内容审核

用户指南

文档版本	01
发布日期	2024-04-23

华为技术有限公司

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。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： https://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：
https://www.huawei.com/cn/psirt/vul-response-process
如企业客户须获取漏洞信息，请参见如下网址：
https://securitybulletin.huawei.com/enterprise/cn/security-advisory

目 录

1 服务使用流程

内容审核（Content Moderation），是基于图像、文本、音频、视频的检测技术，可自动检测涉黄、涉暴、图文违规等内容，用户通过调用API对上传的图片、文字、音视频进行内容审核，获取推理结果，帮助用户打造智能化业务系统提升业务效率。

使用本服务的操作流程如下所示：

图 1-1 使用流程

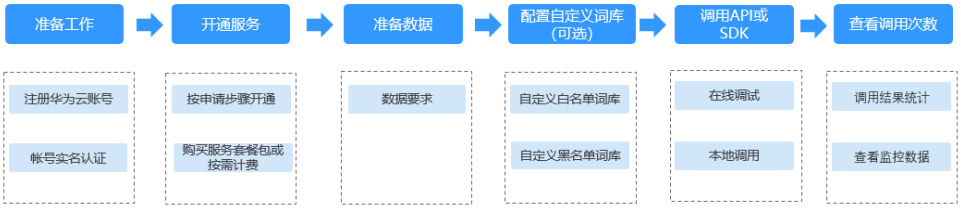


表 1-1 使用流程说明

流程	子任务	说明	详细指导
准备工作	注册华为账号并实名认证	使用内容审核服务之前，您需要注册华为账号并进行实名认证。	**注册华为账号**
开通服务	按申请步骤开通	服务通过工单方式开通，需要您按照步骤操作说明来申请开通服务。	**开通服务**
	购买服务套餐包或按需计费	成功开通服务后需要购买服务，有两种计费方式可供选择。	**购买服务**
准备数据	数据要求	数据格式和调用并发数有相应的约束限制，需要您在使用服务前参考约束准备好待审核的数据。	**准备数据**
配置自定义词库（可选）	自定义白名单词库/自定义黑名单词库	使用**文本内容审核**服务，您可以配置自定义白名单词库或自定义黑名单词库，来帮助您过滤和检测指定文本内容。	**配置自定义词库（可选）**

流程	子任务	说明	详细指导
调用API或SDK	在线调试	以**文本内容审核**为例，介绍如何使用API Explorer调试API。	**在线调试**
	本地调用	介绍使用Moderation SDK在本地进行开发，用户直接调用接口函数即可使用SDK功能。	**本地调用**
查看调用次数	调用结果统计	开始使用服务后，可以在管理控制台上查看服务审核详情和调用次数统计。	**查看调用次数**

2 开通服务

您可以按照如下步骤操作申请开通本服务。

说明

本服务仅面向企业用户开放，个人用户暂不支持开通。

注册华为账号

如果您已完成华为账号注册，可跳过该步骤。

1. 登录**华为云**官方网站。
2. 单击华为云官网右上角“注册”进入注册页面。
3. 在注册页面，根据提示信息完成注册。具体操作可参见**账号注册**。
4. 注册成功后即可自动登录华为云，您需要完成“实名认证”才可以正常使用服务。具体操作可参见**实名认证**。

开通服务

内容审核服务申请开通您可以按照如下步骤操作：

1. 已注册华为账号，并完成实名认证。
2. 登录内容审核管理控制台，控制台左上角默认显示服务部署在“华北-北京四”区域，请您根据业务需要选择对应区域，服务部署的区域具体请参见**终端节点**。
3. 在左侧导航栏选择“服务管理”，进入服务管理页面，进行以下步骤操作：
   a. 单击“申请开通服务”按钮，进入到新建工单页面。

   图 2-1 服务管理页面

   b. 在“我在Moderation遇到问题类型”分类中选择“服务开通”，进入到智能客服对话框中。

图 2-2 服务开通

c. 在对话框中输入“申请开通内容审核服务”，单击“发送”后对话框会出现“转人工”的按钮，选择转人工服务。

图 2-3 转人工

d. 在对话框中智能客服将为您创建工单，输入以下具体信息：

- 问题描述：需要填写：
  1、使用场景（即：申请开通“文本/图像/音频”内容审核）
  2、企业名称（本服务暂只支持企业用户使用）
- 区域：选择想要开通服务的区域。
- 联系方式：输入手机号或邮箱，客服会通过手机或邮箱联系您告知服务开通进展。

输入完成后提交工单，等待客服审核完成后帮您开通本服务。

图 2-4 创建工单

说明

- 服务只需要开通一次即可，后面使用时无需再申请。

4. 商用服务申请成功后，在“服务管理”页面，“我的服务”中显示已经申请开通成功的服务，此时，您可以通过调用API的方式使用内容审核服务。

计费方式

目前内容审核服务提供两种计费模式供您选择：按需计费和预付套餐包计费。具体介绍请参见**计费说明**。

- 按需计费

  如果您想使用按需计费的方式，详细费用价格请参见**内容审核价格详情**。

- 预付套餐包计费

  开通服务后，单击右上角的“预付套餐包”按钮，进入到本服务套餐包购买页面，按需选择想要购买的功能类型和规格，选择完成后单击“立即购买”，确认购买信息无误后完成付款即可开始使用本服务。

图 2-5 预付套餐包

3 准备数据

服务不同功能部署的区域，数据格式和调用并发数有相应的约束限制，需要您在使用服务前参考约束准备好待审核的数据。

服务功能的使用约束请参见**约束与限制**。

例如文本内容审核，输入数据存在以下约束：

- 文本内容审核V2版本：支持“华北-北京一、华北-北京四、华东-上海一”区域，新用户建议使用“华北-北京四”。
- 文本内容审核V3版本：支持“华北-北京一、华北-北京四、华东-上海一”区域，新用户建议使用“华北-北京四”。
- 只支持中文文本内容审核。
- 默认API调用最大并发为50，如需调整更高并发限制请通过工单联系专业工程师为您服务。

说明

服务所支持的区域是指服务部署在该区域下的服务器，用户所在地区与服务部署区域不一致也是可以开通和使用本服务的。有如下两种情况：

1. 如果请求输入的数据是OBS地址方式，就需要使用相同区域的内容审核服务。
   例如：您的OBS请求数据在“华北-北京四”，只能调用“华北-北京四”区域下的内容审核服务，如果本服务不支持该区域则不能调用。
2. 如果请求输入的数据是Base64图片或者公网URL，则不受区域影响。
   例如：您的服务器在“华东-上海一”可以调用“华北-北京四”的内容审核服务接口。

4 配置自定义词库（可选）

使用**文本内容审核**服务前，您可以配置自定义白名单词库或自定义黑名单词库，来帮助您过滤和检测指定文本内容。

配置自定义词库 V2请看**具体操作**。

配置自定义词库 V3请看**具体操作**。

5 调用 API 或 SDK

5.1 在线调试

功能介绍

API Explorer在线调试工具提供API的检索、调试、代码示例生成功能。同时，集成开发环境CloudIDE，可完成代码的构建、调试、运行。

本章节以**文本内容审核**为例，介绍如何使用API Explorer调试API。

前提条件

1. 已注册华为账号，并完成实名认证，账号不能处于欠费、冻结、被注销等异常状态。
2. 了解**文本内容审核约束限制**。
3. 已**开通文本内容审核服务**。

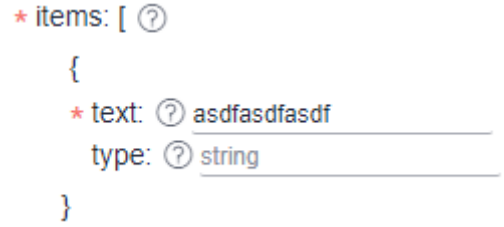
操作步骤

1. 登录**API Explorer**。

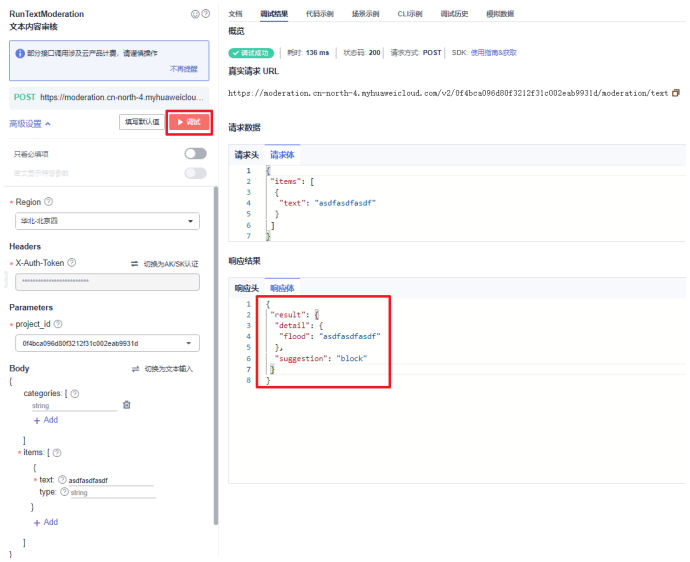
   登录后，“X-Auth-Token”和“project_id”参数会自动填充，无需填写。
2. 填写待检测文本数据，带*号的是必须要填的参数。

   text表示要检测的文本，文本的编码格式为“utf-8”，最多检测5000个字符，具体的文本检测格式要求请参见**文本内容审核API**。

   例如：输入text的值为asdfasdfasdf。

3. 单击“调试”按钮，获取识别结果。



5.2 本地调用

内容审核软件开发工具包（Moderation SDK）是对内容审核提供的REST API进行的封装，以简化用户的开发工作。用户通过添加依赖或下载的方式调用API即可实现使用内容审核业务能力的目的。

本章节以**文本内容审核**为例，介绍如何使用Moderation Python SDK在本地进行开发，用户直接调用接口函数即可使用SDK功能。

其他审核功能及支持的SDK列表可参见**内容审核SDK参考**。

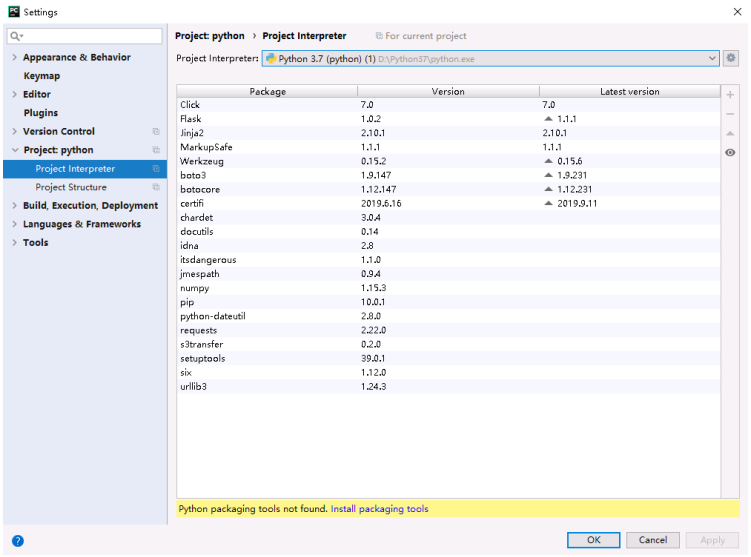
前提条件

- 已注册华为账号，并完成实名认证，账号不能处于欠费、冻结、被注销等异常状态。
- 了解**文本内容审核约束限制**。
- 已**开通文本内容审核服务**。

操作步骤

1. 安装Python环境并获取SDK软件包。
   a. 从**Python官网**下载并安装合适的Python版本。请使用Python3.3以上版本，如下以Python3.7 版本为例进行说明。
   b. 从**PyCharm官网**下载并安装最新版本。
   c. 在PyCharm开发工具中配置Python环境，在菜单依次选择“File > Settings > Project Interpreter”。
   d. 在页面上方选择您的Python安装路径，如图 **PyCharm配置python环境所示**。选择好目标Python之后单击页面下方“Apply”完成配置。

图 5-1 PyCharm 配置 python 环境



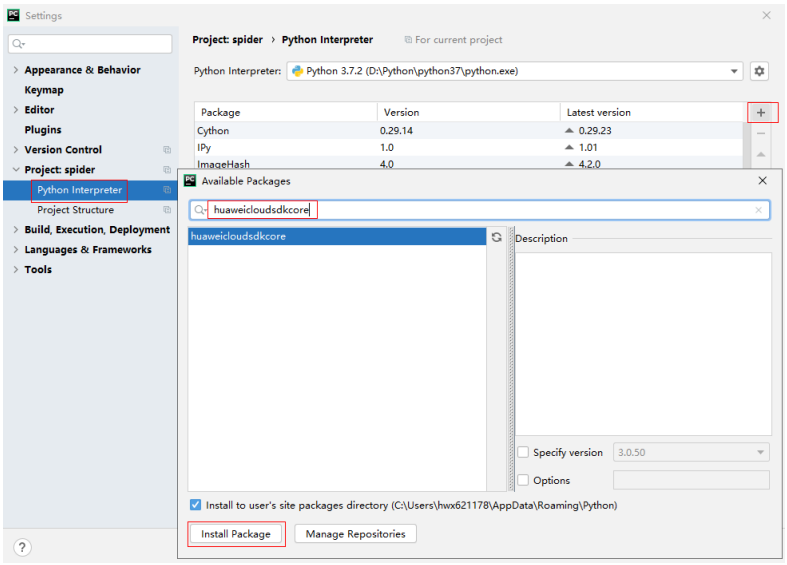
2. 在PyCharm中新建一个项目，并单击左下方“Terminal”按钮。分别执行以下命令安装SDK（该SDK支持Python3及以上版本）。参考方法如下：

```
pip 安装：
# 安装核心库
pip install huaweicloudsdkcore

# 安装Moderation服务库
pip install huaweicloudsdkmoderation
```

在pycharm中，选择“File > Settings > Project > Python Interpreter”单击右上角+，分别搜索huaweicloudsdkcore及huaweicloudsdkmoderation，搜索到包内容单击左下角Install Package完成安装。

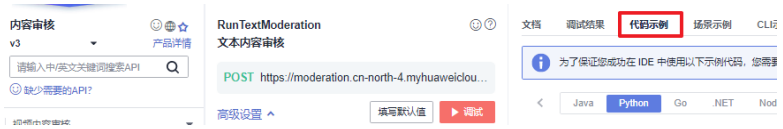
图 5-2 pycharm 安装内容审核 python 版本 sdk 包



3. 获取文本内容审核SDK示例代码。

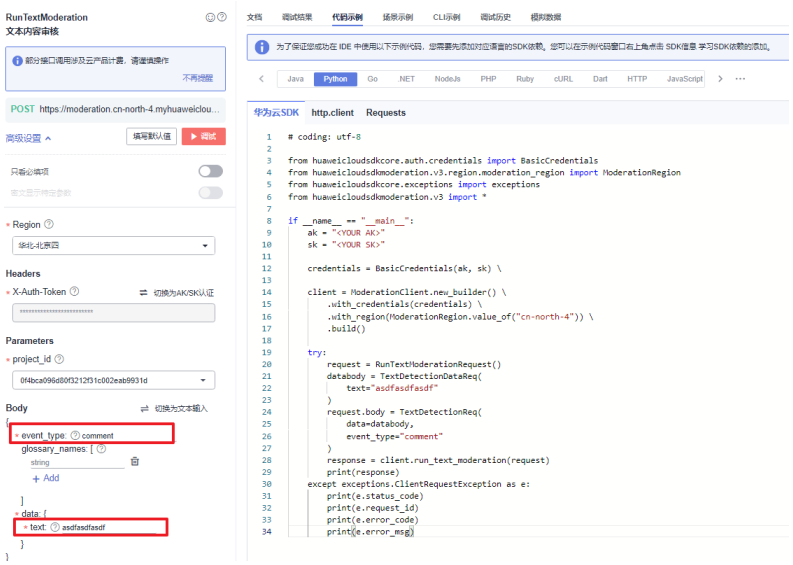
   a. 登录**API Explorer**，在“代码示例”中选择“Python”。

图 5-3 代码示例



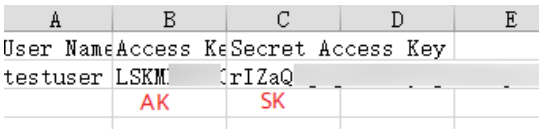
b. 填写请求Body参数：event_type（事件类型）和text（待检测文本）。
例如：event_type输入comment，text输入asdfasdfasdf。

图 5-4 填写参数



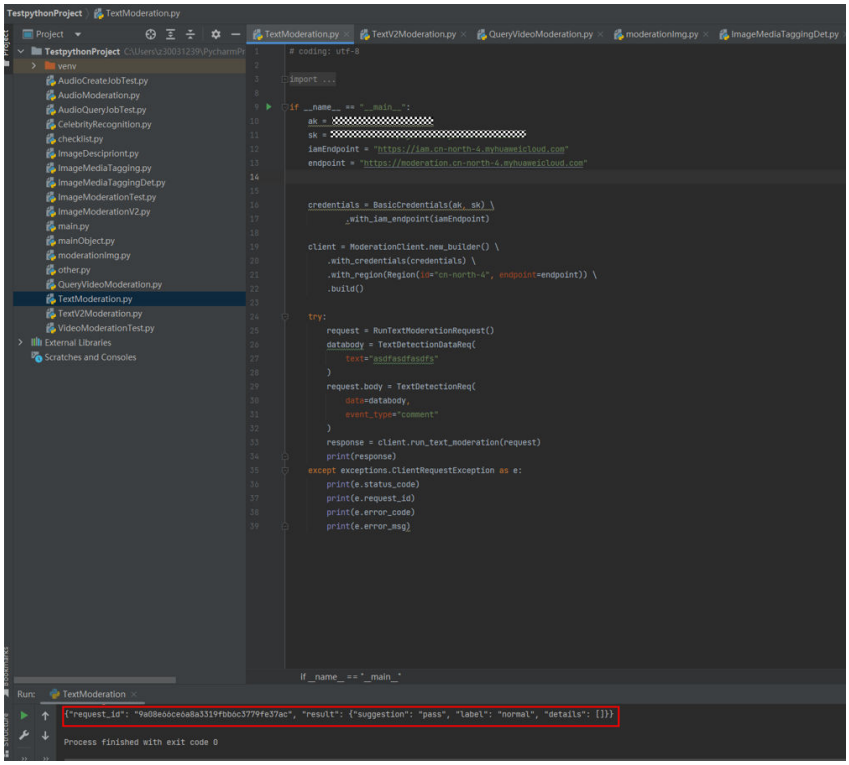
c. 复制代码示例至PyCharm中。

4. 获取AK/SK，替换代码示例中的“<YOUR AK>”、“<YOUR SK>”参数。
登录**访问密钥**页面，新增访问密钥，或使用已有的访问密钥。访问密钥为credentials.csv文件，包含AK/SK信息。



5. 运行代码示例，获取识别结果。您可根据响应参数说明来解读审核结果的含义，具体可参考**文本内容审核结果**。

图 5-5 运行示例



6 查看调用次数

功能介绍

您可以在内容审核服务管理控制台上查看服务审核详情和调用次数统计，帮助您更好了解服务的审核情况和调用情况。

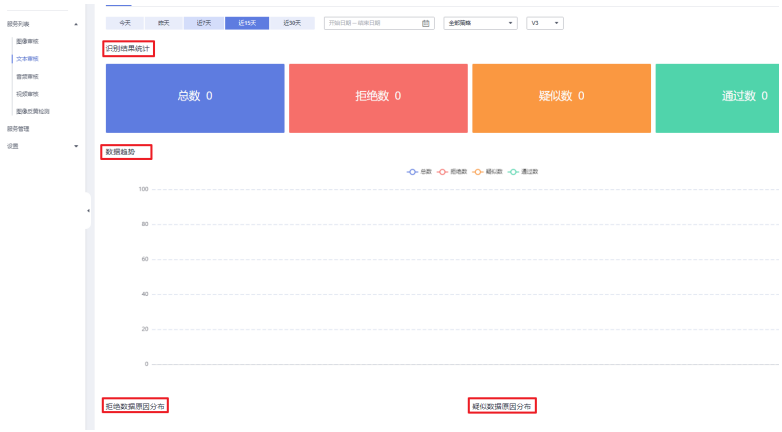
说明

该功能适用于文本/图像/音频/视频审核。

操作步骤

1. 登录内容审核服务管理控制台。
2. 在左侧导航栏中选择“服务列表>文本审核”，可以查看识别统计详情，如图6-1所示。您可以设置时间范围，策略（事件类型）来观察这段时间内的调用次数变化情况。

图 6-1 识别统计



- 识别结果统计：显示一段时间范围，内容审核的调用总数，拒绝数，疑似数和通过数，帮助您更好了解服务的调用情况和审核情况。
  - 总数：指的是审核调用总次数。
  - 拒绝数：指的是block总数，即文本中包含敏感信息，审核不通过的次数。

- 疑似数：指的是review总数，即人工复查审核的次数。
- 通过数：指的是pass总数，即通过审核的次数。

- 数据趋势：显示您设置的这段时间范围内，总数，拒绝数，疑似数和通过数的变化趋势。
- 拒绝数据原因分布：显示您设置的这段时间范围内，审核不通过的检测场景占比数。
- 疑似数据原因分布：显示您设置的这段时间范围内，需要人工复查的检测场景占比数。

