" : "Internal Server Error."
}
```

状态码

状态码	描述
200	请求被服务所理解，正常调用。
500	服务器内部错误或三方服务器内部错误。

错误码

请参见**错误码**。

4.3 知识中心

4.3.1 检索知识库数据

功能介绍

检索知识库数据，根据用户提供的检索信息，返回命中的信息数据。

URI

POST /v1/knowledge-bases/{knowledge_base_id}/embed-datas

表 4-38 路径参数

参数	是否必选	参数类型	描述
knowledge_base_id	是	String	知识库id。

版权所有 © 华为云计算技术有限公司

请求参数

表 4-39 请求 Body 参数

参数	是否必选	参数类型	描述
keyword	否	String	搜索关键字。
similarity_min	否	Float	相似度最小值。
limit	是	Integer	显示的条目数量。 最小值：**1** 最大值：**100**
filter	否	**SearchSqlFilter** object	过滤条件。
order_by	否	**SqlOrder** object	排序规则。

表 4-40 SearchSqlFilter

参数	是否必选	参数类型	描述
group_type	否	Object	枚举值： • **AND** • **OR**
expressions	否	Array of **Expression** objects	过滤条件。

表 4-41 Expression

参数	是否必选	参数类型	描述
field	否	String	过滤字段。
field_type	否	Object	字段类型。 枚举值： • **INT** • **FLOAT** • **BOOLEAN** • **STRING**

版权所有 © 华为云计算技术有限公司

参数	是否必选	参数类型	描述
operator	否	Object	操作符。 枚举值： • **EQUAL** • **NOT_EQUAL** • **GREAT_THAN** • **GREAT_EQUAL** • **LESS_THAN** • **LESS_EQUAL** • **IN** • **NOTIN**
values	否	Array of strings	过滤值。

表 4-42 SqlOrder

参数	是否必选	参数类型	描述
order_items	否	Array of **OrderItem** objects	排序规则。

表 4-43 OrderItem

参数	是否必选	参数类型	描述
field	否	String	排序字段。
field_type	否	Object	字段类型。 枚举值： • **INT** • **FLOAT** • **BOOLEAN** • **STRING**
order_type	否	Object	排序类型。 枚举值： • **ASC** • **DESC**

版权所有 © 华为云计算技术有限公司

响应参数

状态码： 200

表 4-44 响应 Body 参数

参数	参数类型	描述
data	Array of **data** objects	检索知识库数据具体内容。

表 4-45 data

参数	参数类型	描述
id	String	分片id。
document	String	分片数据。
similarity	Float	相似度范围, 从0到1数值越大相似度越高。
metadata	**metadata** object	元数据。
download_address	String	临时下载地址，当知识库数据类型为图片、图片-摘要、视频-摘要时有临时下载地址。

表 4-46 metadata

参数	参数类型	描述
order	Integer	序号。
path	String	文件路径。

状态码： 500

表 4-47 响应 Body 参数

参数	参数类型	描述
error_code	String	异常错误码
error_msg	String	异常错误信息

请求示例

```
{
  "keyword" : "户外",
  "similarity_min" : "0.78",
```


版权所有 © 华为云计算技术有限公司

```
  "limit" : 10,
  "filter" : null,
  "order_by" : null
}
```

响应示例

状态码： 200

OK

```
{
  "data" : [ {
    "id" : "812857ef-e298-4b8e-8bd1-24ba9fd5e95c",
    "document" : "户外运动热度大大带动各相关产业发展",
    "similarity" : 0.79593855,
    "metadata" : {
      "order" : 0,
      "file_name" : "户外运动热度大大带动各相关产业发展.docx",
      "file_id" : "户外运动热度大大带动各相关产业发展",
      "path" : "户外运动热度大大带动各相关产业发展.docx"
    },
    "download_address" : null
  } ]
}
```

状态码： 500

服务器内部错误或三方服务器内部错误

```
{
  "error_code" : "AIAE.00001500",
  "error_msg" : "系统内部错误。"
}
```

状态码

状态码	描述
200	OK
500	服务器内部错误或三方服务器内部错误

错误码

请参见**错误码**。

版权所有 © 华为云计算技术有限公司

5 附录

5.1 状态码

状态码如表5-1所示

表 5-1 状态码

状态码	编码	错误码说明
100	Continue	继续请求。 这个临时响应用来通知客户端，它的部分请求已经被服务器接收，且仍未被拒绝。
101	Switching Protocols	切换协议。只能切换到更高级的协议。 例如，切换到HTTP的新版本协议。
201	Created	创建类的请求完全成功。
202	Accepted	已经接受请求，但未处理完成。
203	Non-Authoritative Information	非授权信息，请求成功。
204	NoContent	请求完全成功，同时HTTP响应不包含响应体。 在响应OPTIONS方法的HTTP请求时返回此状态码。
205	Reset Content	重置内容，服务器处理成功。
206	Partial Content	服务器成功处理了部分GET请求。
300	Multiple Choices	多种选择。请求的资源可包括多个位置，相应可返回一个资源特征与地址的列表用于用户终端（例如：浏览器）选择。
301	Moved Permanently	永久移动，请求的资源已被永久的移动到新的URI，返回信息会包括新的URI。

版权所有 © 华为云计算技术有限公司

状态码	编码	错误码说明
302	Found	资源被临时移动。
303	See Other	查看其它地址。 使用GET和POST请求查看。
304	Not Modified	所请求的资源未修改，服务器返回此状态码时，不会返回任何资源。
305	Use Proxy	所请求的资源必须通过代理访问。
306	Unused	已经被废弃的HTTP状态码。
400	BadRequest	非法请求。 建议直接修改该请求，不要重试该请求。
401	Unauthorized	在客户端提供认证信息后，返回该状态码，表明服务端指出客户端所提供的认证信息不正确或非法。
402	Payment Required	保留请求。
403	Forbidden	请求被拒绝访问。 返回该状态码，表明请求能够到达服务端，且服务端能够理解用户请求，但是拒绝做更多的事情，因为该请求被设置为拒绝访问，建议直接修改该请求，不要重试该请求。
404	NotFound	所请求的资源不存在。 建议直接修改该请求，不要重试该请求。
405	MethodNotAllowed	请求中带有该资源不支持的方法。 建议直接修改该请求，不要重试该请求。
406	Not Acceptable	服务器无法根据客户端请求的内容特性完成请求。
407	Proxy Authentication Required	请求要求代理的身份认证，与401类似，但请求者应当使用代理进行授权。
408	Request Time-out	服务器等候请求时发生超时。 客户端可以随时再次提交该请求而无需进行任何更改。
409	Conflict	服务器在完成请求时发生冲突。 返回该状态码，表明客户端尝试创建的资源已经存在，或者由于冲突请求的更新操作不能被完成。
410	Gone	客户端请求的资源已经不存在。 返回该状态码，表明请求的资源已被永久删除。
411	Length Required	服务器无法处理客户端发送的不带Content-Length的请求信息。

版权所有 © 华为云计算技术有限公司

状态码	编码	错误码说明
412	Precondition Failed	未满足前提条件，服务器未满足请求者在请求中设置的其中一个前提条件。
413	Request Entity Too Large	由于请求的实体过大，服务器无法处理，因此拒绝请求。为防止客户端的连续请求，服务器可能会关闭连接。如果只是服务器暂时无法处理，则会包含一个Retry-After的响应信息。
414	Request-URI Too Large	请求的URI过长（URI通常为网址），服务器无法处理。
415	Unsupported Media Type	服务器无法处理请求附带的媒体格式。
416	Requested range not satisfiable	客户端请求的范围无效。
417	Expectation Failed	服务器无法满足Expect的请求头信息。
422	UnprocessableEntity	请求格式正确，但是由于含有语义错误，无法响应。
429	TooManyRequests	表明请求超出了客户端访问频率的限制或者服务端接收到多于它能处理的请求。建议客户端读取相应的Retry-After首部，然后等待该首部指出的时间后再重试。
500	InternalServerError	表明服务端能被请求访问到，但是不能理解用户的请求。
501	Not Implemented	服务器不支持请求的功能，无法完成请求。
502	Bad Gateway	充当网关或代理的服务器，从远端服务器接收到了一个无效的请求。
503	ServiceUnavailable	被请求的服务无效。 建议直接修改该请求，不要重试该请求。
504	ServerTimeout	请求在给定的时间内无法完成。客户端仅在为请求指定超时（Timeout）参数时会得到该响应。
505	HTTP Version not supported	服务器不支持请求的HTTP协议的版本，无法完成处理。

5.2 错误码

当您调用API时，如果遇到“APIGW”开头的错误码，请参见**API网关错误码**进行处理。

版权所有 © 华为云计算技术有限公司

状态码	错误码	错误信息	描述	处理措施
200	AIAE.22001001	API调用异常	API调用异常	调用接口url、请求方式错误或出现访问其他用户资源的越权问题，请检查后重试
200	AIAE.22001002	IAM认证异常	IAM认证异常	后端服务错误，请联系技术支持
200	AIAE.22001003	认证失败:{reason}	认证失败:{reason}	请参考返回的error message，或联系技术支持
200	AIAE.22001004	参数{parameterName}异常	参数{parameterName}异常	输入的参数有误，请参考返回的error message进行修改后重试
200	AIAE.22001005	{type}类型远程调用失败，错误信息为{error_msg}	{type}类型远程调用失败，错误信息为{error_msg}	请参考返回的error message处理后重试
200	AIAE.22001006	文件上传失败:{reason}	文件上传失败:{reason}	文件上传失败，请参考返回的error message处理后重试
200	AIAE.22001007	技能未设置鉴权信息	技能未设置鉴权信息	在wiseAgent页面为技能设置鉴权信息。
200	AIAE.22001008	用户无权限访问当前资源	用户无权限访问当前资源	请更换用户后访问
200	AIAE.22001009	开启的流无法被删除	开启的流无法被删除	关闭流后重试
200	AIAE.22009001	系统内部错误，请联系管理员	系统内部错误，请联系管理员	系统内部错误，请联系技术支持
400	AIAE.31001106	AK/SK signature verify failed, please check and try again later!	很抱歉，AK/SK签名验证失败!	很抱歉，AK/SK签名验证失败!
400	AIAE.31001601	Sensitive request error, please try again later!	很抱歉，请求内容中包含敏感信息，请重试!	很抱歉，请求内容中包含敏感信息，请重试!

版权所有 © 华为云计算技术有限公司

状态码	错误码	错误信息	描述	处理措施
400	AIAE.31001602	Sensitive response error, please try again later!	很抱歉，返回内容中包含敏感信息，请重试！	很抱歉，返回内容中包含敏感信息，请重试！
400	AIAE.31001603	Sensitive content error, please try again later!	很抱歉，请求或返回内容中包含敏感信息，请重试！	很抱歉，请求或返回内容中包含敏感信息，请重试！
400	AIAE.31001701	Bad request parameter error, please check and try again later!	很抱歉，请求参数异常，请检查后重试！	很抱歉，请求参数异常，请检查后重试！
400	AIAE.40001003	Authentication failed	X-Auth-Token鉴权失败	很抱歉，X-Auth-Token 鉴权失败
400	AIAE.40002605	knowledgeBase status is not ENABLE	很抱歉，知识库未启用，没有权限查询	很抱歉，知识库未启用，没有权限查询
401	AIAE.31001101	User not login, please check and try again later!	很抱歉，用户未登录，请登录后重试！	很抱歉，用户未登录，请登录后重试！
401	AIAE.31001102	AK/SK verify failed, please check and try again later!	很抱歉，AK/SK校验失败！	很抱歉，AK/SK校验失败！
401	AIAE.31001103	Authentication verify failed, please check and try again later!	很抱歉，鉴权失败！	很抱歉，鉴权失败！
401	AIAE.31001104	API Key verify failed, please check and try again later!	很抱歉，API Key校验失败！	很抱歉，API Key校验失败！
401	AIAE.31001201	Tenant id is empty, please check and try again later!	很抱歉，空的租户id，请检查后重试！	很抱歉，空的租户id，请检查后重试！

版权所有 © 华为云计算技术有限公司

状态码	错误码	错误信息	描述	处理措施
401	AIAE.31005001	The third model service authentication is abnormal, please check and try again later!	很抱歉，三方模型服务鉴权异常，请检查您的鉴权信息！	很抱歉，三方模型服务鉴权异常，请检查您的鉴权信息！
401	AIAE.31005002	Invalid third api key, please check and try again later!	很抱歉，三方模型服务鉴权API Key异常，请检查您的鉴权信息！	很抱歉，三方模型服务鉴权API Key异常，请检查您的鉴权信息！
401	AIAE.31005007	The third model service authentication is empty, please set and try again later!	很抱歉，三方模型服务鉴权未设置，请设置后重试！	很抱歉，三方模型服务鉴权未设置，请设置后重试！
402	AIAE.31005005	The third model service exceeded current quota error, please check and try again later!	很抱歉，账户异常，请检查您的账户余额！	很抱歉，账户异常，请检查您的账户余额！
403	AIAE.31001105	Role permission verify failed, please check and try again later!	很抱歉，当前用户不允许该操作！	很抱歉，当前用户不允许该操作！
403	AIAE.31001501	SKU not subscribed to model service, please try again after subscribed!	很抱歉，您未订阅当前模型服务的SKU，请联系管理员订阅！	很抱歉，您未订阅当前模型服务的SKU，请联系管理员订阅！
403	AIAE.31001502	SKU verify failed, please check and try again later!	很抱歉，SKU校验异常，请检查您的SKU！	很抱歉，SKU校验异常，请检查您的SKU！
403	AIAE.40001004	User does not have permission	当前用户没有权限	很抱歉，当前用户没有权限

版权所有 © 华为云计算技术有限公司

状态码	错误码	错误信息	描述	处理措施
404	AIAE.31001202	Model not published, please try again after model published!	很抱歉，模型服务未发布，请发布后重试！	很抱歉，模型服务未发布，请发布后重试！
404	AIAE.31001702	Model not exists, please check and try again later!	很抱歉，模型服务不存在，请检查您输入的模型服务名称！	很抱歉，模型服务不存在，请检查您输入的模型服务名称！
408	AIAE.31001003	Connection timeout, please try again later!	很抱歉，网络连接超时，请稍后重试！	很抱歉，网络连接超时，请稍后重试！
408	AIAE.31005006	The third model service connect timeout, please try again later!	很抱歉，三方服务连接超时，请稍后重试！	很抱歉，三方服务连接超时，请稍后重试！
429	AIAE.31001002	Request too frequent error, please try again later!	很抱歉，您的请求过于频繁，请稍后重试！	很抱歉，您的请求过于频繁，请稍后重试！
429	AIAE.31005003	The third model service rate limit exceeded, please try again later!	很抱歉，您的请求当前已达最大并发数，请稍后重试！	很抱歉，您的请求当前已达最大并发数，请稍后重试！
429	AIAE.31005004	The third model service overload error, please try again later!	很抱歉，服务当前超载，请稍后重试！	很抱歉，服务当前超载，请稍后重试！
500	AIAE.31001001	Internal server error, please try again later!	很抱歉，服务内部出现了问题，请稍后重试！	很抱歉，服务内部出现了问题，请稍后重试！

版权所有 © 华为云计算技术有限公司

状态码	错误码	错误信息	描述	处理措施
500	AIAE.31001004	Ai security governance service error, please try again later or disable ai security governance service!	很抱歉，内容审核服务出现了问题，请稍后重试！	很抱歉，内容审核服务出现了问题，请稍后重试！
500	AIAE.31005000	Invalid third response, please try again later!	很抱歉，调用三方模型服务异常，请稍后重试！	很抱歉，调用三方模型服务异常，请稍后重试！
400	UniModel.Request.0001	请求参数错误	模型服务的请求参数错误	检查模型的请求参数
500	UniDataEmbed.Internal.0001	请求失败	请求向量化服务失败	检查向量知识库配置
500	UniModel.Internal.0001	模型访问失败	无法访问选择的模型	检查模型是否已经正常部署
500	UniModel.Internal.0002	模型返回超时	模型服务返回超时	检查网络情况，或者减少模型返回内容
500	WS.00100001	AUTHENTICATION_ERROR	鉴权错误	检查访问权限
500	WS.00100002	SHA_ERROR	SHA算法错误	检查签名所用的算法
500	WS.00100003	SIGN_ERROR	请求签名错误	检查请求签名
500	WS.00100005	NO_ACCESS_ERROR	无访问权限错误	检查接口访问权限

版权所有 © 华为云计算技术有限公司