智能体平台

API 参考

文档版本 01

发布日期 2025-09-16

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 2025。 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。

华为云计算技术有限公司

地址： 贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址： https://www.huaweicloud.com/

目 录

1 使用前必读.....1
2 API 概览.....3
3 如何调用 API.....4
3.1 构造请求.....4
3.2 认证鉴权.....6
3.3 返回结果.....8
4 API.....10
4.1 工作流.....10
4.1.1 调用工作流应用.....10
4.2 智能体.....19
4.2.1 调用智能体应用.....19
4.2.2 上传文件.....28
5 应用示例.....33
5.1 调用工作流应用示例.....33
5.2 调用智能体应用示例.....34
6 附录.....37
6.1 状态码.....37
6.2 错误码.....39
6.3 获取项目 ID.....43
6.4 获取账号 ID.....44
6.5 获取工作区 ID.....44

 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Versatile。Versatile是一个一站式企业级智能体构建平台，包含应用管理、组件库、知识库、提示词开发、配置管理、模型接入调测等功能模块，覆盖体验设计、代码开发、应用运行、资产管理、数据处理、测试发布、运营监控、安全保障八大面，为企业级用户提供开箱即用的大模型应用开发工具链。依托强大的应用开发工具链，Versatile可支撑客户的个性化应用功能开发需求，智能扩展Agent边界，搭配兼具性能和安全的运行机制，降低开发门槛，使得应用规模化落地，助力各行业企业将大模型应用与实际业务融合，打造企业级专属应用。

您可以使用本文档提供API对Versatile进行相关操作，如调用应用、调用工作流等。支持的全部操作请参见**API概览**。

在调用Versatile API之前，请确保已经充分了解Versatile相关概念，详细信息请参见Versatile《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，Versatile目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。

基本概念

- 账号

  用户注册华为云时的账号，账号对其所拥有的资源及云服务具有完全的访问权限，可以重置用户密码、分配用户权限等。由于账号是付费主体，为了确保账号安全，建议您不要直接使用账号进行日常管理工作，而是创建用户并使用他们进行日常管理工作。

- 用户

  由账号在IAM中创建的用户，是云服务的使用人员，具有身份凭证（密码和访问密钥）。

  在**我的凭证**下，您可以查看账号ID和用户ID。通常在调用API的鉴权过程中，您需要用到账号、用户和密码等信息。

- 区域（Region）

  从地理位置和网络时延维度划分，同一个Region内共享弹性计算、块存储、对象存储、VPC网络、弹性公网IP、镜像等公共服务。Region分为通用Region和专属Region，通用Region指面向公共租户提供通用云服务的Region；专属Region指只承载同一类业务或只面向特定租户提供业务服务的专用Region。

版权所有 © 华为云计算技术有限公司

详情请参见**区域和可用区**。

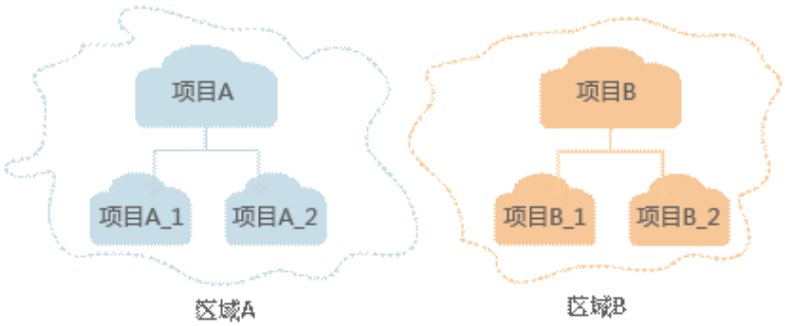
- 可用区（AZ，Availability Zone）

  一个AZ是一个或多个物理数据中心的集合，有独立的风火水电，AZ内逻辑上再将计算、网络、存储等资源划分成多个集群。一个Region中的多个AZ间通过高速光纤相连，以满足用户跨AZ构建高可用性系统的需求。

- 项目

  华为云的区域默认对应一个项目，这个项目由系统预置，用来隔离物理区域间的资源（计算资源、存储资源和网络资源），以默认项目为单位进行授权，用户可以访问您账号中该区域的所有资源。如果您希望进行更加精细的权限控制，可以在区域默认的项目中创建子项目，并在子项目中购买资源，然后以子项目为单位进行授权，使得用户仅能访问特定子项目中资源，使得资源的权限控制更加精确。

  图 1-1 项目隔离模型



  同样在**我的凭证**下，您可以查看项目ID。

- 企业项目

  企业项目是项目的升级版，针对企业不同项目间资源的分组和管理，是逻辑隔离。企业项目中可以包含多个区域的资源，且项目中的资源可以迁入迁出。

  关于企业项目ID的获取及企业项目特性的详细信息，请参见《**企业管理服务用户指南**》。

版权所有 © 华为云计算技术有限公司

2 API 概览

类型	说明
工作流	调用工作流应用接口。
智能体	调用智能体应用、上传文件接口。

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IAM服务的**管理员创建IAM用户**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。

- **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，Versatile目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。
- **resource-path**：资源路径，也即API访问路径。从具体API的URI模块获取，例如“管理员创建IAM用户”接口的resource-path为“/v3.0/OS-USER/users”。
- **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。

例如您需要创建IAM用户，由于IAM为全局服务，则使用任一区域的Endpoint（比如“华北-北京四”区域的Endpoint：“iam.cn-north-4.myhuaweicloud.com”），并在**管理员创建IAM用户**的URI部分找到resource-path（/v3.0/OS-USER/users），拼接起来如下所示。

```
https://iam.cn-north-4.myhuaweicloud.com/v3.0/OS-USER/users
```


版权所有 © 华为云计算技术有限公司

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），它告诉服务正在请求什么类型的操作。

- **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**管理员创建IAM用户**的URI部分，您可以看到其请求方法为“POST”，则其请求为：

```
POST https://iam.cn-north-4.myhuaweicloud.com/v3.0/OS-USER/users
```

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。

如下公共消息头需要添加到请求中。

- **Content-Type**：消息体的类型（格式），必选，默认取值为“application/json”，有其他取值时会在具体接口中专门说明。
- **X-Auth-Token**：用户Token，可选，当使用Token方式认证时，必须填充该字段。用户Token也就是调用**获取用户Token**接口的响应值，该接口是唯一不需要认证的接口。

对于**管理员创建IAM用户**接口，使用AK/SK方式认证时，添加消息头后的请求如下所示。

```
POST https://iam.cn-north-4.myhuaweicloud.com/v3.0/OS-USER/users
Content-Type: application/json
X-Sdk-Date: 20240416T095341Z
Authorization: SDK-HMAC-SHA256 Access=**********, SignedHeaders=content-type;host;x-sdk-date, Signature=**********
```

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，

版权所有 © 华为云计算技术有限公司

并在Content-type中声明字符编码方式，例如：Content-Type: application/json; charset=utf-8。

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**管理员创建IAM用户**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，其中加粗的字段需要根据实际值填写。

- **accountid**为IAM用户所属的账号ID。
- **username**为要创建的IAM用户名。
- **email**为IAM用户的邮箱。
- **************为IAM用户的登录密码。

```
POST https://iam.cn-north-4.myhuaweicloud.com/v3.0/OS-USER/users（中国站）
POST https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users（国际站）
Content-Type: application/json
X-Sdk-Date: 20240416T095341Z
Authorization: SDK-HMAC-SHA256 Access=**********, SignedHeaders=content-type;host;x-sdk-date, Signature=**********

{
    "user": {
        "domain_id": "accountid",
        "name": "username",
        "password": "**********",
        "email": "email",
        "description": "IAM User Description"
    }
}
```

到此为止，一个API请求所需要的内容已经准备完成，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

调用接口有如下两种认证方式，您可以选择其中一种进行认证鉴权。

- AK/SK认证：通过AK（Access Key ID）/SK（Secret Access Key）加密调用请求。
- Token认证：通过Token认证调用请求。

AK/SK 认证

说明

- AK/SK签名认证方式仅支持消息体大小12MB以内，12MB以上的请求请使用Token认证。
- AK/SK既可以使用永久访问密钥中的AK/SK，也可以使用临时访问密钥中的AK/SK，但使用临时访问密钥的AK/SK时需要额外携带“X-Security-Token”字段，字段值为临时访问密钥的security_token。

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。

版权所有 © 华为云计算技术有限公司

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，也可以使用专门的签名SDK对请求进行签名。详细的签名方法和SDK使用方法请参见**API签名指南**。

须知

签名SDK只提供签名功能，与服务提供的SDK不同，使用时请注意。

您也可以通过这个视频教程了解AK/SK认证的使用：**https://bbs.huaweicloud.com/videos/100697** 。

Token 认证

说明

- Token的有效期为24小时，需要使用一个Token鉴权时，可以先缓存起来，避免频繁调用。
- 使用Token前请确保Token离过期有足够的时间，防止调用API的过程中Token过期导致调用API失败。

Token在计算机系统中代表令牌（临时）的意思，拥有Token就代表拥有某种权限。Token认证就是在调用API的时候将Token加到请求消息头，从而通过身份认证，获得操作API的权限。

Token可通过调用**获取用户Token**接口获取，调用本服务API需要project级别的Token，即调用接口时，请求body中auth.scope的取值需要选择project，如下所示。

```
{
    "auth": {
        "identity": {
            "methods": [
                "password"
            ],
            "password": {
                "user": {
                    "name": "username",
                    "password": "********",
                    "domain": {
                        "name": "domainname"
                    }
                }
            }
        },
        "scope": {
            "project": {
                "name": "xxxxxxxx"
            }
        }
    }
}
```

获取Token后，再调用其他接口时，您需要在请求消息头中添加“X-Auth-Token”，其值即为获取到的Token。例如Token值为“ABCDEFG....”，则调用接口时将“X-Auth-Token: ABCDEFG....”加到请求消息头即可，如下所示。

```
POST https://iam.cn-north-4.myhuaweicloud.com/v3.0/OS-USER/users（中国站）
POST https://iam.ap-southeast-1.myhuaweicloud.com/v3.0/OS-USER/users（国际站）
Content-Type: application/json
X-Auth-Token: ABCDEFG....
```

您还可以通过这个视频教程了解如何使用Token认证：**https://bbs.huaweicloud.com/videos/101333** 。

版权所有 © 华为云计算技术有限公司

3.3 返回结果

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

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**管理员创建IAM用户**接口，如果调用后返回状态码为“201”，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，如“Content-type”。

对于**管理员创建IAM用户**接口，返回如**图3-2**所示的消息头。

图 3-2 管理员创建 IAM 用户响应消息头

```
"X-Frame-Options": "SAMEORIGIN",
"X-IAM-ETag-id": "2562365939-d8f6f12921974cb097338ac11fceac8a",
"Transfer-Encoding": "chunked",
"Strict-Transport-Security": "max-age=31536000; includeSubdomains;",
"Server": "api-gateway",
"X-Request-Id": "af2953f2bcc67a42325a69a19e6c32a2",
"X-Content-Type-Options": "nosniff",
"Connection": "keep-alive",
"X-Download-Options": "noopen",
"X-XSS-Protection": "1; mode=block;",
"X-IAM-Trace-Id": "token_[redacted]_null_af2953f2bcc67a42325a69a19e6c32a2",
"Date": "Tue, 21 May 2024 09:03:40 GMT",
"Content-Type": "application/json; charset=utf8"
```

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**管理员创建IAM用户**接口，返回如下消息体。为篇幅起见，这里只展示部分内容。

```
{
    "user": {
        "id": "c131886aec...",
        "name": "IAMUser",
        "description": "IAM User Description",
        "areacode": "",
        "phone": "",
        "email": "***@***.com",
        "status": null,
        "enabled": true,
        "pwd_status": false,
        "access_mode": "default",
        "is_domain_owner": false,
        "xuser_id": "",
        "xuser_type": "",
```


版权所有 © 华为云计算技术有限公司

```
        "password_expires_at": null,
        "create_time": "2024-05-21T09:03:41.000000",
        "domain_id": "d78cbac1..........",
        "xdomain_id": "30086000........",
        "xdomain_type": "",
        "default_project_id": null
    }
}
```

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_msg": "Request body is invalid.",
    "error_code": "IAM.0011"
}
```

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 工作流

4.1.1 调用工作流应用

功能介绍

该接口用于运行场景化应用，支持在指定的项目、工作流和对话上下文中执行工作流逻辑。接口支持流式响应模式，可以根据需要返回增量执行结果，适用于实时交互场景。

适用场景：

- 在项目中运行预定义的工作流。
- 支持调试模式和发布模式，适用于不同开发和生产环境。
- 支持流式响应，适用于需要实时反馈的场景（如聊天机器人、实时数据分析等）。

URI

POST /v1/{project_id}/workflows/{workflow_id}/conversations/{conversation_id}

版权所有 © 华为云计算技术有限公司

表 4-1 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	**参数解释**: 当前租户项目ID。 获取方法请参考**获取项目ID**。 **约束限制**: 不涉及。 **取值范围**: 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**: 不涉及。
workflow_id	是	String	**参数解释**: 工作流应用的ID。 获取方式: 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 应用管理 > 工作流应用”。 3. 在待复制ID的工作流应用卡片上，单击“▪▪▪ > 复制ID”。 **约束限制**: 不涉及。 **取值范围**: 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**: 不涉及。

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
conversation_id	是	String	**参数解释**： 会话ID，每个会话的唯一标识符，可将会话ID设置为任意值，使用标准UUID格式。 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 应用管理 > 工作流应用”。 3. 在待复制调用路径的工作流应用卡片上，单击“▪▪▪ > 调用路径”。 4. 在弹出的“调用路径”页面，单击“复制路径”即可获取调用路径。其中，conversation_id参数为会话ID，唯一标识每个会话的标识符，可将会话ID设置为任意值，使用标准UUID格式。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。

表 4-2 Query 参数

参数	是否必选	参数类型	描述
workspace_id	否	String	**参数解释**： 工作空间ID，用于标识特定的工作空间。 获取方法请参考**获取工作空间ID**。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。

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
version	否	String	**参数解释**: 发布版本号。 获取方式: 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 应用管理 > 工作流应用”。 3. 选择需要查找的工作流应用。 4. 在工作流界面右上角，单击“发布历史”，获取发布版本号。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。

请求参数

表 4-3 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	是	String	**参数解释**: 用户Token。通过调用IAM服务获取用户Token接口获取（响应消息头中X-Subject-Token的值）。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。

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
X-Invoke-Mode	否	String	**参数解释**： 该参数用于标识工作流应用运行的模式。 ● X-Invoke-Mode的值为debug时，工作流应用的运行模式为调试模式。调试模式会生成日志、详细的执行步骤，便于排查问题。 ● X-Invoke-Mode的值为published时，工作流应用的运行模式为发布模式。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。
stream	否	Boolean	**参数解释**： 是否开启流式调用。 ● 当stream为true时，服务器以流式方式逐步返回结果，适合需要实时反馈的场景。 ● 当stream为false时，服务器在处理完成后一次性返回结果，适合处理较小数据或不需要实时反馈的场景。 **约束限制**： 不涉及。 **取值范围**： ● true：开启。 ● false：不开启。 **默认取值**： 不涉及。

版权所有 © 华为云计算技术有限公司

表 4-4 请求 Body 参数

参数	是否必选	参数类型	描述
inputs	是	Map<String,Object>	**参数解释**: 用户提出的问题，作为运行工作流的输入，与工作流开始节点输入参数对应。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
plugin_configs	否	Array of **PluginConfig** objects	**参数解释**: 插件配置信息。 **约束限制**: 不涉及。 **取值范围**: 当工作流关联插件节点，并且插件是“用户级鉴权”时，需要配置对应的鉴权信息。其他情况该参数无需传值，plugin_configs传空数组。 **默认取值**: 不涉及。

版权所有 © 华为云计算技术有限公司

表 4-5 PluginConfig

参数	是否必选	参数类型	描述
plugin_id	否	String	**参数解释**： 插件ID。 获取方式： 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 组件库 > 我的插件”。 3. 在待复制ID的插件卡片上，单击“▪▪▪ > 复制ID”。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。
config	否	Map<String,String>	**参数解释**： 配置插件信息。当工作流关联插件节点，并且插件是“用户级鉴权”时，需要在此配置对应的鉴权信息。例如，针对如下插件，config可以配成：{"key2": "value"}。其他情况该参数无需传值，plugin_configs传空数组即可。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。

响应参数

状态码：200

版权所有 © 华为云计算技术有限公司

表 4-6 响应 Body 参数

参数	参数类型	描述
event	Map<String,Object>	**参数解释**: 工作流最终输出内容表示工作流运行。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
data	Array of **Message** objects	**参数解释**: 工作流助手回复内容。例如，提问器节点问题消息。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
status	Map<String,Object>	**参数解释**: 状态信息。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
start_time	Long	**参数解释**: 开始时间。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。

版权所有 © 华为云计算技术有限公司

参数	参数类型	描述
end_time	Long	**参数解释**： 结束时间。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。

表 4-7 Message

参数	参数类型	描述
role	String	**参数解释**： 会话角色。 **约束限制**： 不涉及。 **取值范围**： ● user：用户输入的消息，包括提示词和上下文信息。 ● assistant：模型生成的回复内容。 **默认取值**： 不涉及。
content	String	**参数解释**： 会话内容。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。

请求示例

```
{
  "method" : "POST",
  "url" : "https://api.example.com/v1/12345/workflows/67890/conversations/67890",
  "headers" : {
    "Content-Type" : "application/json",
    "X-Auth-Token" :
"MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQExDTALBglghkgBZQMEAgEwgguVBgkqhkiG...",
    "stream" : true
```


版权所有 © 华为云计算技术有限公司

```
  },
  "body" : {
    "inputs" : {
      "query" : "你好"
    },
    "plugin_configs" : [ {
      "plugin_id" : "xxxxxxxxx",
      "config" : {
        "key" : "value"
      }
    } ]
  }
}
```

响应示例

状态码：200

成功响应。

```
{
  "event" : "workflow_finished",
  "data" : {
    "status" : {
      "code" : 1,
      "desc" : "succeeded"
    },
    "outputs" : {
      "responseContent" : "《朝花夕拾》是关于某篇文章、某个人物或某个主题，欢迎继续提问。"
    },
    "start_time" : 1757729064202,
    "end_time" : 1757729120126
  }
}
```

状态码

状态码	描述
200	成功响应。

错误码

请参见**错误码**。

4.2 智能体

4.2.1 调用智能体应用

功能介绍

该接口用于运行知识型智能体应用，支持单智能体和多智能体，支持在指定的项目、智能体和对话上下文中执行智能体逻辑。接口支持流式响应模式，可以根据需要返回增量执行结果，适用于实时交互场景。

适用场景：

版权所有 © 华为云计算技术有限公司

- 在项目中运行预定义的知识型智能体应用。
- 支持调试模式和发布模式，适用于不同开发和生产环境。
- 支持流式响应，适用于需要实时反馈的场景（如聊天机器人、实时数据分析等）。

URI

POST /v1/{project_id}/agents/{agent_id}/conversations/{conversation_id}

表 4-8 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	**参数解释**： 当前租户项目ID。 获取方法请参考**获取项目ID**。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。
agent_id	是	String	**参数解释**： 智能体应用ID。 获取方式： 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 应用管理 > 单智能体应用”或选择“开发中心 > 应用管理 > 多智能体应用”。 3. 在待复制ID的智能体应用卡片上，单击“▪▪▪ > 复制ID”。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。

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
conversation_id	是	String	**参数解释**： 会话ID，每个会话的唯一标识符，可将会话ID设置为任意值，使用标准UUID格式。 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 应用管理 > 单智能体应用”或选择“开发中心 > 应用管理 > 多智能体应用”。 3. 在待复制调用路径的智能体应用卡片上，单击“▪▪▪ > 调用路径”。 4. 在弹出的“调用路径”页面，单击“复制路径”即可获取调用路径。其中，conversation_id参数为会话ID，唯一标识每个会话的标识符，可将会话ID设置为任意值，使用标准UUID格式。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。

表 4-9 Query 参数

参数	是否必选	参数类型	描述
workspace_id	是	String	**参数解释**： 工作空间ID，用于标识特定的工作空间。 获取方法请参考**获取工作空间ID**。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。

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
version	否	String	**参数解释**： 发布版本号。 获取方式： 1. 进入Versatile智能体平台。 2. 在左侧导航选择“开发中心 > 应用管理 > 单智能体应用”或选择“开发中心 > 应用管理 > 多智能体应用”。 3. 选择需要查找的智能体应用。 4. 在智能体界面右上角，单击“发布历史”，获取发布版本号。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。
type	否	String	**参数解释**： 该参数允许调用者指定智能体应用的执行类型，支持不同的执行方式和返回模式。 **约束限制**： 不涉及。 **取值范围**： ● 实时处理：设置type为AGENT，使用流式返回，适用于需要实时反馈的场景，如实时聊天。 ● 批量处理：设置type为CONTROLLER，使用非流式返回，适用于处理完成后一次性返回结果的场景，如数据处理。 **默认取值**： Constant.AppType.AGENT。

版权所有 © 华为云计算技术有限公司

请求参数

表 4-10 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	是	String	**参数解释**: 用户Token。通过调用IAM服务获取用户Token接口获取（响应消息头中X-Subject-Token的值）。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
X-Invoke-Mode	否	String	**参数解释**: 该参数用于标识工作流应用运行的模式。 ● X-Invoke-Mode的值为debug时，工作流应用的运行模式为调试模式。调试模式会生成日志、详细的执行步骤，便于排查问题。 ● X-Invoke-Mode的值为published时，工作流应用的运行模式为发布模式。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。

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
stream	否	Boolean	**参数解释**： 是否开启流式调用。当前智能体应用只支持流式调用。 ● 当stream为true时，服务器以流式方式逐步返回结果，适合需要实时反馈的场景。 ● 当stream为false时，服务器在处理完成后一次性返回结果，适合处理较小数据或不需要实时反馈的场景。 **约束限制**： 不涉及。 **取值范围**： ● true：开启。 ● false：不开启。 **默认取值**： 不涉及。

表 4-11 请求 Body 参数

参数	是否必选	参数类型	描述
query	是	String	**参数解释**： 用户问题，作为运行智能体的输入。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。

响应参数

状态码：200

版权所有 © 华为云计算技术有限公司

表 4-12 响应 Body 参数

参数	参数类型	描述
data	String	当请求参数“stream”值为“true”时，智能体的消息以流式形式返回。生成的内容以增量的方式逐步发送回来，每个data字段均包含一部分生成的内容，直到所有data返回，响应结束。
event	String	**参数解释**： 数据单元类型。 **约束限制**： 不涉及。 **取值范围**： ● start：开始节点，表示开始调用模型进行会话。 ● message：消息节点，表示模型返回的消息。 ● plugin_start：插件调用请求节点，表示调用插件的请求信息。 ● plugin_end：插件调用响应节点，表示调用插件的响应信息。 ● statistic_data：执行数据节点，包含本次调用的耗时信息。 ● summary_response：消息总结节点，包含本次调用的全量响应信息。 ● done：流式调用结束节点，表示流式响应结束。 **默认取值**： 不涉及。
content	Object	**参数解释**： 消息块内容。“event”参数类型不同，内容结构不同。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。

版权所有 © 华为云计算技术有限公司

参数	参数类型	描述
createdTime	Long	**参数解释**: 消息块返回的时间戳。例如，1733817348963。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
latency	**latency** object	**参数解释**: 耗时，包括以下三个元素: • plugin： 插件调用耗时。 • model： 模型调用耗时。 • overall： 总耗时。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。
plugin	**plugin** object	**参数解释**: 插件请求信息，包括以下两个元素: • name：插件名。 • arguments：插件入参名。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。

表 4-13 latency

参数	参数类型	描述
plugin	Long	插件调用耗时。
model	Long	模型调用耗时。

版权所有 © 华为云计算技术有限公司

参数	参数类型	描述
overall	Long	总耗时。

表 4-14 plugin

参数	参数类型	描述
name	String	插件名。
arguments	Object	插件入参名。

请求示例

```
{
  "method" : "POST",
  "url" : "https://{endpoint}/v1/{project_id}/agents/{agent_id}/conversations/{conversation_id}",
  "headers" : {
    "Content-Type" : "application/json",
    "X-Auth-Token" :
"MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQExDTALBglghkgBZQMEAgEwgguVBgkqhkiG...",
    "stream" : true
  },
  "body" : {
    "query" : "查询A12会议室在9:00到10:00的状态"
  }
}
```

响应示例

状态码：200

流式响应，返回模型生成内容的增量数据块。

```
data:{"event":"start","createdTime":1735558575017}

data:{"event":"message","content":"好的","createdTime":1735558576300}

data:{"event":"message","content":"，","createdTime":1735558576301}

data:{"event":"message","content":"我将","createdTime":1735558576301}

data:{"event":"message","content":"调用","createdTime":1735558576302}

data:{"event":"message","content":"query","createdTime":1735558576302}

data:{"event":"statistic_data","latency":{"overall":1.97},"createdTime":1735558576986}

data:{"event":"summary_response","content":"A12会议室在9:00到10:00的时间段内是空闲的。
","role":"assistant","createdTime":1735558576987}

data:{"event":"done","createdTime":1735558577011}
```

状态码

状态码	描述
200	流式响应，返回模型生成内容的增量数据块。

版权所有 © 华为云计算技术有限公司

错误码

请参见**错误码**。

4.2.2 上传文件

功能介绍

该接口用于智能体上传文件，支持多种图片、文档、表格等多种格式的文件上传。接口返回临时下载路径，可用于临时下载文件。

适用场景：在智能体应用中上传文件。

格式要求：

- 办公文档：DOC、DOCX、XLS、XLSX、PPT、PPTX、PDF、Numbers、CSV。
- 图像文件：JPG、JPEG、PNG、GIF、WEBP、HEIC、HEIF、BMP、PCD、TIFF。
- 音频文件：WAV、MP3、FLAC、M4A、AAC、OGG、WMA、MIDI。
- 文本文件：JS、CPP、PY、JAVA、C、TXT、CSS、JAVASCRIPT、HTML、JSON、MD。

URI

POST /v1/{project_id}/agent-runtime/upload-file

表 4-15 路径参数

参数	是否必选	参数类型	描述
project_id	是	String	**参数解释**： 当前租户项目ID。 获取方法请参考**获取项目ID**。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。

版权所有 © 华为云计算技术有限公司

表 4-16 Query 参数

参数	是否必选	参数类型	描述
workspace_id	是	String	**参数解释**： 工作空间ID，用于标识特定的工作空间。 获取方法请参考**获取工作空间ID**。 **约束限制**： 不涉及。 **取值范围**： 由英文，数字，“-”，“_”组成，不超过64位字符。 **默认取值**： 不涉及。
file	是	Object	**参数解释**： 上传的文件。 **约束限制**： 不涉及。 **取值范围**： 大小不超过60MB。 **默认取值**： 不涉及。
expires	否	Integer	**参数解释**： 访问授权过期时间（天）。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。
is_image	否	Boolean	**参数解释**： 是否是图片上传。 **约束限制**： 不涉及。 **取值范围**： • 是：文件为图片格式。 • 否：文件为非图片格式。 **默认取值**： 不涉及。

版权所有 © 华为云计算技术有限公司

请求参数

表 4-17 请求 Header 参数

参数	是否必选	参数类型	描述
X-Auth-Token	是	String	**参数解释**: 用户Token。通过调用IAM服务获取用户Token接口获取（响应消息头中X-Subject-Token的值）。 **约束限制**: 不涉及。 **取值范围**: 不涉及。 **默认取值**: 不涉及。

表 4-18 请求 Body 参数

参数	是否必选	参数类型	描述
file	是	String	**参数解释**: 用户上传的文档。 **约束限制**: 不涉及。 **取值范围**: 文件大小小于60MB。 **默认取值**: 不涉及。
is_image	否	Boolean	**参数解释**: 用户上传的文档是否是图片。 **约束限制**: 不涉及。 **取值范围**: • 是：文件为图片格式。 • 否：文件为非图片格式。 **默认取值**: 不涉及。

版权所有 © 华为云计算技术有限公司

响应参数

状态码：200

表 4-19 响应 Body 参数

参数	参数类型	描述
url	String	**参数解释**： 临时有效，用于访问存储在华为云 OBS 上的文件的下载地址。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。
headers	Object	**参数解释**： 请求访问的域名，是华为云OBS签名验证的关键信息。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。
file_name	String	**参数解释**： 文件名。 **约束限制**： 不涉及。 **取值范围**： 不涉及。 **默认取值**： 不涉及。

请求示例

```
{
  "method" : "POST",
  "url" : "https://api.example.com/v1/{project_id}/agent-runtime/upload-file",
  "headers" : {
    "Content-Type" : "application/json",
    "X-Auth-Token" :
"MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQExDTALBglghkgBZQMEAgEwgguVBgkqhkiG...",
    "stream" : true
  },
```


版权所有 © 华为云计算技术有限公司

```
  "body" : {
    "file" : "C:\\Users\\Desktop\\market-CFrwA1xu.png"
  }
}
```

响应示例

状态码：200

Agent文件上传结束的响应体。

```
{
  "url" : "https://test-agent-poc.obs.cn-north-7.ulanqab.huawei.com:443/file/3fd960a8-ca5d-4423-b8da-bb8866e21c28.docx?AccessKeyId=8SL1ZFP1ELHHMAWYJHCJ&Expires=1758282352&Signature=r02Qxi3%2Bhv1FtnMo3XcCvReBQGo%3D",
  "headers" : [ {
    "Host" : "test-agent-poc.obs.cn-north-7.ulanqab.huawei.com:443"
  } ]
}
```

状态码

状态码	描述
200	Agent文件上传结束的响应体。

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 应用示例

5.1 调用工作流应用示例

操作场景

该接口用于运行场景化应用，支持在指定的项目、工作流和对话上下文中执行工作流逻辑。接口支持流式响应模式，可以根据需要返回增量执行结果，适用于实时交互场景。

适用场景：

- 在项目中运行预定义的工作流。
- 支持调试模式和发布模式，适用于不同开发和生产环境。
- 支持流式响应，适用于需要实时反馈的场景（如聊天机器人、实时数据分析等）。

下面介绍如何**调用工作流应用**API使用智能体应用，API的调用方法请参见**如何调用API**。

前提条件

您需要规划Versatile所在的区域信息，并根据区域确定调用API的Endpoint，Versatile目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。

调用工作流

如下示例是调用工作流应用的配置。

```
POST https://api.example.com/v1/12345/workflows/67890/conversations/67890

Request Header:
Content-Type: application/json
X-Auth-Token: MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQExDTALBglghkgBZQMEAgEwgguVBgkqhkiG...
stream: true

Request Body:
{
  "inputs": {
```


版权所有 © 华为云计算技术有限公司

```
  "query": "你好"
 },
 "plugin_configs": [
  {
   "plugin_id": "xxxxxxxxx",
   "config": {
    "key": "value"
   }
  }
 ]
}
```

- endpoint：终端节点，Versatile目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。
- project_id：当前项目ID，获取方法请参考**获取项目ID**。
- workflow_id：工作流应用ID，获取方法如下：
  a. 进入Versatile智能体平台。
  b. 在左侧导航选择“开发中心 > 应用管理 > 工作流应用”。
  c. 在待复制ID的工作流应用卡片上，单击“▪▪▪ > 复制ID”。
- conversation_id：会话ID，每个会话的唯一标识符，可将会话ID设置为任意值，使用标准UUID格式。
  a. 进入Versatile智能体平台。
  b. 在左侧导航选择“开发中心 > 应用管理 > 工作流应用”。
  c. 在待复制调用路径的工作流应用卡片上，单击“▪▪▪ > 调用路径”。
  d. 在弹出的“调用路径”页面，单击“复制路径”即可获取调用路径，如**图5-1**所示。

  其中，conversation_id参数为会话ID，唯一标识每个会话的标识符，可将会话ID设置为任意值，使用标准UUID格式。

图 5-1 获取应用调用路径

5.2 调用智能体应用示例

操作场景

该接口用于运行知识型智能体应用（单智能体应用、多智能体应用），支持在指定的项目、智能体和对话上下文中执行智能体逻辑。接口支持流式响应模式，可以根据需要返回增量执行结果，适用于实时交互场景。

版权所有 © 华为云计算技术有限公司

适用场景：

- 在项目中运行预定义的知识型智能体应用。
- 支持调试模式和发布模式，适用于不同开发和生产环境。
- 支持流式响应，适用于需要实时反馈的场景（如聊天机器人、实时数据分析等）。

下面介绍如何**调用智能体应用**API使用智能体应用，API的调用方法请参见**如何调用API**。

前提条件

您需要规划Versatile所在的区域信息，并根据区域确定调用API的Endpoint，Versatile目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。

调用应用

如下示例是调用智能体应用的配置。

```
POST https://{endpoint}/v1/{project_id}/agents/{agent_id}/conversations/{conversation_id}

Request Header:
Content-Type: application/json
X-Auth-Token: MIINRwYJKoZIhvcNAQcCoIINODCCDTQCAQExDTALBglghkgBZQMEAgEwgguVBgkqhkiG...
stream: true

Request Body:
{
  "query": "查询A12会议室在9:00到10:00的状态"
}
```

- endpoint：终端节点，Versatile目前仅部署在“**华北-北京四**”区域，Endpoint为“aiae.appstage.myhuaweicloud.com”。
- project_id：当前项目ID，获取方法请参考**获取项目ID**。
- agent_id：智能体应用ID，获取方法如下：
  a. 进入Versatile智能体平台。
  b. 在左侧导航选择“开发中心 > 应用管理 > 单智能体应用”或选择“开发中心 > 应用管理 > 多智能体应用”。
  c. 在待复制ID的智能体应用卡片上，单击“▪▪▪ > 复制ID”。
- conversation_id：会话ID，每个会话的唯一标识符，可将会话ID设置为任意值，使用标准UUID格式。
  a. 进入Versatile智能体平台。
  b. 在左侧导航选择“开发中心 > 应用管理 > 单智能体应用”或选择“开发中心 > 应用管理 > 多智能体应用”。
  c. 在待复制调用路径的智能体应用卡片上，单击“▪▪▪ > 调用路径”。
  d. 在弹出的“调用路径”页面，单击“复制路径”即可获取调用路径，如**图5-2**所示。

     其中，conversation_id参数为会话ID，唯一标识每个会话的标识符，可将会话ID设置为任意值，使用标准UUID格式。

版权所有 © 华为云计算技术有限公司

图 5-2 获取应用调用路径

版权所有 © 华为云计算技术有限公司

6 附录

6.1 状态码

状态码如表6-1所示。

表 6-1 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。
200	OK	服务器已成功处理了请求。
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。

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
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。
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。

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
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。
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。

6.2 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
400	IIT.0201	请求参数不合法。	请求参数不合法。	请检查请求参数是否填写正确。
400	IIT.0203	请求参数不合法，输入参数中的数据长度小于训练所用数据长度。	请求参数不合法，输入参数中的数据长度小于训练所用数据长度。	请确认请求body中特征名称、特征数量是否与训练数据中的特征一致。
400	PANGU.0010	请求参数错误。	请求参数错误。	请参考《API文档》输入正确的请求参数，并重新调试API。
400	PANGU.3278	请求参数丢失。	请求参数丢失。	请检查调用API时请求参数是否填写完整、是否有拼写错误、取值是否正确。
400	PANGU.3306	访问的API模型与实例模型不匹配。	访问的API模型与实例模型不匹配。	请检查参数是否正确，或联系服务技术支持协助解决。
400	PANGU.3308	访问的API与已开通的API服务不匹配。	访问的API与已开通的API服务不匹配。	请确认调用的API是否填写错误。
400	PANGU.3317	最大token不合法。	最大token不合法。	请参考《API文档》检查请求参数中输入的token数值是否不在范围内，并重新调试API。
400	PANGU.3318	Content长度不合法。	Content长度不合法。	请参考《API文档》检查请求参数中输入的Content参数长度是否不在范围内，并重新调试API。
400	PANGU.3320	非流式调用推理服务传的参数只能是1或者2。	非流式调用推理服务传的参数只能是1或者2。	请使用正确的取值：1或者2。
400	PANGU.3321	流式调用推理服务n只能取1。	流式调用推理服务n只能取1。	请使用正确的取值：1。

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	APIG.0301	IAM身份验证信息不正确：decrypt token fail：token解析失败。 token expires：token过期。 verify aksk signature fail：AK/SK认证失败。 x-auth-token not found：未找到x-auth-token参数。	IAM身份验证信息不正确：decrypt token fail：token解析失败。 token expires：token过期。 verify aksk signature fail：AK/SK认证失败。 x-auth-token not found：未找到x-auth-token参数。	token解析失败，请检查获取token的方法，请求体信息是否填写正确，token是否正确；检查获取token的环境与调用的环境是否一致。 token超时（token expires），请重新获取token，使用不过期的token。 请检查AK/SK是否正确（AK对应的SK错误，不匹配；AK/SK中多填了空格）。 AK/SK频繁出现鉴权出错，连续错误5次以上，被锁定5分钟（5分钟内，则一直认为其是异常的鉴权请求），5分钟后解锁重新认证。 检查账号权限，是否欠费，被冻结等。 检查调用API时，请求header参数X-Auth-Token是否拼写正确。
401	PANGU.0011	认证失败。	认证失败。	认证鉴权失败，请参考《API文档》“认证鉴权”章节重新进行认证。
401	PANGU.0012	服务内部异常。	服务内部异常。	请联系服务技术支持协助解决。
401	PANGU.3305	获取token错误。	获取token错误。	请检查调用API时使用的token是否完整，是否存在错误。
403	PANGU.3307	账号未开通该API服务。	账号未开通该API服务。	请确认是否已开通该API服务。

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
403	PANGU.3315	API模型实例未公开。	API模型实例未公开。	请检查是否具备使用权限，或联系服务运维人员协助解决。
403	PANGU.3319	权限错误。	权限错误。	请联系服务技术支持协助解决。
404	APIG.0101	访问的API不存在或尚未在环境中发布。	访问的API不存在或尚未在环境中发布。	请检查API的URL是否拼写正确，例如，URL中是否缺少project_id。 HTTP请求方法（POST，GET等）是否正确。
404	PANGU.3254	资源不存在。	资源不存在。	请检查调用API时projectId是否填写正确。
429	APIG.0308	发送请求超过了服务的默认配置限流。	发送请求超过了服务的默认配置限流。	通过重试机制，在代码里检查返回值，碰到并发错误可以延时一小段时间（如2-5s）重试请求。 后端检查上一个请求结果，上一个请求返回之后再发送下一个请求，避免请求过于频繁。
429	PANGU.3267	QPS超出限制。	QPS超出限制。	请降低请求频率。
500	IIT.0202	内部错误。	内部错误。	请联系服务技术支持协助解决。
500	PANGU.0001	未知错误。	未知错误。	请联系服务技术支持协助解决。
500	PANGU.3316	创建代理失败。	创建代理失败。	请联系服务运维人员协助解决。
503	PANGU.3259	推理服务状态异常。	推理服务状态异常。	请检查调用API时deploymentId是否正确，并检查模型的部署状态是否存在异常，如果仍无法解决请联系服务技术支持协助解决。

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
504	APIG.0201	请求超时。	请求超时。	请检查原调用请求是否过于频繁，如果是并发过大，可以通过重试机制解决，在代码里检查返回值，碰到这个并发错误可以延时一小段时间（如2-5s）重试请求；也可以后端检查上一个请求结果，上一个请求返回之后再发送下一个请求，避免请求过于频繁。

6.3 获取项目 ID

调用 API 获取项目 ID

项目ID还可通过调用**查询指定条件下的项目信息**API获取。

获取项目ID的接口为“GET https://{Endpoint}/v3/projects”，其中{Endpoint}为IAM的终端节点，可以从**地区和终端节点**获取。接口的认证鉴权请参见**认证鉴权**。

响应示例如下，其中projects下的“id”即为项目ID。

```
{
    "projects": [
        {
            "domain_id": "65382450e8f64ac0870cd180d14e684b",
            "is_domain": false,
            "parent_id": "65382450e8f64ac0870cd180d14e684b",
            "name": "cn-north-4",
            "description": "",
            "links": {
                "next": null,
                "previous": null,
                "self": "https://www.example.com/v3/projects/a4a5d4098fb4474fa22cd05f897d6b99"
            },
            "id": "a4a5d4098fb4474fa22cd05f897d6b99",
            "enabled": true
        }
    ],
    "links": {
        "next": null,
        "previous": null,
        "self": "https://www.example.com/v3/projects"
    }
}
```

从控制台获取项目 ID

在调用接口的时候，部分URL中需要填入项目编号，所以需要获取到项目编号。项目编号获取步骤如下：

版权所有 © 华为云计算技术有限公司

步骤1 登录管理控制台。

步骤2 鼠标移动到右上角的用户名处，在下拉列表中单击“我的凭证”。

步骤3 在“API凭证”页面的项目列表中查看项目ID。

图 6-1 查看项目 ID

多项目时，展开“所属区域”，从“项目ID”列获取子项目ID。

----结束

6.4 获取账号 ID

在调用接口的时候，部分URL中需要填入账号ID，所以需要先在管理控制台上获取到账号ID。账号ID获取步骤如下：

步骤1 登录管理控制台。

步骤2 鼠标移动到右上角的用户名处，在下拉列表中单击“我的凭证”。

步骤3 在“API凭证”页面中查看账号ID。

图 6-2 获取账号 ID

----结束

6.5 获取工作区 ID

在调用接口的时候，部分URL中需要填入工作区ID，工作区ID获取方法如下。

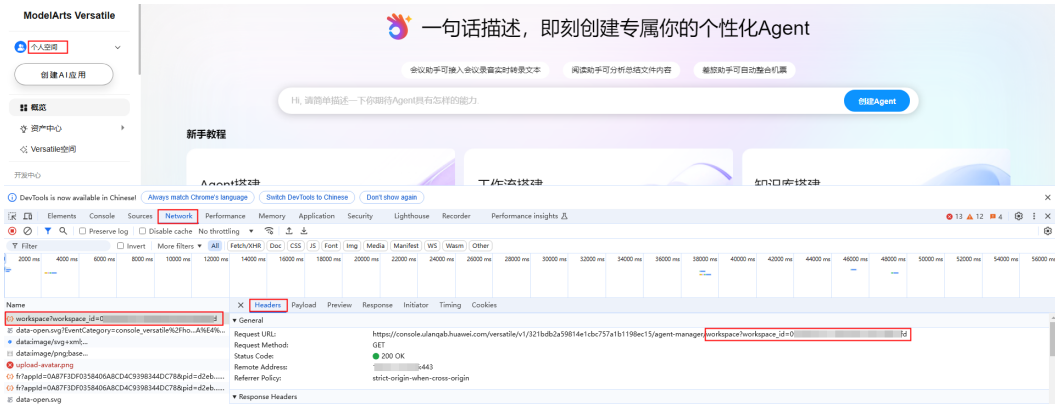
步骤1 进入Versatile智能体平台。

步骤2 打开F12，选择“Network”，单击任意页面，例如，个人空间。

步骤3 可以在接口调用中看到workspace_id=xxx。xxx为工作区ID的值。

版权所有 © 华为云计算技术有限公司

图 6-3 获取工作区 ID



----结束

版权所有 © 华为云计算技术有限公司